

Supporting Information

Multi-Assay-based Compound Prioritization via Assistance Utilization: a Machine Learning Framework

Junfeng Liu[†] and Xia Ning^{*,†,‡}

[†]*Indiana University-Purdue University, Indianapolis*

723 West Michigan St, SL 280, Indianapolis, IN 46202, USA

[‡]*Center of Computational Biology and Bioinformatics, Indiana University School of
Medicine*

410 West 10th St, HITS 5000, Indianapolis, IN 46202, USA

E-mail: xning@iupui.edu

S1 Ranking List Alignment

Algorithm S1 presents the algorithm to align two ranking lists. The algorithm takes the essential ideas from sequence alignment with a key difference on the recurrence as in Equation S3. In specific, a discount is applied in Equation S3 (i.e., d_i and d_j), which promotes alignment over the top-ranked compounds in the lists. In addition, the match score over two non-identical compounds is allowed as in Algorithm S2, in which the match score can be calculated from compound similarities. The default match/mismatch values are in Table S1.

Table S1: Default Scores in Ranking Alignment

match (s_m)	mismatch (s_s)	gap (s_g)
1.0	0.5	0.2

Algorithm S1 Ranking List Alignment Algorithm

```

1: function RANKING_ALIGNMENT( $r_1, r_2, s_g, \text{IsDset}$ )
2:
3:    $m \leftarrow \text{size}(r_1) + 1$ 
4:    $n \leftarrow \text{size}(r_2) + 1$ 
5:    $S \leftarrow m \times n$  matrix
6:
7:   for  $i$  from 0 to  $m - 1$  do
8:      $S[i][0] \leftarrow i \times s_g$ 
9:   end for
10:
11:  for  $j$  from 0 to  $n - 1$  do
12:     $S[0][j] \leftarrow j \times s_g$ 
13:  end for
14:
15:   $d_i \leftarrow 1.0$ 
16:   $d_j \leftarrow 1.0$ 
17:
18:  for  $i$  from 1 to  $m - 1$  do
19:    for  $j$  from 1 to  $n - 1$  do
20:      if  $\text{IsDset}$  then
21:
22:
23:      end if
24:       $s_{mch} \leftarrow \text{match\_score}(r_1[i - 1], r_2[j - 1], \text{mode})$   $\triangleright$  as in Algorithm S2
25:
26:       $S[i][j] \leftarrow \max \left( \begin{array}{l} S[i - 1][j] + d_i s_g \\ S[i][j - 1] + d_j s_g \\ S[i - 1][j - 1] + \sqrt{d_i d_j} s_{mch} \end{array} \right)$  (S3)
26:    end for
27:  end for
28:
29:  return  $S[m - 1][n - 1]$ 
30: end function

```

Algorithm S2 Match Score Calculation

```
1: function MATCH_SCORE( $c_1, c_2, \text{mode}$ )
2:   if  $\text{mode} == \text{alinS}_{\text{cidn}}$  then
3:     
$$s \leftarrow \begin{cases} s_m & \text{if } c_1 == c_2 \\ s_s & \text{otherwise} \end{cases} \quad (\text{S4})$$

4:   end if
5:   if  $\text{mode} == \text{alinS}_{\text{csim}}$  then
6:     
$$s \leftarrow \text{Tanimoto}(c_1, c_2) \quad (\text{S5})$$

7:   end if
8:
9:   return  $s$ 
10: end function
```

S2 Aggregated Compound Similarities

The aggregated compound similarities are defined as follows.

$$\text{cSim}^{\max}(c, \mathcal{C}_i) = \max_{c' \in \mathcal{C}_i} \text{Tanimoto}(c, c') \quad (\text{S6})$$

$$\text{cSim}^{\min}(c, \mathcal{C}_i) = \min_{c' \in \mathcal{C}_i} \text{Tanimoto}(c, c') \quad (\text{S7})$$

$$\text{cSim}^{\text{avg}}(c, \mathcal{C}_i) = \frac{1}{|\mathcal{C}_i|} \sum_{c' \in \mathcal{C}_i} \text{Tanimoto}(c, c') \quad (\text{S8})$$

$$\text{cSim}^{\text{pos}}(c, \mathcal{C}_i) = \frac{1}{|\mathcal{C}_i|} \sum_{i=1, \dots, |\mathcal{C}_i|} \frac{1}{i} \text{Tanimoto}(c, c_i) \quad (\text{S9})$$

Algorithm S3 Assistance Compound Score Assignment

```
1: function ASSIGN_SCORE( $B_i, RM_i, \mathcal{C}_i, \mathcal{C}_i^+$ )
2:
3:    $\tilde{r}_i^+ \leftarrow$  apply  $RM_i$  on  $\mathcal{C}_i^+$ 
4:
5:   for  $k$  from 1 to  $\text{size}(\tilde{r}_i^+)$  do
6:     if  $c_k \in \mathcal{C}_i^+ \setminus \mathcal{C}_i$  then  $\triangleright c_k$  is an assistance compound
7:        $p \leftarrow \arg \max_{c_p \in \mathcal{C}_i, c_p \succ_{\tilde{r}_i^+} c_k} p$   $\triangleright$  find  $c_p \in \mathcal{C}_i$  that is closest to  $c_k$  among higher-ranked
       compounds than  $c_k$  in  $\tilde{r}_i^+$ 
8:        $q \leftarrow \arg \min_{c_q \in \mathcal{C}_i, c_k \succ_{\tilde{r}_i^+} c_q} q$   $\triangleright$  find  $c_q \in \mathcal{C}_i$  that is closest to  $c_k$  among lower-ranked
       compounds than  $c_k$  in  $\tilde{r}_i^+$ 
9:        $c_k$ 's score  $\leftarrow$  linearly interpolation from  $c_p$  and  $c_q$ 's true scores in  $B_i$ 
10:    end if
11:  end for
12:
13: end function
```

S3 Decision Tree Learning

Decision trees represent a type of rule-based classifiers that can be used for multi-class classification. Each internal node in a decision tree represents a decision rule over a certain feature of testing instances that directs the classification for the instances. The classification process on a decision tree starts from routing the testing instances from the tree root. The leaf nodes that testing instances are ended at following the decision rules along the tree indicate the instance classes. An outstanding advantage of decision trees is on their interpretability, as the paths that testing instances are routed along provides the reasoning for the classification results.

We utilized a DecisionTreeClassifier function from Scikit-learn package ¹ in Python. We specified two options: the maximum number of levels in the tree, and the minimum leaf node size (i.e. minimum number of samples that fall into the leaf node), with all the other options as their default values. We set the maximum level of the tree as 6, and minimum number of samples in each leaf node as 5.

¹<http://scikit-learn.org/stable/>

S4 Assay Information

Table S2 lists the IDs of assays that are used in the experiments, their protein target and encoding genes.

Table S2: Bioassay Information

ID	protein target	gene
690	placental-like alkaline phosphatase preproprotein	ALPPL2
30868	Acetylcholinesterase	ACHE
35397	Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1
37646	Angiotensin-converting enzyme	ACE
37647	Angiotensin-converting enzyme	ACE
37659	Angiotensin-converting enzyme	ACE
37660	Angiotensin-converting enzyme	ACE
37667	Angiotensin-converting enzyme	ACE
37669	Angiotensin-converting enzyme	ACE
39498	Type-1 angiotensin II receptor	AGTR1
39648	Type-1 angiotensin II receptor	AGTR1
39801	Type-1 angiotensin II receptor	AGTR1
39807	Type-1 angiotensin II receptor	AGTR1
39808	Type-1 angiotensin II receptor	AGTR1
39809	Type-1 angiotensin II receptor	AGTR1
39812	Type-1 angiotensin II receptor	AGTR1
39820	Type-1 angiotensin II receptor	AGTR1
42316	Translocator protein	TSPO
47369	Glucocorticoid receptor	NR3C1
50708	Beta-lactamase	AMPC

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
55683	dihydrofolate reductase [Escherichia coli]	-
55702	Dihydrofolate reductase	-
55820	Dihydrofolate reductase	-
55827	Dihydrofolate reductase	-
55830	Dihydrofolate reductase	-
55833	Dihydrofolate reductase	-
55843	Dihydrofolate reductase	-
55844	Dihydrofolate reductase	-
55855	Dihydrofolate reductase	-
56169	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56172	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56178	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56183	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56184	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56186	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56189	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56310	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56314	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56318	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56325	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56330	Bifunctional dihydrofolate reductase-thymidylate synthase	-
56345	Bifunctional dihydrofolate reductase-thymidylate synthase	-
57074	Dihydrofolate reductase	DFR1
57075	Dihydrofolate reductase	DFR1

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
57093	dihydrofolate reductase [Escherichia coli]	DFRA17
57094	dihydrofolate reductase [Escherichia coli]	DFRA17
57097	dihydrofolate reductase [Escherichia coli]	DFRA17
57107	Dihydrofolate reductase	DFR1
58100	Dihydrofolate reductase	DFR1
58101	Dihydrofolate reductase	FOLA
78155	Histamine H1 receptor	HRH1
86581	Histamine H1 receptor	HRH1
90359	HD2 type histone deacetylase HDA106	HDA106
90682	HD2 type histone deacetylase HDA106	HDA106
90687	HD2 type histone deacetylase HDA106	HDA106
91424	Integrase	POL
91425	Integrase	POL
91426	Integrase	POL
92730	Disintegrin and metalloproteinase domain-containing protein 17	ADAM17
104448	Reverse transcriptase/RNaseH	POL
143393	Reverse transcriptase/RNaseH	POL
144372	NH(3)-dependent NAD(+) synthetase	NADE
149742	Mu-type opioid receptor	OPRM1
149865	Mu-type opioid receptor	OPRM1
157306	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
160735	Prostaglandin G/H synthase 2	PTGS2
161281	Potassium voltage-gated channel subfamily H member 2	KCNH2
161283	Potassium voltage-gated channel subfamily H member 2	KCNH2

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
162143	Prostaglandin G/H synthase 1	PGHS1
162144	Prostaglandin G/H synthase 1	PGHS1
162176	Prostaglandin G/H synthase 1	PGHS1
179927	Sodium-dependent serotonin transporter	SLC6A4
180514	Voltage-dependent L-type calcium channel subunit alpha-1C	CACNA1C
196387	Sodium-dependent serotonin transporter	SLC6A4
197775	Reverse transcriptase/RNaseH	POL
197777	Reverse transcriptase/RNaseH	POL
197782	Reverse transcriptase/RNaseH	POL
197795	Reverse transcriptase/RNaseH	POL
197925	Reverse transcriptase/RNaseH	POL
197932	Reverse transcriptase/RNaseH	POL
197933	Reverse transcriptase/RNaseH	POL
197934	Reverse transcriptase/RNaseH	POL
197941	Reverse transcriptase/RNaseH	POL
197943	Reverse transcriptase/RNaseH	POL
197944	Reverse transcriptase/RNaseH	POL
197946	Reverse transcriptase/RNaseH	POL
197948	Reverse transcriptase/RNaseH	POL
198086	Reverse transcriptase/RNaseH	POL
198094	Reverse transcriptase/RNaseH	POL
198247	Reverse transcriptase/RNaseH	POL
198385	Reverse transcriptase/RNaseH	POL
198409	Reverse transcriptase/RNaseH	POL

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
198410	Reverse transcriptase/RNaseH	POL
198411	Reverse transcriptase/RNaseH	POL
212727	Trypanothione reductase	TPR
224348	Platelet-activating factor receptor	PTAFR
240604	Thymidylate synthase	TYMS
240624	Estrogen receptor beta	ESR2
240749	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
240835	Bifunctional dihydrofolate reductase-thymidylate synthase	-
240892	Dihydrofolate reductase	FOLA
240926	Dihydrofolate reductase	-
240961	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
241195	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
241201	Estrogen receptor beta	ESR2
241231	Estrogen receptor	ESR1
241308	Prostaglandin G/H synthase 2	PTGS2
241337	Estrogen receptor beta	ESR2
241475	Cannabinoid receptor 1	CNR1
241574	Carbonic anhydrase 5A, mitochondrial	CA5A
241604	Carbonic anhydrase 2	CA2
241692	Acetylcholinesterase	ACHE
241793	3-hydroxy-3-methylglutaryl-coenzyme A reductase	HMGCR
241825	Estrogen receptor beta	ESR2
241852	Estrogen receptor	ESR1
241865	Estrogen receptor beta	ESR2

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
241889	Cytochrome P450 2D6	CYP2D6
241899	Estrogen receptor	ESR1
242043	Estrogen receptor beta	ESR2
242065	MO15-related protein kinase Pfmrk	PFMRK
242068	Estrogen receptor	ESR1
242135	Estrogen receptor beta	ESR2
242151	Estrogen receptor	ESR1
242353	Estrogen receptor beta	ESR2
242369	Estrogen receptor	ESR1
242577	Cannabinoid receptor 1	CNR1
242795	Cytochrome P450 11B2, mitochondrial	CYP11B2
242797	Cytochrome P450 11B1, mitochondrial	CYP11B1
254777	Estrogen receptor beta	ESR2
254783	Estrogen receptor	ESR1
254970	Vacuolar-type proton translocating pyrophosphatase 1	PPASE1
255079	Adenosine receptor A3	ADORA3
255080	Cytochrome P450 11B1, mitochondrial	CYP11B1
255141	Vascular endothelial growth factor receptor 2	KDR
257779	Androgen receptor	AR
258959	Dipeptidyl peptidase 4	DPP4
258960	Dipeptidyl peptidase 8	DPP8
259059	Serine/threonine-protein kinase B-raf	BRAF
259060	Serine/threonine-protein kinase B-raf	BRAF
260151	Estrogen receptor	ESR1

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
260152	Estrogen receptor beta	ESR2
260157	Androgen receptor	AR
260895	Epidermal growth factor receptor	EGFR
260896	Receptor tyrosine-protein kinase erbB-2	ERBB2
260898	Epidermal growth factor receptor	EGFR
261405	Prostaglandin G/H synthase 2	PTGS2
261732	DNA topoisomerase 4 subunit A	PARC
262046	Sodium-dependent serotonin transporter	SLC6A4
262047	Sodium-dependent dopamine transporter	SLC6A3
262048	Sodium-dependent noradrenaline transporter	SLC6A2
262172	Estrogen receptor	ESR1
262173	Estrogen receptor beta	ESR2
262754	Acetylcholinesterase	ACHE
262755	Cholinesterase	BCHE
262948	Estrogen receptor beta	ESR2
262950	Estrogen receptor	ESR1
263980	Epidermal growth factor receptor	EGFR
263981	Receptor tyrosine-protein kinase erbB-2	ERBB2
264807	Epidermal growth factor receptor	EGFR
264808	Epidermal growth factor receptor	EGFR
265578	Androgen receptor	AR
266160	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
267726	Inosine-5'-monophosphate dehydrogenase 2	IMPDH2
267730	Estrogen receptor	ESR1

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
267731	Estrogen receptor beta	ESR2
267771	Glucocorticoid receptor	NR3C1
267774	Glucocorticoid receptor	NR3C1
268271	Sodium-dependent serotonin transporter	SLC6A4
268714	Sodium-dependent serotonin transporter	SLC6A4
268715	Sodium-dependent noradrenaline transporter	SLC6A2
269184	Estrogen receptor	ESR1
269185	Estrogen receptor beta	ESR2
269691	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
269937	Sodium-dependent serotonin transporter	SLC6A4
269938	Sodium-dependent noradrenaline transporter	SLC6A2
270404	Estrogen receptor	ESR1
270405	Estrogen receptor beta	ESR2
270514	Substance-P receptor	TACR1
270616	Receptor tyrosine-protein kinase erbB-2	ERBB2
270617	Epidermal growth factor receptor	EGFR
271466	Substance-P receptor	TACR1
271946	Tyrosine-protein kinase Lck	LCK
273141	Mu-type opioid receptor	OPRM1
273143	Delta-type opioid receptor	OPRD1
273145	Kappa-type opioid receptor	OPRK1
273166	HD2 type histone deacetylase HDA106	HDA106
273168	Histone deacetylase	HD1B
274015	Histone deacetylase 1	HDAC1

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
274062	Histone deacetylase 1	HDAC1
274110	Histone deacetylase 1	HDAC1
274396	Mu-type opioid receptor	OPRM1
274572	Cannabinoid receptor 1	CNR1
275442	Sodium-dependent dopamine transporter	SLC6A3
276317	Androgen receptor	AR
276767	Glucocorticoid receptor	NR3C1
282431	Prostaglandin G/H synthase 1	PTGS1
282432	Prostaglandin G/H synthase 2	PTGS2
282900	Aromatase	CYP19A1
284206	Androgen receptor	AR
284208	Androgen receptor	AR
289392	Cholinesterase	BCHE
290229	Androgen receptor	AR
290288	Cannabinoid receptor 2	CNR2
290289	Cannabinoid receptor 1	CNR1
290408	Vascular endothelial growth factor receptor 2	KDR
292040	Inosine-5'-monophosphate dehydrogenase 2	IMPDH2
293006	Glucocorticoid receptor	NR3C1
293101	Voltage-dependent N-type calcium channel subunit alpha-1B	CACNA1B
294498	Progesterone receptor	PGR
294499	Glucocorticoid receptor	NR3C1
296092	Dipeptidyl peptidase 4	DPP4
296876	Equilibrative nucleoside transporter 1	SLC29A1

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
297967	Neuronal acetylcholine receptor subunit alpha-7	CHRNA7
298278	Acetylcholinesterase	ACHE
298279	Cholinesterase	BCHE
299948	Thyroid hormone receptor beta	THRB
300916	Glucocorticoid receptor	NR3C1
300918	Mineralocorticoid receptor	NR3C2
301059	Substance-P receptor	TACR1
304223	Prostaglandin G/H synthase 1	PTGS1
304251	NAD(P)H dehydrogenase [quinone] 1	NQO1
304252	NAD(P)H dehydrogenase [quinone] 1	NQO1
306462	Estrogen receptor	ESR1
306463	Estrogen receptor beta	ESR2
307629	Steroid hormone receptor ERR1	ESRRA
307630	Steroid hormone receptor ERR2	ESRRB
308376	Histone deacetylase 1	HDAC1
309702	3-hydroxy-3-methylglutaryl-coenzyme A reductase	HMGCR
312172	3-hydroxy-3-methylglutaryl-coenzyme A reductase	HMGCR
313111	D(1A) dopamine receptor	DRD1
313112	D(2) dopamine receptor	DRD2
314091	Acetylcholinesterase	ACHE
314412	Sodium channel protein type 9 subunit alpha	SCN9A
315276	Intermediate conductance calcium-activated potassium channel protein 4	KCNN4
315992	Prostaglandin G/H synthase 2	PTGS2

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
316695	Androgen receptor	AR
316896	Histone deacetylase 1	HDAC1
317007	Cannabinoid receptor 1	CNR1
317008	Cannabinoid receptor 2	CNR2
318593	Farnesyl pyrophosphate synthase	FDPS
318594	Farnesyl pyrophosphate synthase	FDPS
319592	Androgen receptor	AR
319593	Sodium-dependent noradrenaline transporter	SLC6A2
319594	Sodium-dependent serotonin transporter	SLC6A4
319764	Sodium-dependent noradrenaline transporter	SLC6A2
319765	Sodium-dependent serotonin transporter	SLC6A4
321308	Sodium-dependent noradrenaline transporter	SLC6A2
321309	Sodium-dependent serotonin transporter	SLC6A4
322229	Cannabinoid receptor 1	CNR1
325959	Mu-type opioid receptor	OPRM1
326367	Multidrug resistance protein 1	ABCB1
326368	Multidrug resistance protein 1	ABCB1
326396	Progesterone receptor	PGR
326467	Indoleamine 2,3-dioxygenase 1	IDO1
330796	Type-1A angiotensin II receptor	AGTR1A
339393	Sodium-dependent noradrenaline transporter	SLC6A2
339521	Androgen receptor	AR
340414	Aromatase	CYP19A1
341470	Bifunctional dihydrofolate reductase-thymidylate synthase	-

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
341476	Dihydrofolate reductase	DHFR
343578	Sodium-dependent dopamine transporter	SLC6A3
343579	Sodium-dependent serotonin transporter	SLC6A4
343580	Sodium-dependent noradrenaline transporter	SLC6A2
343839	Norepinephrine transporter	SLC6A2
343840	Sodium-dependent serotonin transporter	SLC6A4
343841	Sodium-dependent dopamine transporter	SLC6A3
345340	Cannabinoid receptor 1	CNR1
345341	Cannabinoid receptor 2	CNR2
345876	Epidermal growth factor receptor	EGFR
345893	Progesterone receptor	PGR
346980	Amine oxidase [flavin-containing] A	MAOA
347221	Prostaglandin G/H synthase 1	PTGS1
348792	Histone deacetylase 1	HDAC1
349306	Dipeptidyl peptidase 4	DPP4
352215	Sodium-dependent noradrenaline transporter	SLC6A2
352217	Sodium-dependent serotonin transporter	SLC6A4
353146	Renin	REN
353290	Hydroxycarboxylic acid receptor 2	HCAR2
353291	Hydroxycarboxylic acid receptor 2	HCAR2
353308	Hydroxycarboxylic acid receptor 2	HCAR2
354019	Fructose-1,6-bisphosphatase 1	FBP1
354245	Cannabinoid receptor 1	CNR1
359909	Sodium-dependent dopamine transporter	SLC6A3

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
361789	Acetylcholinesterase	ACHE
361790	Butyrylcholinesterase	BCHE
362025	Sodium-dependent noradrenaline transporter	SLC6A2
362026	Sodium-dependent serotonin transporter	SLC6A4
362062	Sodium-dependent noradrenaline transporter	SLC6A2
362063	Sodium-dependent serotonin transporter	SLC6A4
362471	Steroid 17-alpha-hydroxylase/17,20 lyase	CYP17A1
362929	Cannabinoid receptor 2	CNR2
362931	Cannabinoid receptor 1	CNR1
364560	Dipeptidyl peptidase 4	DPP4
364670	Acetylcholinesterase	ACHE
364887	ATP-binding cassette sub-family G member 2	ABCG2
367003	Amine oxidase [flavin-containing] A	MAOA
367273	Amine oxidase [flavin-containing] B	MAOB
371780	Renin	REN
371781	Renin	REN
372021	Arachidonate 5-lipoxygenase	ALOX5
372022	Arachidonate 5-lipoxygenase	ALOX5
384316	Histone deacetylase	HD1B
384740	3-hydroxy-3-methylglutaryl-coenzyme A reductase	HMGCR
386533	Cannabinoid receptor 1	CNR1
386535	Cannabinoid receptor 1	CNR1
386625	Solute carrier family 22 member 1	SLC22A1
386739	Cholinesterase	BCHE

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
388703	Sodium-dependent serotonin transporter	SLC6A4
389657	Kappa-type opioid receptor	OPRK1
390722	Hormone-sensitive lipase	LIPE
391371	Geranylgeranyl pyrophosphate synthase	GGPS1
393050	C-C chemokine receptor type 5	CCR5
393561	Dipeptidyl peptidase 4	DPP4
395483	C-C chemokine receptor type 5	CCR5
395657	Acetylcholinesterase	ACHE
395658	Butyrylcholinesterase	BCHE
397266	Cannabinoid receptor 1	CNR1
397267	Cannabinoid receptor 2	CNR2
397743	Potassium voltage-gated channel subfamily H member 2	KCNH2
405937	Progesterone receptor	PGR
408340	Potassium voltage-gated channel subfamily H member 2	KCNH2
409942	Amine oxidase [flavin-containing] A	MAOA
410404	Sodium-dependent noradrenaline transporter	SLC6A2
410405	Sodium-dependent serotonin transporter	SLC6A4
410413	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
410466	Cannabinoid receptor 1	CNR1
411213	Sodium-dependent dopamine transporter	SLC6A3
411890	Potassium voltage-gated channel subfamily A member 5	KCNA5
412780	Aromatase	CYP19A1
414717	Histone deacetylase 1	HDAC1
414980	Histone deacetylase	HDAC1

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
416572	Amine oxidase [flavin-containing] A	MAOA
416955	Cannabinoid receptor 1	CNR1
417910	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	PDE4B
419199	Prostaglandin E2 receptor EP2 subtype	PTGER2
419781	Heat shock protein HSP 90-beta	HSP90AB1
420358	Glucocorticoid receptor	NR3C1
420668	Potassium voltage-gated channel subfamily H member 2	KCNH2
423861	Sodium-dependent noradrenaline transporter	SLC6A2
423862	Sodium-dependent serotonin transporter	SLC6A4
429130	Glucocorticoid receptor	NR3C1
431881	Sodium-dependent noradrenaline transporter	SLC6A2
432010	Arachidonate 5-lipoxygenase-activating protein	ALOX5AP
432545	Prostaglandin G/H synthase 2	PTGS2
433783	Cannabinoid receptor 1	CNR1
436917	Dihydrofolate reductase	DFR1
436918	Dihydrofolate reductase	DHFR
441347	Sodium-dependent serotonin transporter	SLC6A4
441348	Sodium-dependent noradrenaline transporter	SLC6A2
441648	NAD(P)H dehydrogenase [quinone] 1	NQO1
441649	NAD(P)H dehydrogenase [quinone] 1	NQO1
441697	Monoglyceride lipase	MGLL
441756	Glutathione S-transferase P	GSTP1
444031	Cholinesterase	BCHE
444520	Sodium-dependent dopamine transporter	SLC6A3

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
444522	Sodium-dependent noradrenaline transporter	SLC6A2
445786	Cannabinoid receptor 1	CNR1
448555	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	PDE4B
452686	Cannabinoid receptor 1	CNR1
452689	Renin	REN
452690	Renin	REN
453456	Fructose-1,6-bisphosphatase 1	FBP1
453457	Fructose-1,6-bisphosphatase 1	FBP1
454342	Acetylcholinesterase	ACHE
454488	Voltage-dependent calcium channel alpha-2 delta subunit precursor	
454554	Free fatty acid receptor 2	FFAR2
455690	Cannabinoid receptor 1	CNR1
456091	Receptor tyrosine-protein kinase erbB-2	ERBB2
456100	Receptor tyrosine-protein kinase erbB-2	ERBB2
459247	Proto-oncogene tyrosine-protein kinase receptor Ret	RET
459280	Amine oxidase [flavin-containing] A	MAOA
459282	Cytochrome P450 11B2, mitochondrial	CYP11B2
459283	Cytochrome P450 11B1, mitochondrial	CYP11B1
459375	Sodium-dependent serotonin transporter	SLC6A4
459711	Sodium-dependent dopamine transporter	SLC6A3
459712	Sodium-dependent noradrenaline transporter	SLC6A2
459948	Sodium-dependent serotonin transporter	SLC6A4
459949	Sodium-dependent noradrenaline transporter	SLC6A2

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
461131	Cannabinoid receptor 1	CNR1
461809	Aromatase	CYP19A1
462782	Acetylcholinesterase	ACHE
464690	Cannabinoid receptor 1	CNR1
465032	Cholinesterase	BCHE
466821	Sodium-dependent noradrenaline transporter	SLC6A2
467823	Estradiol 17-beta-dehydrogenase 1	HSD17B1
469220	Glycogen phosphorylase, muscle form	PYGM
469633	Xanthine dehydrogenase/oxidase	XDH
470602	Cytochrome P450 3A4	CYP3A4
471116	Prostaglandin G/H synthase 1	PTGS1
471117	Prostaglandin G/H synthase 2	PTGS2
471239	Estrogen receptor	ESR1
471240	Estrogen receptor	ESR1
471630	Sodium-dependent serotonin transporter	SLC6A4
472619	Dihydrofolate reductase	DHFR
477863	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
479369	Aromatase	CYP19A1
479548	Prostaglandin G/H synthase 2	PTGS2
480068	Histone deacetylase 2	HDAC2
480448	Acetylcholinesterase	ACHE
480449	Cholinesterase	BCHE
482940	Histone deacetylase 1	HDAC1
483451	Sodium-dependent noradrenaline transporter	SLC6A2

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
483452	Sodium-dependent noradrenaline transporter	SLC6A2
484757	Hydroxycarboxylic acid receptor 2	HCAR2
484805	cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A	PDE10A
485980	Cholesteryl ester transfer protein	CETP
489474	Methionine aminopeptidase 1	METAP1
489476	Methionine aminopeptidase 2	METAP2
491169	Sodium-dependent dopamine transporter	SLC6A3
491170	Sodium-dependent noradrenaline transporter	SLC6A2
492320	Epidermal growth factor receptor	EGFR
493501	Androgen receptor	AR
493503	Androgen receptor	AR
494980	Progesterone receptor	PGR
497342	Amine oxidase [flavin-containing] A	MAOA
497575	Cysteinyl leukotriene receptor 1	CYSLTR1
498477	Phospholipase D1	PLD1
498478	Phospholipase D2	PLD2
498482	Phospholipase D2	PLD2
498484	Phospholipase D1	PLD1
499539	Steroid 17-alpha-hydroxylase/17,20 lyase	CYP17A1
501798	Cannabinoid receptor 1	CNR1
509549	Amine oxidase [flavin-containing] A	MAOA
509564	Prolyl endopeptidase FAP	FAP
511766	Acetylcholinesterase	ACHE

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
512126	Membrane primary amine oxidase	AOC3
514141	Amine oxidase [flavin-containing] A	MAOA
515517	Glucocorticoid receptor	NR3C1
538478	Acetylcholinesterase	ACHE
540067	Ribosyldihyronicotinamide dehydrogenase [quinone]	NQO2
550507	cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A	PDE10A
3550776	Sodium-dependent serotonin transporter	SLC6A4
550777	Sodium-dependent noradrenaline transporter	SLC6A2
550778	Sodium-dependent dopamine transporter	SLC6A3
553065	Cytochrome P450 11B1, mitochondrial	CYP11B1
553066	Cytochrome P450 11B2, mitochondrial	CYP11B2
568216	Glucocorticoid receptor	NR3C1
568218	Glucocorticoid receptor	NR3C1
570116	P2X purinoceptor 2	P2RX2
576612	Potassium voltage-gated channel subfamily H member 2	KCNH2
578229	Dipeptidyl peptidase 4	DPP4
578685	Xanthine dehydrogenase/oxidase	XDH
580195	Dipeptidyl peptidase 4	DPP4
589578	Vascular endothelial growth factor receptor 2	KDR
590252	Cannabinoid receptor 1	CNR1
591316	RAF proto-oncogene serine/threonine-protein kinase	RAF1
592803	Bifunctional epoxide hydrolase 2	EPHX2
593714	Solute carrier family 22 member 12	SLC22A12

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
594415	Prostaglandin G/H synthase 2	PTGS2
595800	Histamine H1 receptor	HRH1
598149	Muscarinic acetylcholine receptor M3	CHRM3
598150	Muscarinic acetylcholine receptor M2	CHRM2
598151	Muscarinic acetylcholine receptor M1	CHRM1
598348	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	PDE4B
600984	Cholinesterase	BCHE
602103	Aromatase	CYP19A1
605421	Dihydrofolate reductase	DHFR
611101	Prostaglandin D2 receptor	PTGDR
612501	Cannabinoid receptor 1	CNR1
614370	Xanthine dehydrogenase/oxidase	XDH
619830	Sodium-dependent noradrenaline transporter	SLC6A2
619831	Sodium-dependent noradrenaline transporter	SLC6A2
620239	Epidermal growth factor receptor	EGFR
620240	Receptor tyrosine-protein kinase erbB-2	ERBB2
620273	Arachidonate 5-lipoxygenase	ALOX5
621312	Fatty acid synthase	FASN
623021	Histone deacetylase 8	HDAC8
625146	Arachidonate 15-lipoxygenase	ALOX15
625150	Amine oxidase [flavin-containing] A	MAOA
625151	Muscarinic acetylcholine receptor M1	CHRM1
625152	Muscarinic acetylcholine receptor M2	CHRM2
625153	Muscarinic acetylcholine receptor M3	CHRM3

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
625154	Muscarinic acetylcholine receptor M4	CHRM4
625155	Muscarinic acetylcholine receptor M5	CHRM5
625161	Delta-type opioid receptor	OPRD1
625162	Kappa-type opioid receptor	OPRK1
625163	Mu-type opioid receptor	OPRM1
625171	Potassium voltage-gated channel subfamily H member 2	KCNH2
625172	Progesterone receptor	PGR
625181	Mitogen-activated protein kinase 1	MAPK1
625182	Mitogen-activated protein kinase 14	MAPK14
625184	Epidermal growth factor receptor	EGFR
625185	Tyrosine-protein kinase Fyn	FYN
625186	Receptor tyrosine-protein kinase erbB-2	ERBB2
625187	Tyrosine-protein kinase Lck	LCK
625191	5-hydroxytryptamine receptor 1B	HTR1B
625192	5-hydroxytryptamine receptor 2A	HTR2A
625193	Acetylcholinesterase	ACHE
625195	Adenosine receptor A2a	ADORA2A
625196	Adenosine receptor A3	ADORA3
625198	Alpha-1A adrenergic receptor	ADRA1A
625199	Alpha-1B adrenergic receptor	ADRA1B
625200	Alpha-1D adrenergic receptor	ADRA1D
625201	Alpha-2A adrenergic receptor	ADRA2A
625202	Alpha-2B adrenergic receptor	ADRA2B
625203	Alpha-2C adrenergic receptor	ADRA2C

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
625204	Beta-1 adrenergic receptor	ADRB1
625205	Beta-2 adrenergic receptor	ADRB2
625206	Beta-3 adrenergic receptor	ADRB3
625207	Sodium-dependent noradrenaline transporter	SLC6A2
625208	Aldose reductase	AKR1B1
625217	5-hydroxytryptamine receptor 2B	HTR2B
625218	5-hydroxytryptamine receptor 2C	HTR2C
625221	5-hydroxytryptamine receptor 6	HTR6
625222	Sodium-dependent serotonin transporter	SLC6A4
625223	Sigma non-opioid intracellular receptor 1	SIGMAR1
625227	Substance-K receptor	TACR2
625228	Androgen receptor	AR
625229	Thromboxane-A synthase	TBXAS1
625243	Prostaglandin G/H synthase 1	PTGS1
625244	Prostaglandin G/H synthase 2	PTGS2
625245	Cytochrome P450 1A2	CYP1A2
625247	Cytochrome P450 2C19	CYP2C19
625248	Cytochrome P450 2C9	CYP2C9
625249	Cytochrome P450 2D6	CYP2D6
625251	Cytochrome P450 3A4	CYP3A4
625252	D(1A) dopamine receptor	DRD1
625253	D(2) dopamine receptor	DRD2
625254	D(3) dopamine receptor	DRD3
625256	Sodium-dependent dopamine transporter	SLC6A3

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
625258	Estrogen receptor	ESR1
625263	Glucocorticoid receptor	NR3C1
625269	Histamine H1 receptor	HRH1
625270	Histamine H2 receptor	HRH2
626140	Glucocorticoid receptor	NR3C1
626141	Progesterone receptor	PGR
626142	Mineralocorticoid receptor	NR3C2
626146	Glucocorticoid receptor	NR3C1
629788	cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A	PDE10A
629915	Glucocorticoid receptor	NR3C1
631741	Muscarinic acetylcholine receptor M3	CHRM3
632046	Arachidonate 5-lipoxygenase-activating protein	ALOX5AP
632047	Arachidonate 5-lipoxygenase-activating protein	ALOX5AP
635462	Mineralocorticoid receptor	NR3C2
639076	Androgen receptor	AR
639078	Androgen receptor	AR
640508	Glucocorticoid receptor	NR3C1
641618	Prostaglandin G/H synthase 2	PTGS2
643321	Multidrug resistance protein 1	ABCB1
645197	Type-1 angiotensin II receptor	AGTR1
649305	Epidermal growth factor receptor	EGFR
649306	Receptor tyrosine-protein kinase erbB-2	ERBB2
Continued on next page		

Table S2 – continued from previous page

ID	protein target	gene
650201	cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A	PDE10A
650842	Aromatase	CYP19A1
652428	Cytochrome P450 11B1, mitochondrial	CYP11B1
652429	Cytochrome P450 11B2, mitochondrial	CYP11B2
658648	Prostaglandin G/H synthase 2	PTGS2
659183	Prolyl endopeptidase FAP	FAP
661902	Arachidonate 5-lipoxygenase	ALOX5
661903	Arachidonate 5-lipoxygenase	ALOX5
666573	Acetylcholinesterase	ACHE
670012	Histone deacetylase 6	HDAC6
670013	Histone deacetylase 1	HDAC1
672956	Calmodulin-domain protein kinase 1	CDPK1
677013	Acetylcholinesterase	ACHE
685045	Dipeptidyl peptidase 4	DPP4
688689	Dipeptidyl peptidase 4	DPP4
694143	Androgen receptor	AR
698652	Arachidonate 5-lipoxygenase	ALOX5
700026	Sodium channel protein type 5 subunit alpha	SCN5A
700027	Sodium channel protein type 9 subunit alpha	SCN9A
700028	Sodium channel protein type 9 subunit alpha	SCN9A
701329	Cholinesterase	BCHE
701330	Acetylcholinesterase	ACHE
701506	Fatty-acid amide hydrolase 1	FAAH

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
703400	Aldo-keto reductase family 1 member C3	AKR1C3
709142	P2X purinoceptor 4	P2RX4
709834	cGMP-specific 3',5'-cyclic phosphodiesterase	PDE5A
711102	Sucrase-isomaltase, intestinal	SI
711104	Lysosomal alpha-glucosidase	GAA
712520	Sodium-dependent serotonin transporter	SLC6A4
721751	Solute carrier family 22 member 2	SLC22A2
721752	Multidrug and toxin extrusion protein 2	SLC47A2
721754	Multidrug and toxin extrusion protein 1	SLC47A1
722207	Epidermal growth factor receptor	EGFR
722623	Sodium- and chloride-dependent GABA transporter 1	SLC6A1
723193	Cytochrome P450 11B2, mitochondrial	CYP11B2
723194	Cytochrome P450 11B1, mitochondrial	CYP11B1
723195	Aromatase	CYP19A1
725964	Receptor tyrosine-protein kinase erbB-2	ERBB2
725966	Epidermal growth factor receptor	EGFR
727074	D(2) dopamine receptor	DRD2
727393	Nuclear receptor ROR-gamma	RORC
728035	Xanthine dehydrogenase/oxidase	XDH
728046	Type-1 angiotensin II receptor	AGTR1
728454	Potassium voltage-gated channel subfamily H member 2	KCNH2
729258	Aldo-keto reductase family 1 member C3	AKR1C3
729259	Aldo-keto reductase family 1 member C2	AKR1C2
729339	Amine oxidase [flavin-containing] A	MAOA

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
731054	Sodium-dependent serotonin transporter	SLC6A4
731055	Sodium-dependent noradrenaline transporter	SLC6A2
732855	Mast/stem cell growth factor receptor Kit	KIT
732858	Epithelial discoidin domain-containing receptor 1	DDR1
734521	Prostaglandin G/H synthase 2	PTGS2
734522	Prostaglandin G/H synthase 2	PTGS2
734754	Proteasome subunit beta type-5	PSMB5
734801	Farnesyl pyrophosphate synthase	FDPS
734803	Farnesyl pyrophosphate synthase	FDPS
736008	Bifunctional epoxide hydrolase 2	EPHX2
736608	Aromatase	CYP19A1
738129	Myeloperoxidase	MPO
739023	Substance-P receptor	TACR1
739861	Proliferating cell nuclear antigen	PCNA
747252	Histone deacetylase 1	HDAC1
747610	Dihydrofolate reductase	-
747611	Dihydrofolate reductase	DHFR
747614	Dihydrofolate reductase	DHFR
749115	Sodium- and chloride-dependent GABA transporter 1	SLC6A1
749367	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	PDE4B
749667	Aurora kinase A	AURKA
749668	Vascular endothelial growth factor receptor 2	KDR
749669	Receptor-type tyrosine-protein kinase FLT3	FLT3
749871	Delta-type opioid receptor	OPRD1

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
750737	G-protein coupled receptor 55	GPR55
751624	Amine oxidase [flavin-containing] A	MAOA
755442	Voltage-dependent L-type calcium channel subunit alpha-1C	CACH2C
755443	Voltage-dependent L-type calcium channel subunit alpha-1D	CACNA1D
759595	cAMP-specific 3',5'-cyclic phosphodiesterase 4A	PDE4A
761384	Mineralocorticoid receptor	NR3C2
761443	Carbonic anhydrase 9	CA9
761444	Carbonic anhydrase 2	CA2
764602	Sodium- and chloride-dependent GABA transporter 1	SLC6A1
765521	Cytochrome P450 11B2, mitochondrial	CYP11B2
765522	Cytochrome P450 11B1, mitochondrial	CYP11B1
767289	Amine oxidase [flavin-containing] A	MAOA
767814	cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A	PDE10A
771350	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	PDE4B
773015	Mast/stem cell growth factor receptor Kit	KIT
773016	Platelet-derived growth factor receptor beta	PDGFRB
773017	Vascular endothelial growth factor receptor 2	KDR
780495	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	PDE4B
780586	Histone deacetylase 6	HDAC6
780587	Histone deacetylase 3	HDAC3
780588	Histone deacetylase 1	HDAC1
1054736	Muscarinic acetylcholine receptor M3	CHRM3
1055728	Urease	-

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
1057357	Prolyl endopeptidase FAP	FAP
1057358	Dipeptidyl peptidase 9	DPP9
1057359	Dipeptidyl peptidase 8	DPP8
1057360	Dipeptidyl peptidase 2	DPP7
1057361	Dipeptidyl peptidase 4	DPP4
1057565	Prostaglandin E synthase	PTGES
1057566	Arachidonate 5-lipoxygenase	ALOX5
1057934	Epidermal growth factor receptor	EGFR
1057953	ATP-binding cassette sub-family G member 2	ABCG2
1058575	Mineralocorticoid receptor	NR3C2
1058577	Glucocorticoid receptor	NR3C1
1061153	Vascular endothelial growth factor receptor 2	KDR
1064827	ATP-dependent 6-phosphofructokinase	PKF
1065631	Histone deacetylase 4	HDAC4
1065638	Histone deacetylase 4	HDAC4
1067106	5-hydroxytryptamine receptor 6	HTR6
1068575	Botulinum neurotoxin type A	BOTA
1069013	Cytochrome P450 3A4	CYP3A4
1069523	Prostaglandin G/H synthase 1	PTGS1
1072705	Pantothenate synthetase	PANC
1076089	Fructose-1,6-bisphosphatase 1	FBP1
1077092	Acetylcholinesterase	ACHE
1085881	Voltage-dependent L-type calcium channel subunit alpha-1C	CACNA1C
1094446	Voltage-dependent L-type calcium channel subunit alpha-1C	CACNA1C

Continued on next page

Table S2 – continued from previous page

ID	protein target	gene
1105813	4-aminobutyrate aminotransferase, mitochondrial	ABAT
1106449	Epidermal growth factor receptor	EGFR
1107227	Prostaglandin G/H synthase 2	PTGS2
1113337	Acetylcholinesterase	ACHE
1113849	Acetylcholinesterase	ACHE

The column corresponding to “ID” has the bioassay AIDs from PubChem. The column corresponding to “protein target” has protein target of the corresponding bioassay. The column corresponding to “gene” has the encoding gene on the protein target. “-” in column “gene” indicates no gene information available.

S5 Best Performance Analysis

Table S3 presents the best performance for each of the 665 bioassays in the dataset.

Table S3: Overall Performance Comparison

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
619831	15	0.267	0.533	100.00	$bSim_p^{al}$	$alinS_{cSim}^{rpos}$	2	$cSim^{max}$	13
349306	21	0.420	0.680	61.90	$bSim_x^{al}$	$alinS_{cSim}^{rpos}$	6	$cSim^{pos}$	35
261405	16	0.533	0.800	50.00	$bSim_p^{cs}$	-	9	$cSim^{avg}$	43
625258	20	0.400	0.600	50.00	$bSim_x^{al}$	$alinS_{cSim}^{rpos}$	4	$cSim^{pos}$	28
55843	22	0.460	0.680	47.83	$bSim_x^{al}$	$alinS_{cSim}^{rpos}$	4	$cSim^{pos}$	37
240749	23	0.453	0.660	45.59	$bSim_x^{al}$	$alinS_{cSim}^{rpos}$	3	$cSim^{avg}$	34
749667	22	0.467	0.667	42.86	$bSim_x^{al}$	$alinS_{cSim}^{rpos}$	9	$cSim^{max}$	38
149865	22	0.533	0.747	40.00	$bSim_x^{ci}$	-	6	$cSim^{max}$	29
498478	21	0.340	0.473	39.21	$bSim_p^{cs}$	-	10	$cSim^{max}$	26
264807	18	0.600	0.833	38.89	$bSim_p^{cs}$	-	3	$cSim^{avg}$	24

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
459949	26	0.367	0.507	38.18	bSim _p ^{cs}	-	6	cSim ^{pos}	26
477863	19	0.267	0.367	37.50	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	14
162143	18	0.367	0.500	36.36	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{avg}	35
723193	24	0.480	0.647	34.72	bSim _x ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{min}	38
274062	22	0.653	0.867	32.65	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{max}	42
241231	21	0.633	0.840	32.63	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	28
626142	26	0.513	0.680	32.47	bSim _p ^{al}	alinS _{csim}	7	cSim ^{min}	18
389657	26	0.587	0.773	31.82	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{max}	35
312172	23	0.407	0.533	31.15	bSim _x ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	23
761444	19	0.433	0.567	30.77	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	24
701506	20	0.467	0.600	28.57	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{max}	23
1094446	24	0.440	0.560	27.27	bSim _x ^{al}	alinS _{csim} ^{rpos}	12	cSim ^{max}	50
260896	22	0.553	0.700	26.51	bSim _x ^{ci}	-	4	cSim ^{max}	32
319592	20	0.633	0.800	26.32	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{pos}	44
255080	20	0.633	0.800	26.31	bSim _p ^{al}	alinS _{csim} ^{rpos}	3	cSim ^{max}	30
773016	25	0.540	0.680	25.93	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{avg}	25
241201	21	0.707	0.887	25.47	bSim _x ^{al}	alinS _{csim}	6	cSim ^{max}	38
491170	23	0.447	0.560	25.37	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	13
258960	23	0.527	0.660	25.32	bSim _p ^{cs}	-	5	cSim ^{avg}	24
417910	21	0.607	0.760	25.27	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{pos}	27
37667	17	0.400	0.500	25.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{avg}	29
160735	19	0.667	0.833	25.00	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{min}	4
625150	23	0.533	0.667	25.00	bSim _x ^{al}	alinS _{csim}	6	cSim ^{max}	41
47369	22	0.353	0.440	24.53	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	35

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
721752	21	0.553	0.687	24.10	bSim _p ^{cs}	-	12	cSim ^{min}	27
489476	23	0.453	0.560	23.53	bSim _x ^{al}	alinS _{cidn} ^{rpos}	9	cSim ^{max}	24
352215	24	0.400	0.493	23.33	bSim _p ^{al}	alinS _{csim}	4	cSim ^{max}	27
197943	32	0.442	0.545	23.27	bSim _x ^{al}	alinS _{cidn}	3	cSim ^{min}	9
143393	24	0.547	0.673	23.17	bSim _x ^{ci}	-	6	cSim ^{max}	31
441347	32	0.398	0.489	22.97	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{avg}	4
749668	24	0.493	0.607	22.97	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	21
625245	24	0.553	0.680	22.89	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{min}	35
626140	24	0.587	0.720	22.73	bSim _x ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{avg}	30
480068	16	0.600	0.733	22.22	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
652429	31	0.678	0.829	22.19	bSim _x ^{al}	alinS _{csim}	11	cSim ^{avg}	36
265578	24	0.560	0.680	21.43	bSim _x ^{al}	alinS _{cidn} ^{rpos}	10	cSim ^{max}	33
431881	19	0.467	0.567	21.43	bSim _x ^{al}	alinS _{csim} ^{rpos}	3	cSim ^{max}	34
454488	20	0.467	0.567	21.43	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{pos}	18
432545	22	0.567	0.687	21.18	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{min}	35
625187	19	0.633	0.767	21.05	bSim _x ^{ci}	-	8	cSim ^{min}	28
262046	24	0.580	0.700	20.69	bSim _p ^{cs}	-	5	cSim ^{pos}	32
1057359	30	0.520	0.627	20.51	bSim _x ^{al}	alinS _{csim}	5	cSim ^{max}	40
340414	24	0.700	0.840	20.00	bSim _x ^{ci}	-	6	cSim ^{min}	26
367273	15	0.667	0.800	20.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{avg}	28
641618	22	0.567	0.680	20.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{max}	40
498477	25	0.520	0.620	19.23	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{max}	36
353291	22	0.487	0.580	19.18	bSim _p ^{al}	alinS _{csim} ^{rpos}	3	cSim ^{max}	23
491169	21	0.667	0.793	19.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{avg}	31

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
273168	23	0.600	0.713	18.89	bSim _p ^{cs}	-	7	cSim ^{max}	42
625249	38	0.505	0.600	18.87	bSim _p ^{al}	alinS _{cSim}	6	cSim ^{min}	32
90687	20	0.533	0.633	18.75	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{max}	5
294499	25	0.540	0.640	18.52	bSim _x ^{al}	alinS _{cSim} ^{rpos}	3	cSim ^{min}	19
625181	21	0.580	0.687	18.39	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{min}	29
734522	36	0.521	0.617	18.27	bSim _x ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{pos}	41
538478	22	0.733	0.867	18.18	bSim _x ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{min}	34
90682	22	0.593	0.700	17.98	bSim _x ^{ci}	-	6	cSim ^{avg}	37
722207	22	0.413	0.487	17.74	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	22
773017	23	0.640	0.753	17.71	bSim _p ^{cs}	-	8	cSim ^{pos}	31
732855	31	0.678	0.798	17.70	bSim _p ^{al}	alinS _{cSim}	12	cSim ^{min}	27
367003	17	0.567	0.667	17.65	bSim _p ^{cs}	-	9	cSim ^{max}	31
751624	20	0.567	0.667	17.65	bSim _p ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{avg}	23
625186	24	0.613	0.720	17.39	bSim _x ^{ci}	-	4	cSim ^{max}	21
395483	23	0.733	0.860	17.27	bSim _x ^{ci}	-	5	cSim ^{min}	30
659183	29	0.580	0.680	17.24	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{min}	31
362063	23	0.587	0.687	17.05	bSim _x ^{ci}	-	5	cSim ^{min}	33
56330	22	0.553	0.647	16.87	bSim _x ^{al}	alinS _{cSim} ^{rpos}	4	cSim ^{avg}	28
39648	21	0.633	0.740	16.84	bSim _x ^{al}	alinS _{cSim}	6	cSim ^{max}	38
639078	22	0.673	0.787	16.83	bSim _x ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{pos}	31
240835	19	0.800	0.933	16.67	bSim _x ^{al}	alinS _{cSim} ^{rpos}	3	cSim ^{max}	39
595800	20	0.600	0.700	16.67	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	13
605421	20	0.600	0.700	16.67	bSim _x ^{ci}	-	3	cSim ^{max}	13
780495	19	0.600	0.700	16.67	bSim _x ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{avg}	28

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
625227	41	0.489	0.570	16.56	bSim _p ^{al}	alinS _{csim}	9	cSim ^{max}	29
395658	23	0.693	0.807	16.35	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
264808	21	0.653	0.760	16.33	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{avg}	21
621312	21	0.613	0.713	16.30	bSim _x ^{ci}	-	4	cSim ^{min}	34
625161	22	0.613	0.713	16.30	bSim _x ^{ci}	-	6	cSim ^{min}	27
56178	24	0.700	0.813	16.19	bSim _x ^{ci}	-	4	cSim ^{avg}	37
727074	23	0.707	0.820	16.04	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	22
732858	33	0.512	0.594	15.99	bSim _p ^{cs}	-	10	cSim ^{pos}	45
241574	19	0.633	0.733	15.79	bSim _x ^{ci}	-	6	cSim ^{min}	31
269937	18	0.633	0.733	15.79	bSim _p ^{cs}	-	11	cSim ^{max}	25
471240	30	0.507	0.587	15.79	bSim _x ^{al}	alinS _{csim}	6	cSim ^{avg}	34
353290	22	0.600	0.693	15.55	bSim _x ^{ci}	-	7	cSim ^{max}	24
180514	21	0.747	0.860	15.18	bSim _p ^{cs}	-	13	cSim ^{avg}	31
414980	18	0.667	0.767	15.00	bSim _x ^{ci}	-	4	cSim ^{max}	33
469633	27	0.667	0.767	15.00	bSim _x ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{min}	37
494980	19	0.667	0.767	15.00	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{min}	6
620273	16	0.667	0.767	15.00	bSim _p ^{cs}	-	6	cSim ^{avg}	35
666573	20	0.667	0.767	15.00	bSim _x ^{al}	alinS _{csim}	3	cSim ^{avg}	32
1061153	19	0.667	0.767	15.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{pos}	32
345893	33	0.638	0.733	14.92	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{min}	30
454554	38	0.562	0.645	14.83	bSim _x ^{al}	alinS _{csim}	4	cSim ^{avg}	51
309702	22	0.453	0.520	14.71	bSim _p ^{cs}	-	3	cSim ^{avg}	35
268271	25	0.420	0.480	14.29	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	32
739023	20	0.700	0.800	14.29	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	24

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
1058577	19	0.700	0.800	14.29	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	32
493503	27	0.613	0.700	14.13	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{pos}	26
198086	45	0.600	0.683	13.89	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{avg}	30
55844	23	0.627	0.713	13.83	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{pos}	20
224348	21	0.627	0.713	13.83	bSim _p ^{cs}	-	10	cSim ^{max}	42
56310	24	0.580	0.660	13.79	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	23
267771	26	0.680	0.773	13.72	bSim _p ^{cs}	-	7	cSim ^{min}	32
632047	23	0.633	0.720	13.68	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	24
257779	23	0.587	0.667	13.64	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	32
414717	19	0.733	0.833	13.64	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{max}	33
423861	26	0.587	0.667	13.64	bSim _x ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{pos}	30
459375	19	0.733	0.833	13.64	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{avg}	29
625208	20	0.733	0.833	13.64	bSim _p ^{al}	alinS _{csim}	9	cSim ^{pos}	23
728035	20	0.733	0.833	13.64	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	21
348792	29	0.687	0.780	13.59	bSim _x ^{al}	alinS _{csim}	5	cSim ^{max}	29
723194	28	0.593	0.673	13.48	bSim _x ^{al}	alinS _{csim}	9	cSim ^{max}	33
359909	26	0.693	0.787	13.46	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	22
600984	38	0.567	0.643	13.45	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{max}	12
290288	28	0.647	0.733	13.40	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{max}	32
484757	28	0.600	0.680	13.33	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	13
444522	22	0.653	0.740	13.27	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	33
57074	23	0.507	0.573	13.16	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	22
321309	26	0.660	0.747	13.13	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{pos}	7
725964	31	0.653	0.739	13.12	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{pos}	43

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
711104	22	0.407	0.460	13.11	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{max}	14
650201	26	0.560	0.633	13.09	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	23
728046	23	0.713	0.807	13.09	bSim _x ^{ci}	-	7	cSim ^{min}	29
625248	20	0.767	0.867	13.04	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{avg}	29
780588	24	0.613	0.693	13.04	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	22
568218	22	0.667	0.753	13.00	bSim _x ^{al}	alinS _{csim}	7	cSim ^{max}	38
416955	21	0.773	0.873	12.93	bSim _p ^{al}	alinS _{csim} ^{rpos}	2	cSim ^{min}	30
465032	22	0.773	0.873	12.93	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{avg}	7
598149	26	0.673	0.760	12.87	bSim _p ^{cs}	-	11	cSim ^{avg}	25
352217	23	0.627	0.707	12.77	bSim _p ^{cs}	-	11	cSim ^{pos}	31
626141	24	0.627	0.707	12.77	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{avg}	29
1077092	28	0.627	0.707	12.77	bSim _p ^{cs}	-	4	cSim ^{pos}	28
198410	35	0.600	0.676	12.70	bSim _p ^{al}	alinS _{csim}	4	cSim ^{min}	26
35397	20	0.533	0.600	12.50	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
56169	20	0.533	0.600	12.50	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
241604	20	0.800	0.900	12.50	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{min}	27
269938	18	0.533	0.600	12.50	bSim _p ^{cs}	-	12	cSim ^{max}	30
276317	25	0.640	0.720	12.50	bSim _x ^{ci}	-	4	cSim ^{pos}	24
410413	16	0.533	0.600	12.50	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	20
731055	17	0.533	0.600	12.50	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{min}	31
734803	25	0.640	0.720	12.50	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{avg}	28
453456	28	0.647	0.727	12.37	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{max}	40
670012	32	0.651	0.731	12.28	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{min}	28
469220	28	0.600	0.673	12.22	bSim _p ^{cs}	-	8	cSim ^{avg}	25

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
480448	23	0.660	0.740	12.12	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{max}	13
1057357	26	0.660	0.740	12.12	bSim _p ^{al}	alinS _{csim}	7	cSim ^{avg}	17
313112	19	0.833	0.933	12.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	22
263981	23	0.613	0.687	11.96	bSim _p ^{cs}	-	10	cSim ^{pos}	20
339393	30	0.560	0.627	11.91	bSim _p ^{al}	alinS _{csim}	6	cSim ^{min}	30
464690	22	0.560	0.627	11.91	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{min}	17
198247	33	0.545	0.610	11.89	bSim _x ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	32
625184	38	0.545	0.610	11.79	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	32
241692	20	0.567	0.633	11.76	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	27
364887	25	0.680	0.760	11.76	bSim _p ^{al}	alinS _{csim}	5	cSim ^{max}	23
492320	20	0.567	0.633	11.76	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{max}	13
661902	31	0.731	0.817	11.72	bSim _p ^{cs}	-	13	cSim ^{max}	32
241793	26	0.573	0.640	11.63	bSim _x ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	32
362471	30	0.693	0.773	11.54	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{max}	30
411890	27	0.760	0.847	11.40	bSim _x ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{avg}	39
322229	28	0.647	0.720	11.34	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{pos}	38
240624	19	0.600	0.667	11.11	bSim _x ^{ci}	-	6	cSim ^{avg}	31
241865	26	0.720	0.800	11.11	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	24
258959	28	0.600	0.667	11.11	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{pos}	5
273141	23	0.540	0.600	11.11	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
761443	20	0.600	0.667	11.11	bSim _p ^{cs}	-	8	cSim ^{min}	31
512126	38	0.691	0.767	11.03	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{min}	29
1067106	32	0.606	0.672	11.01	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{max}	34
30868	33	0.598	0.663	10.83	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{max}	4

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
330796	28	0.620	0.687	10.75	bSim _p ^{cs}	-	6	cSim ^{avg}	40
241825	18	0.633	0.700	10.53	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	27
297967	18	0.633	0.700	10.53	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{max}	20
304252	19	0.633	0.700	10.53	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	27
1106449	23	0.700	0.773	10.48	bSim _x ^{al}	alinS _{cidn}	3	cSim ^{avg}	8
273166	26	0.573	0.633	10.47	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{min}	7
625229	37	0.640	0.707	10.41	bSim _p ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{min}	25
241195	23	0.833	0.920	10.40	bSim _x ^{ci}	-	5	cSim ^{avg}	36
268714	25	0.580	0.640	10.34	bSim _x ^{ci}	-	1	cSim ^{pos}	11
1058575	21	0.773	0.853	10.34	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	31
162176	23	0.647	0.713	10.31	bSim _x ^{ci}	-	3	cSim ^{avg}	27
462782	22	0.713	0.787	10.28	bSim _x ^{al}	alinS _{csim}	7	cSim ^{max}	25
625191	38	0.698	0.769	10.24	bSim _p ^{al}	alinS _{csim}	5	cSim ^{avg}	34
640508	26	0.587	0.647	10.23	bSim _p ^{al}	alinS _{csim}	4	cSim ^{min}	29
290289	30	0.653	0.720	10.20	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	19
765522	30	0.653	0.720	10.20	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	37
1057360	23	0.527	0.580	10.13	bSim _x ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{pos}	18
498484	24	0.593	0.653	10.11	bSim _p ^{cs}	-	12	cSim ^{avg}	30
499539	22	0.660	0.727	10.10	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{min}	46
242068	20	0.667	0.733	10.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	21
260895	20	0.667	0.733	10.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{pos}	21
364670	18	0.667	0.733	10.00	bSim _p ^{al}	alinS _{csim}	4	cSim ^{avg}	28
453457	19	0.667	0.733	10.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{min}	31
721751	26	0.600	0.660	10.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{pos}	22

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
197775	34	0.516	0.568	9.96	bSim _x ^{al}	alinS _{cSim}	3	cSim ^{pos}	34
734521	39	0.552	0.607	9.91	bSim _x ^{al}	alinS _{cSim}	8	cSim ^{min}	26
364560	28	0.673	0.740	9.90	bSim _p ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{max}	38
274396	41	0.581	0.638	9.84	bSim _x ^{al}	alinS _{cSim}	7	cSim ^{min}	39
55855	24	0.820	0.900	9.76	bSim _x ^{ci}	-	4	cSim ^{max}	31
635462	29	0.627	0.687	9.57	bSim _p ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{pos}	30
37646	30	0.560	0.613	9.52	bSim _p ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{avg}	27
341470	17	0.700	0.767	9.52	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
353308	20	0.700	0.767	9.52	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{pos}	14
570116	17	0.700	0.767	9.52	bSim _p ^{cs}	-	9	cSim ^{max}	36
736608	20	0.700	0.767	9.52	bSim _p ^{cs}	-	7	cSim ^{min}	34
1057566	18	0.700	0.767	9.52	bSim _p ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{min}	22
298279	31	0.583	0.638	9.48	bSim _p ^{al}	alinS _{cSim} ^{rpos}	7	cSim ^{avg}	27
589578	22	0.633	0.693	9.47	bSim _x ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{avg}	29
266160	25	0.640	0.700	9.38	bSim _p ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{min}	26
37659	49	0.608	0.664	9.32	bSim _x ^{ci}	-	8	cSim ^{min}	38
242151	64	0.596	0.652	9.30	bSim _p ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{avg}	37
260152	24	0.647	0.707	9.28	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
263980	24	0.647	0.707	9.28	bSim _p ^{cs}	-	9	cSim ^{avg}	22
625195	23	0.647	0.707	9.28	bSim _x ^{al}	alinS _{cSim}	6	cSim ^{min}	31
149742	20	0.367	0.400	9.09	bSim _p ^{al}	alinS _{cSim} ^{rpos}	4	cSim ^{avg}	26
242043	20	0.733	0.800	9.09	bSim _p ^{al}	alinS _{cSim} ^{rpos}	7	cSim ^{avg}	25
454342	25	0.660	0.720	9.09	bSim _x ^{al}	alinS _{cSim}	5	cSim ^{min}	34
550777	30	0.587	0.640	9.09	bSim _p ^{cs}	-	7	cSim ^{avg}	23

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
594415	16	0.733	0.800	9.09	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{avg}	24
267731	31	0.630	0.688	9.06	bSim _x ^{ci}	-	4	cSim ^{max}	26
625251	22	0.813	0.887	9.02	bSim _x ^{al}	alinS _{cidn} ^{rpos}	9	cSim ^{avg}	31
729259	33	0.592	0.646	9.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	2	cSim ^{pos}	27
471116	34	0.680	0.741	8.96	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	30
78155	39	0.745	0.812	8.94	bSim _p ^{cs}	-	8	cSim ^{min}	39
316695	27	0.747	0.813	8.93	bSim _p ^{cs}	-	6	cSim ^{max}	33
371781	25	0.680	0.740	8.82	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	29
625205	25	0.680	0.740	8.82	bSim _x ^{ci}	-	4	cSim ^{avg}	17
274110	26	0.613	0.667	8.70	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{min}	27
550778	30	0.613	0.667	8.70	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	20
307629	19	0.767	0.833	8.69	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	2
731054	28	0.620	0.673	8.60	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{max}	12
314412	36	0.638	0.693	8.58	bSim _p ^{cs}	-	7	cSim ^{pos}	27
242065	28	0.700	0.760	8.57	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	35
631741	26	0.700	0.760	8.57	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{pos}	23
441648	26	0.627	0.680	8.51	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{avg}	35
652428	39	0.588	0.638	8.50	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{pos}	24
393050	34	0.676	0.733	8.45	bSim _p ^{cs}	-	3	cSim ^{pos}	36
448555	21	0.793	0.860	8.40	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	14
198094	16	0.800	0.867	8.33	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	5
198409	33	0.731	0.792	8.33	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	5
317007	15	0.800	0.867	8.33	bSim _p ^{cs}	-	4	cSim ^{min}	29
429130	15	0.800	0.867	8.33	bSim _p ^{cs}	-	10	cSim ^{min}	34

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
612501	29	0.807	0.873	8.26	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{min}	8
650842	21	0.807	0.873	8.26	bSim _x ^{ci}	-	6	cSim ^{avg}	40
625163	35	0.695	0.752	8.22	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{max}	30
161281	45	0.611	0.661	8.18	bSim _p ^{al}	alinS _{csim}	5	cSim ^{max}	35
420668	30	0.653	0.707	8.16	bSim _p ^{al}	alinS _{csim}	5	cSim ^{min}	37
441649	27	0.653	0.707	8.16	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	25
509564	22	0.653	0.707	8.16	bSim _x ^{ci}	-	5	cSim ^{avg}	31
480449	25	0.740	0.800	8.11	bSim _x ^{al}	alinS _{csim}	6	cSim ^{max}	33
598150	35	0.705	0.762	8.11	bSim _p ^{cs}	-	10	cSim ^{min}	23
700026	28	0.580	0.627	8.05	bSim _x ^{ci}	-	5	cSim ^{min}	27
307630	20	0.833	0.900	8.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{min}	37
317008	18	0.833	0.900	8.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	40
643321	30	0.667	0.720	8.00	bSim _p ^{cs}	-	7	cSim ^{max}	35
314091	28	0.753	0.813	7.96	bSim _p ^{cs}	-	12	cSim ^{min}	28
343580	22	0.673	0.727	7.92	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{avg}	19
386535	22	0.673	0.727	7.92	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
441348	28	0.673	0.727	7.92	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{pos}	14
709142	31	0.634	0.684	7.81	bSim _p ^{cs}	-	12	cSim ^{max}	33
540067	29	0.687	0.740	7.77	bSim _x ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{min}	28
326368	26	0.773	0.833	7.76	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	35
270616	20	0.867	0.933	7.69	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	37
276767	17	0.867	0.933	7.69	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{avg}	38
319764	30	0.693	0.747	7.69	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	31
361789	28	0.607	0.653	7.69	bSim _x ^{al}	alinS _{csim}	6	cSim ^{min}	27

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
629915	19	0.867	0.933	7.69	bSim _x ^{ci}	-	5	cSim ^{min}	25
698652	22	0.700	0.753	7.62	bSim _p ^{cs}	-	14	cSim ^{pos}	42
639076	23	0.613	0.660	7.61	bSim _p ^{al}	alinS _{cSim} ^{rpos}	2	cSim ^{avg}	27
391371	36	0.564	0.607	7.59	bSim _p ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{pos}	30
267730	32	0.653	0.703	7.58	bSim _x ^{al}	alinS _{cSim} ^{rpos}	2	cSim ^{max}	15
354245	39	0.698	0.750	7.51	bSim _x ^{ci}	-	7	cSim ^{min}	27
343841	25	0.800	0.860	7.50	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
688689	45	0.594	0.639	7.48	bSim _p ^{al}	alinS _{cSim} ^{rpos}	2	cSim ^{min}	34
1057361	26	0.627	0.673	7.45	bSim _x ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{max}	33
384316	38	0.707	0.759	7.41	bSim _x ^{al}	alinS _{cSim} ^{rpos}	8	cSim ^{min}	42
700027	31	0.617	0.663	7.41	bSim _p ^{al}	alinS _{cSim} ^{rpos}	4	cSim ^{max}	32
747614	27	0.720	0.773	7.41	bSim _p ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{min}	25
750737	29	0.633	0.680	7.37	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	23
721754	38	0.550	0.591	7.36	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{pos}	50
294498	21	0.733	0.787	7.27	bSim _p ^{cs}	-	11	cSim ^{avg}	38
419199	24	0.733	0.787	7.27	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
598151	23	0.553	0.593	7.23	bSim _x ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{avg}	29
493501	28	0.740	0.793	7.21	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{pos}	11
780587	23	0.740	0.793	7.21	bSim _x ^{al}	alinS _{cSim} ^{rpos}	8	cSim ^{min}	29
738129	24	0.840	0.900	7.14	bSim _x ^{ci}	-	9	cSim ^{avg}	47
712520	36	0.605	0.648	7.09	bSim _p ^{cs}	-	14	cSim ^{pos}	23
304223	21	0.753	0.807	7.08	bSim _p ^{cs}	-	5	cSim ^{min}	31
56345	23	0.660	0.707	7.07	bSim _p ^{al}	alinS _{cSim} ^{rpos}	8	cSim ^{avg}	16
471630	40	0.607	0.650	7.06	bSim _x ^{ci}	-	2	cSim ^{max}	26

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
619830	21	0.567	0.607	7.06	bSim _p ^{al}	alinS _{csim} ^{rpos}	2	cSim ^{avg}	13
441697	32	0.649	0.695	7.04	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{avg}	27
410405	23	0.853	0.913	7.03	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	13
672956	31	0.623	0.667	7.03	bSim _x ^{ci}	-	6	cSim ^{min}	31
293101	38	0.712	0.762	7.02	bSim _p ^{al}	alinS _{csim}	6	cSim ^{max}	33
436918	31	0.815	0.872	7.01	bSim _p ^{al}	alinS _{csim} ^{rpos}	12	cSim ^{max}	22
483452	22	0.667	0.713	7.00	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	3
362026	30	0.573	0.613	6.98	bSim _p ^{al}	alinS _{csim}	5	cSim ^{pos}	17
755443	34	0.600	0.642	6.98	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{min}	29
591316	28	0.767	0.820	6.96	bSim _x ^{al}	alinS _{csim}	8	cSim ^{avg}	37
497342	27	0.480	0.513	6.94	bSim _p ^{cs}	-	11	cSim ^{min}	41
1069013	26	0.773	0.827	6.90	bSim _x ^{ci}	-	7	cSim ^{avg}	28
241337	20	0.500	0.533	6.67	bSim _x ^{al}	alinS _{csim}	3	cSim ^{avg}	29
259059	26	0.600	0.640	6.67	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{min}	8
304251	24	0.500	0.533	6.67	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{avg}	6
625162	29	0.607	0.647	6.59	bSim _p ^{al}	alinS _{csim}	9	cSim ^{avg}	29
625206	29	0.713	0.760	6.54	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{min}	31
755442	30	0.613	0.653	6.52	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	29
393561	25	0.620	0.660	6.45	bSim _x ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{min}	31
626146	25	0.620	0.660	6.45	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{avg}	33
1068575	40	0.664	0.707	6.45	bSim _x ^{al}	alinS _{cidn} ^{rpos}	9	cSim ^{min}	36
39807	27	0.727	0.773	6.42	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{avg}	28
362929	41	0.759	0.807	6.38	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{max}	44
515517	24	0.627	0.667	6.38	bSim _x ^{al}	alinS _{csim}	9	cSim ^{pos}	39

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
645197	28	0.627	0.667	6.38	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	23
625146	43	0.620	0.659	6.27	bSim _p ^{al}	alinS _{csim}	5	cSim ^{avg}	33
55820	20	0.533	0.567	6.25	bSim _p ^{al}	alinS _{csim} ^{rpos}	3	cSim ^{avg}	23
269691	20	0.533	0.567	6.25	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{max}	30
459712	28	0.647	0.687	6.19	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{max}	34
602103	28	0.647	0.687	6.19	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	35
269184	33	0.623	0.661	6.12	bSim _x ^{ci}	-	6	cSim ^{max}	33
197944	32	0.596	0.632	6.07	bSim _x ^{ci}	-	8	cSim ^{avg}	37
747611	25	0.660	0.700	6.06	bSim _p ^{al}	alinS _{csim} ^{rpos}	11	cSim ^{min}	11
255079	29	0.667	0.707	6.00	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{max}	9
362062	21	0.560	0.593	5.95	bSim _x ^{al}	alinS _{csim}	9	cSim ^{avg}	37
773015	23	0.673	0.713	5.94	bSim _x ^{al}	alinS _{csim}	5	cSim ^{avg}	30
91425	22	0.680	0.720	5.88	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{max}	5
197932	18	0.567	0.600	5.88	bSim _x ^{al}	alinS _{csim} ^{rpos}	11	cSim ^{avg}	42
590252	25	0.680	0.720	5.88	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{max}	5
339521	34	0.722	0.764	5.81	bSim _x ^{al}	alinS _{cidn} ^{rpos}	11	cSim ^{max}	32
1057934	29	0.693	0.733	5.77	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{max}	15
649305	37	0.705	0.745	5.74	bSim _x ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	34
734754	26	0.813	0.860	5.74	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	27
670013	33	0.737	0.779	5.69	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{pos}	28
55702	31	0.758	0.800	5.53	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{min}	42
319765	27	0.607	0.640	5.49	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{pos}	6
625172	44	0.766	0.808	5.49	bSim _p ^{al}	alinS _{csim}	9	cSim ^{max}	31
728454	34	0.703	0.741	5.42	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{pos}	29

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
198411	25	0.740	0.780	5.41	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	27
254783	24	0.740	0.780	5.41	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{min}	24
749115	30	0.493	0.520	5.41	bSim _x ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{min}	21
625247	23	0.747	0.787	5.36	bSim _p ^{al}	alinS _{csim}	6	cSim ^{avg}	24
50708	25	0.760	0.800	5.26	bSim _x ^{al}	alinS _{csim}	7	cSim ^{avg}	27
416572	27	0.633	0.667	5.26	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{avg}	11
432010	22	0.633	0.667	5.26	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{avg}	31
270514	26	0.767	0.807	5.22	bSim _p ^{cs}	-	11	cSim ^{avg}	26
39808	26	0.773	0.813	5.17	bSim _x ^{ci}	-	6	cSim ^{avg}	28
275442	27	0.773	0.813	5.17	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	28
661903	26	0.773	0.813	5.17	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{pos}	30
326467	29	0.780	0.820	5.13	bSim _p ^{al}	alinS _{csim} ^{rpos}	3	cSim ^{min}	26
467823	25	0.780	0.820	5.13	bSim _p ^{cs}	-	8	cSim ^{avg}	28
268715	22	0.653	0.687	5.10	bSim _p ^{al}	alinS _{csim} ^{rpos}	3	cSim ^{pos}	23
749871	22	0.787	0.827	5.08	bSim _x ^{al}	alinS _{csim}	4	cSim ^{avg}	33
372021	35	0.752	0.790	5.06	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{min}	2
444031	21	0.660	0.693	5.05	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	13
55683	18	0.667	0.700	5.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{avg}	28
56314	20	0.667	0.700	5.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	29
408340	45	0.667	0.700	5.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{pos}	36
362025	37	0.683	0.717	4.88	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	29
55827	23	0.693	0.727	4.81	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	5
1065638	28	0.693	0.727	4.81	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{min}	14
58101	39	0.745	0.781	4.79	bSim _x ^{ci}	-	5	cSim ^{min}	28

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
372022	37	0.798	0.836	4.78	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	14
270617	20	0.700	0.733	4.76	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{avg}	20
354019	21	0.700	0.733	4.76	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{min}	13
568216	20	0.700	0.733	4.76	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	14
592803	17	0.700	0.733	4.76	bSim _p ^{cs}	-	8	cSim ^{min}	35
598348	18	0.700	0.733	4.76	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
734801	20	0.700	0.733	4.76	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	22
625171	35	0.610	0.638	4.69	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	36
729339	22	0.853	0.893	4.69	bSim _p ^{cs}	-	3	cSim ^{avg}	33
262755	28	0.713	0.747	4.67	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
284208	23	0.713	0.747	4.67	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{max}	38
343839	36	0.562	0.588	4.66	bSim _p ^{al}	alinS _{csim}	5	cSim ^{min}	19
254970	43	0.741	0.775	4.60	bSim _x ^{ci}	-	7	cSim ^{max}	27
550776	28	0.580	0.607	4.60	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	12
37647	19	0.733	0.767	4.55	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
56325	19	0.733	0.767	4.55	bSim _x ^{al}	alinS _{csim}	9	cSim ^{max}	44
267726	20	0.733	0.767	4.55	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{avg}	28
313111	18	0.733	0.767	4.55	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{max}	2
410404	16	0.733	0.767	4.55	bSim _p ^{cs}	-	11	cSim ^{min}	22
620240	30	0.880	0.920	4.55	bSim _p ^{cs}	-	9	cSim ^{avg}	32
747252	28	0.733	0.767	4.55	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	24
343578	16	0.733	0.767	4.54	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
614370	33	0.758	0.792	4.52	bSim _p ^{al}	alinS _{csim}	9	cSim ^{avg}	34
620239	31	0.888	0.928	4.51	bSim _x ^{al}	alinS _{csim}	6	cSim ^{pos}	30

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
284206	22	0.593	0.620	4.50	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{pos}	36
384740	21	0.740	0.773	4.50	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	30
459247	29	0.740	0.773	4.50	bSim _x ^{ci}	-	2	cSim ^{max}	36
470602	26	0.740	0.773	4.50	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	13
700028	45	0.617	0.644	4.50	bSim _p ^{al}	alinS _{csim}	5	cSim ^{max}	41
765521	29	0.747	0.780	4.46	bSim _x ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{avg}	40
57107	39	0.750	0.783	4.44	bSim _p ^{cs}	-	5	cSim ^{max}	26
162144	18	0.767	0.800	4.35	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{avg}	3
197933	19	0.767	0.800	4.35	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	1
346980	19	0.767	0.800	4.35	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{pos}	27
658648	20	0.767	0.800	4.35	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{avg}	22
388703	42	0.659	0.687	4.34	bSim _x ^{ci}	-	9	cSim ^{avg}	36
632046	46	0.717	0.748	4.34	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	30
269185	32	0.573	0.598	4.32	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	24
301059	21	0.773	0.807	4.31	bSim _p ^{al}	alinS _{csim}	9	cSim ^{max}	31
444520	26	0.773	0.807	4.31	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
86581	34	0.754	0.787	4.29	bSim _x ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{pos}	24
479369	43	0.647	0.675	4.29	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{max}	11
57097	21	0.787	0.820	4.24	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
498482	23	0.633	0.660	4.21	bSim _x ^{ci}	-	8	cSim ^{max}	28
37669	54	0.703	0.733	4.20	bSim _x ^{ci}	-	7	cSim ^{max}	30
345341	33	0.682	0.711	4.19	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{min}	36
55830	18	0.800	0.833	4.17	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	29
56189	29	0.800	0.833	4.17	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{min}	37

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
260898	20	0.800	0.833	4.17	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
197948	39	0.752	0.783	4.11	bSim _x ^{ci}	-	6	cSim ^{min}	31
341476	21	0.813	0.847	4.10	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
625198	66	0.668	0.695	4.06	bSim _p ^{cs}	-	12	cSim ^{pos}	22
625185	52	0.598	0.623	4.05	bSim _p ^{al}	alinS _{csim}	6	cSim ^{pos}	33
459283	43	0.730	0.759	4.02	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{max}	13
326396	21	0.667	0.693	4.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	31
459280	20	0.833	0.867	4.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	29
479548	20	0.833	0.867	4.00	bSim _x ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{min}	18
550507	18	0.833	0.867	4.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	14
57075	37	0.781	0.812	3.96	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{avg}	9
271946	44	0.841	0.875	3.96	bSim _p ^{cs}	-	5	cSim ^{max}	47
39498	73	0.656	0.682	3.95	bSim _x ^{ci}	-	6	cSim ^{pos}	30
242797	35	0.724	0.752	3.95	bSim _p ^{cs}	-	8	cSim ^{max}	34
1105813	29	0.860	0.893	3.88	bSim _x ^{al}	alinS _{csim}	3	cSim ^{min}	33
292040	18	0.867	0.900	3.85	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	30
419781	19	0.867	0.900	3.85	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	14
420358	22	0.867	0.900	3.85	bSim _p ^{cs}	-	7	cSim ^{avg}	41
514141	46	0.692	0.719	3.85	bSim _x ^{al}	alinS _{csim} ^{rpos}	11	cSim ^{avg}	44
694143	30	0.693	0.720	3.85	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	22
701330	31	0.693	0.720	3.85	bSim _p ^{cs}	-	11	cSim ^{avg}	36
274015	27	0.880	0.913	3.79	bSim _x ^{ci}	-	8	cSim ^{min}	40
625254	89	0.671	0.696	3.77	bSim _x ^{al}	alinS _{csim}	7	cSim ^{max}	28
625222	91	0.611	0.634	3.76	bSim _p ^{al}	alinS _{cidn}	4	cSim ^{min}	12

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
58100	41	0.806	0.836	3.74	bSim _p ^{al}	alinS _{csim}	7	cSim ^{max}	34
296092	32	0.611	0.634	3.74	bSim _x ^{ci}	-	4	cSim ^{avg}	29
761384	38	0.705	0.731	3.72	bSim _x ^{ci}	-	4	cSim ^{pos}	31
262950	30	0.720	0.747	3.70	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{pos}	32
625269	49	0.661	0.686	3.70	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{max}	33
90359	29	0.733	0.760	3.64	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{max}	32
161283	27	0.733	0.760	3.64	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{pos}	5
316896	33	0.680	0.705	3.64	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	4
725966	29	0.733	0.760	3.64	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{avg}	12
242577	33	0.792	0.821	3.61	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
764602	21	0.740	0.767	3.60	bSim _p ^{al}	alinS _{csim}	7	cSim ^{avg}	24
270404	22	0.560	0.580	3.57	bSim _p ^{cs}	-	8	cSim ^{min}	40
397743	58	0.684	0.708	3.55	bSim _p ^{cs}	-	9	cSim ^{max}	39
1085881	29	0.753	0.780	3.54	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{avg}	32
649306	37	0.745	0.771	3.51	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	38
747610	27	0.760	0.787	3.51	bSim _p ^{cs}	-	10	cSim ^{avg}	28
625244	30	0.773	0.800	3.45	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{max}	8
767289	18	0.967	1.000	3.45	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	23
55833	59	0.793	0.820	3.44	bSim _p ^{cs}	-	8	cSim ^{max}	23
553065	28	0.780	0.807	3.42	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{min}	34
262172	49	0.689	0.712	3.39	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{avg}	5
483451	29	0.593	0.613	3.37	bSim _p ^{al}	alinS _{csim}	5	cSim ^{avg}	19
56318	58	0.812	0.839	3.36	bSim _x ^{al}	alinS _{csim}	3	cSim ^{avg}	35
300918	26	0.800	0.827	3.33	bSim _p ^{cs}	-	10	cSim ^{min}	35

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
625221	64	0.708	0.731	3.33	bSim _x ^{ci}	-	12	cSim ^{max}	15
711102	22	0.600	0.620	3.33	bSim _p ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{avg}	14
197795	33	0.808	0.834	3.30	bSim _p ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{avg}	31
625204	30	0.813	0.840	3.28	bSim _x ^{al}	alinS _{cSim}	7	cSim ^{max}	30
315992	34	0.585	0.604	3.26	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	33
489474	25	0.620	0.640	3.23	bSim _x ^{ci}	-	4	cSim ^{min}	14
390722	78	0.719	0.742	3.21	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{avg}	9
625153	63	0.704	0.726	3.21	bSim _p ^{cs}	-	13	cSim ^{max}	16
197782	27	0.840	0.867	3.17	bSim _p ^{al}	alinS _{cSim}	5	cSim ^{avg}	41
739861	21	0.840	0.867	3.17	bSim _x ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{min}	33
157306	29	0.853	0.880	3.13	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{min}	30
701329	32	0.731	0.754	3.13	bSim _p ^{al}	alinS _{cSim} ^{rpos}	7	cSim ^{max}	24
1113337	28	0.860	0.887	3.10	bSim _p ^{cs}	-	7	cSim ^{max}	31
690	82	0.615	0.634	3.06	bSim _p ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{max}	38
353146	45	0.733	0.756	3.03	bSim _p ^{al}	alinS _{cSim}	6	cSim ^{avg}	43
412780	77	0.679	0.699	3.02	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{avg}	10
326367	29	0.893	0.920	2.99	bSim _x ^{al}	alinS _{cSim} ^{rpos}	4	cSim ^{max}	27
625228	50	0.751	0.773	2.96	bSim _x ^{al}	alinS _{cSim}	5	cSim ^{pos}	33
452686	48	0.641	0.660	2.95	bSim _x ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{avg}	55
273145	23	0.680	0.700	2.94	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
274572	25	0.680	0.700	2.94	bSim _x ^{al}	alinS _{cSim} ^{rpos}	5	cSim ^{min}	24
461809	36	0.648	0.667	2.94	bSim _x ^{ci}	-	6	cSim ^{avg}	39
723195	22	0.680	0.700	2.94	bSim _p ^{al}	alinS _{cSim} ^{rpos}	6	cSim ^{max}	21
255141	48	0.720	0.741	2.93	bSim _x ^{al}	alinS _{cSim}	6	cSim ^{pos}	36

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
242795	39	0.657	0.676	2.90	bSim _x ^{ci}	-	5	cSim ^{pos}	30
197941	31	0.659	0.678	2.89	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{max}	27
767814	25	0.700	0.720	2.86	bSim _p ^{cs}	-	10	cSim ^{min}	27
625182	26	0.707	0.727	2.83	bSim _p ^{al}	alinS _{cidn}	3	cSim ^{min}	4
196387	56	0.709	0.729	2.82	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{avg}	41
37660	23	0.713	0.733	2.80	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{min}	5
92730	64	0.732	0.753	2.80	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{avg}	33
423862	21	0.720	0.740	2.78	bSim _x ^{ci}	-	4	cSim ^{min}	20
198385	27	0.727	0.747	2.75	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{min}	31
273143	22	0.727	0.747	2.75	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
296876	33	0.623	0.640	2.75	bSim _p ^{al}	alinS _{csim}	4	cSim ^{avg}	34
456091	22	0.727	0.747	2.75	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{pos}	26
300916	32	0.834	0.857	2.74	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{max}	5
452689	26	0.733	0.753	2.73	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{max}	26
325959	26	0.747	0.767	2.68	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{max}	26
410466	60	0.685	0.703	2.65	bSim _p ^{cs}	-	7	cSim ^{avg}	28
625207	104	0.614	0.630	2.63	bSim _p ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{pos}	37
625223	65	0.685	0.703	2.62	bSim _p ^{al}	alinS _{cidn}	3	cSim ^{avg}	14
56186	24	0.767	0.787	2.61	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{pos}	10
240926	23	0.767	0.787	2.61	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	5
282431	24	0.767	0.787	2.61	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{min}	2
318593	24	0.767	0.787	2.61	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
459711	28	0.767	0.787	2.61	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{pos}	18
240892	22	0.773	0.793	2.59	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{min}	13

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
386739	25	0.780	0.800	2.56	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	13
271466	33	0.524	0.537	2.54	bSim _x ^{ci}	-	3	cSim ^{min}	39
409942	22	0.793	0.813	2.52	bSim _x ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	31
197777	24	0.800	0.820	2.50	bSim _x ^{al}	alinS _{csim}	7	cSim ^{max}	37
1065631	42	0.766	0.785	2.49	bSim _p ^{al}	alinS _{csim}	4	cSim ^{pos}	41
436917	29	0.807	0.827	2.48	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{avg}	38
445786	22	0.807	0.827	2.48	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{max}	31
511766	36	0.676	0.693	2.47	bSim _p ^{al}	alinS _{csim}	7	cSim ^{min}	37
629788	68	0.757	0.776	2.47	bSim _p ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{pos}	23
262048	33	0.699	0.716	2.45	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
625256	82	0.654	0.669	2.40	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	27
241889	24	0.840	0.860	2.38	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{max}	34
459282	41	0.701	0.718	2.38	bSim _x ^{ci}	-	7	cSim ^{max}	31
749367	27	0.847	0.867	2.36	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	23
1072705	21	0.853	0.873	2.34	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	24
260157	23	0.860	0.880	2.33	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{avg}	31
39801	51	0.835	0.855	2.32	bSim _p ^{cs}	-	4	cSim ^{min}	34
456100	27	0.867	0.887	2.31	bSim _p ^{cs}	-	13	cSim ^{avg}	31
749669	31	0.743	0.760	2.31	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	24
1057358	31	0.579	0.592	2.30	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{avg}	1
625196	57	0.661	0.676	2.29	bSim _p ^{cs}	-	12	cSim ^{pos}	20
722623	37	0.729	0.745	2.29	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{pos}	33
259060	21	0.600	0.613	2.22	bSim _p ^{cs}	-	8	cSim ^{pos}	31
593714	23	0.907	0.927	2.21	bSim _p ^{cs}	-	7	cSim ^{max}	26

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
319593	35	0.876	0.895	2.17	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{avg}	30
625201	110	0.680	0.694	2.17	bSim _p ^{al}	alinS _{cidn} ^{rpos}	9	cSim ^{max}	37
484805	60	0.703	0.718	2.15	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	43
290408	41	0.748	0.764	2.12	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{min}	38
625203	79	0.695	0.710	2.12	bSim _p ^{al}	alinS _{cidn}	4	cSim ^{max}	8
197925	30	0.640	0.653	2.08	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{avg}	41
262047	32	0.655	0.669	2.03	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{max}	5
1113849	43	0.830	0.847	2.01	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{min}	37
57093	37	0.726	0.741	1.97	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{max}	46
1054736	26	0.680	0.693	1.96	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{pos}	5
179927	48	0.740	0.754	1.95	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	5
241308	23	0.687	0.700	1.94	bSim _x ^{al}	alinS _{csim}	6	cSim ^{avg}	35
386625	27	0.687	0.700	1.94	bSim _x ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{pos}	33
509549	34	0.695	0.709	1.92	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{avg}	9
1107227	31	0.598	0.610	1.91	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{min}	40
345876	118	0.757	0.771	1.90	bSim _p ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{max}	25
625199	72	0.764	0.778	1.90	bSim _p ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{pos}	37
625200	73	0.683	0.696	1.89	bSim _p ^{al}	alinS _{cidn} ^{rpos}	11	cSim ^{avg}	37
39809	84	0.695	0.708	1.88	bSim _p ^{cs}	-	5	cSim ^{max}	37
485980	66	0.775	0.789	1.84	bSim _p ^{al}	alinS _{csim}	12	cSim ^{min}	46
306463	21	0.727	0.740	1.83	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	34
371780	30	0.733	0.747	1.82	bSim _p ^{cs}	-	5	cSim ^{min}	31
321308	33	0.739	0.752	1.80	bSim _x ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{max}	27
625155	63	0.765	0.779	1.80	bSim _x ^{al}	alinS _{csim}	9	cSim ^{max}	41

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
57094	27	0.747	0.760	1.79	bSim _x ^{al}	alinS _{cidn} ^{rpos}	3	cSim ^{max}	31
298278	32	0.752	0.766	1.77	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{avg}	33
625217	120	0.699	0.712	1.76	bSim _p ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{min}	22
386533	31	0.762	0.775	1.75	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	13
625253	59	0.763	0.776	1.75	bSim _x ^{al}	alinS _{csim}	10	cSim ^{avg}	29
343840	33	0.769	0.783	1.73	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	1
56172	37	0.693	0.705	1.72	bSim _x ^{ci}	-	4	cSim ^{min}	26
319594	40	0.864	0.879	1.65	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{min}	4
625192	92	0.732	0.744	1.64	bSim _p ^{al}	alinS _{cidn} ^{rpos}	11	cSim ^{pos}	37
578685	32	0.935	0.951	1.63	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{max}	29
780586	21	0.833	0.847	1.60	bSim _x ^{ci}	-	4	cSim ^{avg}	23
241475	46	0.847	0.860	1.57	bSim _x ^{al}	alinS _{csim}	5	cSim ^{avg}	39
455690	29	0.853	0.867	1.56	bSim _p ^{al}	alinS _{csim} ^{rpos}	2	cSim ^{max}	27
625252	63	0.718	0.729	1.56	bSim _p ^{al}	alinS _{csim} ^{rpos}	11	cSim ^{min}	14
345340	75	0.777	0.789	1.47	bSim _p ^{al}	alinS _{csim}	5	cSim ^{min}	39
262173	55	0.745	0.756	1.46	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{max}	50
433783	57	0.554	0.562	1.42	bSim _p ^{al}	alinS _{csim}	7	cSim ^{max}	44
685045	30	0.947	0.960	1.41	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	24
625154	64	0.693	0.703	1.38	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{avg}	2
290229	37	0.693	0.702	1.37	bSim _x ^{ci}	-	6	cSim ^{max}	29
553066	31	0.699	0.709	1.36	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{avg}	3
625151	67	0.765	0.774	1.24	bSim _x ^{al}	alinS _{csim}	10	cSim ^{min}	28
625152	60	0.742	0.751	1.23	bSim _x ^{ci}	-	5	cSim ^{avg}	22
1055728	24	0.553	0.560	1.21	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{avg}	29

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
212727	35	0.800	0.809	1.19	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
625270	42	0.605	0.612	1.18	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{avg}	5
625218	97	0.752	0.761	1.15	bSim _x ^{al}	alinS _{cidn}	4	cSim ^{pos}	7
625202	100	0.739	0.747	1.14	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{avg}	27
241852	22	0.647	0.653	1.03	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{avg}	4
709834	21	0.700	0.707	0.95	bSim _x ^{al}	alinS _{cidn} ^{rpos}	9	cSim ^{avg}	24
771350	40	0.750	0.757	0.95	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{pos}	43
262948	45	0.617	0.622	0.90	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	8
501798	24	0.740	0.747	0.90	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{avg}	34
466821	28	0.767	0.773	0.87	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	14
625243	44	0.725	0.731	0.77	bSim _x ^{al}	alinS _{cidn} ^{rpos}	7	cSim ^{avg}	27
703400	59	0.789	0.795	0.77	bSim _p ^{cs}	-	14	cSim ^{pos}	27
39812	66	0.737	0.742	0.69	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{pos}	14
242353	82	0.681	0.686	0.68	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{avg}	5
729258	54	0.815	0.820	0.64	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{avg}	34
242369	101	0.613	0.617	0.53	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{avg}	46
242135	59	0.702	0.705	0.52	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{pos}	2
461131	98	0.783	0.787	0.51	bSim _p ^{cs}	-	5	cSim ^{min}	31
497575	66	0.783	0.786	0.47	bSim _x ^{ci}	-	8	cSim ^{min}	37
299948	71	0.782	0.785	0.36	bSim _p ^{cs}	-	9	cSim ^{max}	30
197946	52	0.602	0.602	0.07	bSim _p ^{cs}	-	5	cSim ^{min}	33
42316	20	0.667	0.667	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
56183	20	0.767	0.767	0.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	20
56184	20	0.800	0.800	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{min}	7

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
91424	20	0.900	0.900	0.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	4	cSim ^{max}	26
91426	18	0.767	0.767	0.00	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	3
144372	15	0.800	0.800	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
197934	19	0.633	0.633	0.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{max}	23
240604	19	0.667	0.667	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
241899	16	0.767	0.767	0.00	bSim _p ^{al}	alinS _{cidn}	2	cSim ^{min}	3
260151	20	0.633	0.633	0.00	bSim _p ^{cs}	-	6	cSim ^{max}	28
270405	19	0.567	0.567	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
282900	59	0.753	0.753	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{min}	27
289392	19	0.900	0.900	0.00	bSim _p ^{cs}	-	13	cSim ^{pos}	24
308376	20	0.833	0.833	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	5
315276	18	0.567	0.567	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
343579	21	0.733	0.733	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
347221	16	0.733	0.733	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
362931	40	0.643	0.643	0.00	bSim _x ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{max}	20
395657	30	0.840	0.840	0.00	bSim _x ^{al}	alinS _{cidn} ^{rpos}	6	cSim ^{pos}	25
397267	25	0.580	0.580	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{max}	13
411213	20	0.767	0.767	0.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	8	cSim ^{max}	26
441756	21	0.760	0.760	0.00	bSim _p ^{cs}	-	9	cSim ^{avg}	47
459948	26	0.773	0.773	0.00	bSim _x ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{avg}	32
472619	20	0.633	0.633	0.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	9	cSim ^{max}	27
576612	105	0.711	0.711	0.00	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{max}	2
580195	26	0.640	0.640	0.00	bSim _x ^{ci}	-	6	cSim ^{max}	29
611101	19	0.800	0.800	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14

Continued on next page

Table S3 – continued from previous page

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ \mathcal{B}_i^+ $	cSim	$ \mathcal{C}_i^+ $
623021	30	0.880	0.880	0.00	bSim _x ^{al}	alinS _{cidn}	2	cSim ^{pos}	8
625193	30	0.787	0.787	0.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	10	cSim ^{max}	28
625263	45	0.883	0.883	0.00	bSim _p ^{al}	alinS _{cidn}	3	cSim ^{max}	4
727393	16	0.800	0.800	0.00	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	24
736008	22	0.680	0.680	0.00	bSim _p ^{al}	alinS _{csim}	5	cSim ^{avg}	19
759595	43	0.878	0.878	0.00	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	4
1057953	30	0.800	0.800	0.00	bSim _x ^{al}	alinS _{cidn}	3	cSim ^{min}	8
1064827	34	0.838	0.838	0.00	bSim _x ^{ci}	-	10	cSim ^{avg}	30
1069523	21	0.913	0.913	0.00	bSim _p ^{al}	alinS _{csim} ^{rpos}	4	cSim ^{avg}	22
471117	32	0.842	0.840	-0.23	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	14
482940	32	0.726	0.724	-0.26	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{min}	17
39820	38	0.667	0.664	-0.36	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
471239	36	0.588	0.583	-0.81	bSim _x ^{al}	alinS _{cidn} ^{rpos}	8	cSim ^{max}	26
397266	26	0.693	0.687	-0.96	bSim _p ^{al}	alinS _{cidn}	1	cSim ^{avg}	2
1076089	22	0.607	0.600	-1.10	bSim _x ^{al}	alinS _{cidn} ^{rpos}	5	cSim ^{avg}	28
405937	35	0.724	0.714	-1.32	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{pos}	14
578229	36	0.700	0.691	-1.36	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	14
293006	29	0.820	0.807	-1.63	bSim _x ^{ci}	-	5	cSim ^{max}	38
677013	37	0.852	0.838	-1.68	bSim _x ^{al}	alinS _{cidn}	1	cSim ^{avg}	9
318594	24	0.780	0.760	-2.56	bSim _x ^{ci}	-	5	cSim ^{min}	38
282432	21	0.767	0.747	-2.61	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
254777	24	0.747	0.727	-2.68	bSim _p ^{al}	alinS _{csim} ^{rpos}	6	cSim ^{max}	26
240961	31	0.804	0.781	-2.84	bSim _x ^{ci}	-	9	cSim ^{max}	32
1057565	17	0.867	0.833	-3.85	bSim _p ^{al}	alinS _{csim} ^{rpos}	5	cSim ^{max}	21

Continued on next page

Table S3 – continued from previous page

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	bSim	alinS	$ B_i^+ $	cSim	$ C_i^+ $
452690	27	0.747	0.713	-4.47	bSim _p ^{al}	alinS _{csim} ^{rpos}	7	cSim ^{min}	21
267774	17	0.733	0.700	-4.54	bSim _p ^{al}	alinS _{cidn} ^{rpos}	1	cSim ^{avg}	5
306462	22	0.727	0.693	-4.59	bSim _p ^{al}	alinS _{cidn} ^{rpos}	2	cSim ^{avg}	5
361790	28	0.773	0.733	-5.17	bSim _x ^{al}	alinS _{cidn}	3	cSim ^{avg}	21
262754	28	0.587	0.553	-5.68	bSim _p ^{al}	alinS _{csim}	5	cSim ^{pos}	25
104448	22	0.907	0.853	-5.88	bSim _x ^{al}	alinS _{cidn} ^{rpos}	9	cSim ^{max}	34
261732	30	0.653	0.613	-6.12	bSim _x ^{ci}	-	2	cSim ^{max}	10

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ $|C_i|$ ” has the bioassay size. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “bSim” and “cSim”, respectively, show the assistance bioassay selection method and the assistance compound selection method that result in the best improvement. The column corresponding to “alinS” has the ranking list alignment scoring schemes used in cSim, if applicable. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

S6 Bioassay Assistance Relations

To visualize the assistance relations among bioassays, we generated an assistance relation graph as in Figure S1 using Cytoscape^{2,3}. The directed graph was generated from the best performing method for each bioassay. In Figure S1, each node represents a bioassay, and each edge represents an assistance relation that the bioassay represented by the source node of the edge provide assistance compounds to the bioassay represented by the target node of the edge. The colors of the edges represent the corresponding bioassay selection methods (color codes are available in Figure S1). The width of the edges indicates the number of assistance compounds (wider edges correspond to more assistance compounds). Figure S1

²<http://www.cytoscape.org/download.php>

³http://chianti.ucsd.edu/cytoscape-3.4.0/Cytoscape_3_4_0_unix.sh

only presents the edges that correspond to more than 10 assistance compounds, and the largest connected component among all the connected components based on the assistance relations.

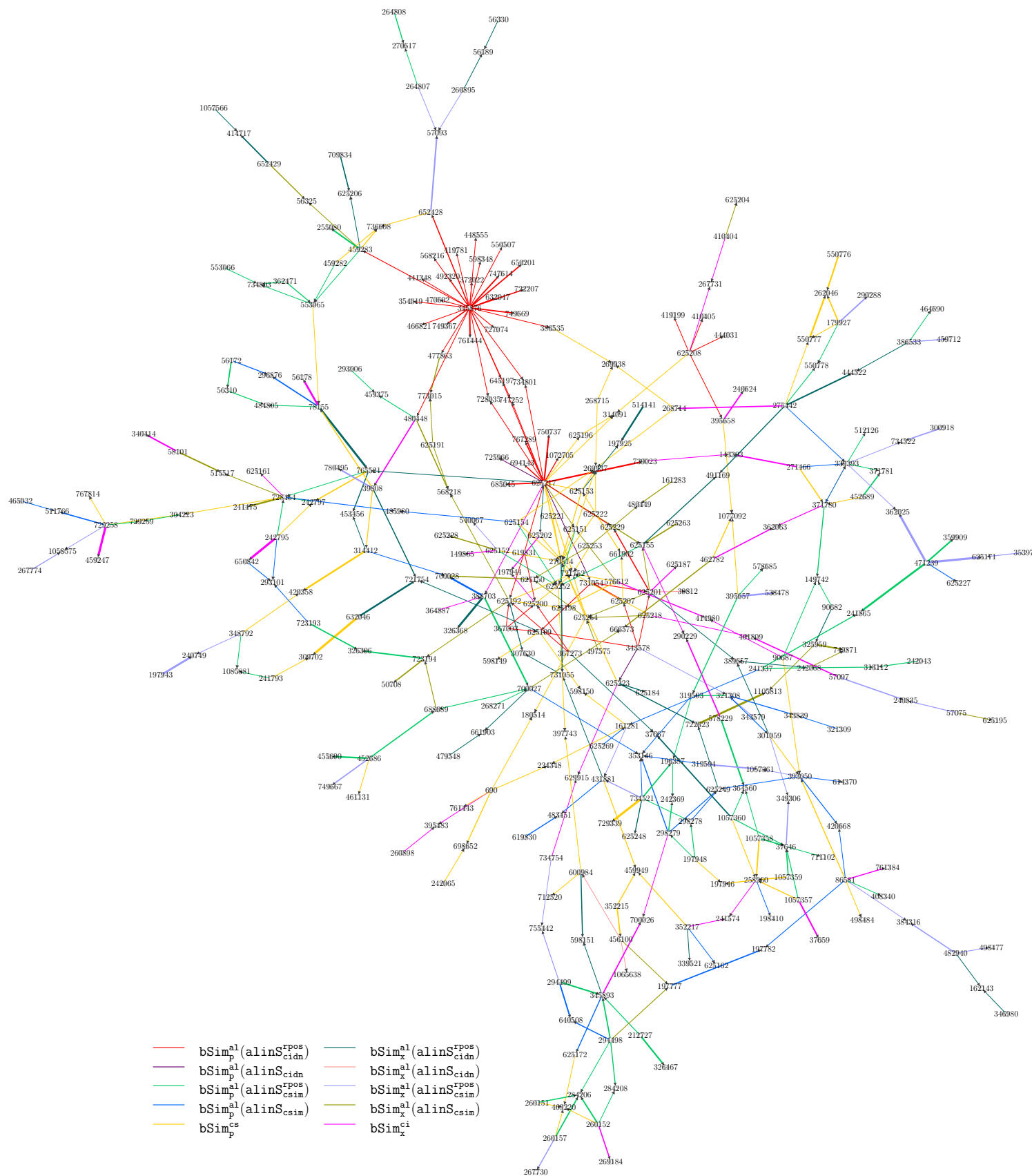


Figure S1: Assistance relations among bioassays

S7 Assistance Bioassay and Compound Selection Method Comparison

Table S4 presents the performance comparison for all the assistance bioassay selection and assistance compound selection methods.

Table S4: Results Summary

bSim	alinS	cSim	imprv(%)	imprv-std	#+	imprv(+%)	#-	imprv(-%)	#0
bSim _x ^{al}	alinS _{cSim}	cSim ^{max}	0.68	7.68	296	6.38	282	-5.09	87
bSim _x ^{al}	alinS _{cSim} ^{rpos}	cSim ^{pos}	0.60	8.12	286	6.67	295	-5.11	84
bSim _x ^{al}	alinS _{cidn} ^{rpos}	cSim ^{max}	0.51	7.37	284	6.32	294	-4.97	87
bSim _x ^{al}	alinS _{cSim} ^{rpos}	cSim ^{max}	0.49	7.56	279	6.62	287	-5.30	99
bSim _x ^{al}	alinS _{cSim}	cSim ^{pos}	0.48	7.62	286	6.34	291	-5.14	88
bSim _x ^{ci}	-	cSim ^{min}	0.48	7.38	281	6.38	295	-4.99	89
bSim _x ^{al}	alinS _{cSim} ^{rpos}	cSim ^{avg}	0.47	7.87	276	6.77	302	-5.15	87
bSim _x ^{al}	alinS _{cSim}	cSim ^{avg}	0.47	7.44	285	6.24	293	-5.02	87
bSim _x ^{ci}	-	cSim ^{max}	0.42	7.23	282	6.31	298	-5.02	85
bSim _p ^{al}	alinS _{cSim}	cSim ^{pos}	0.38	7.35	281	6.03	290	-4.97	94
bSim _x ^{al}	alinS _{cidn}	cSim ^{max}	0.33	6.21	261	5.37	283	-4.17	121
bSim _x ^{al}	alinS _{cidn}	cSim ^{pos}	0.33	5.89	265	5.19	276	-4.18	124
bSim _x ^{al}	alinS _{cSim}	cSim ^{min}	0.27	7.25	283	6.04	299	-5.12	83
bSim _p ^{al}	alinS _{cidn} ^{rpos}	cSim ^{max}	0.26	6.53	279	5.27	274	-4.74	112
bSim _p ^{cs}	-	cSim ^{max}	0.25	7.75	272	6.24	286	-5.34	107
bSim _x ^{al}	alinS _{cidn}	cSim ^{avg}	0.25	5.97	252	5.35	289	-4.09	124
bSim _x ^{al}	alinS _{cidn} ^{rpos}	cSim ^{avg}	0.23	7.47	276	6.33	298	-5.34	91
bSim _p ^{al}	alinS _{cidn} ^{rpos}	cSim ^{min}	0.18	6.67	262	5.68	296	-4.62	107

Continued on next page

Table S4 – continued from previous page

bSim	alinS	cSim	imprv(%)	imprv-std	#+	imprv(+%)	#–	imprv(–%)	#0
bSim _x ^{al}	alinS _{cSim} ^{rpos}	cSim ^{min}	0.17	7.50	274	6.26	306	-5.25	85
bSim _p ^{al}	alinS _{cSim} ^{rpos}	cSim ^{min}	0.10	7.62	275	6.30	310	-5.37	80
bSim _p ^{al}	alinS _{cSim} ^{rpos}	cSim ^{pos}	0.10	7.15	275	5.93	298	-5.24	92
bSim _p ^{al}	alinS _{cSim}	cSim ^{max}	0.10	8.37	260	6.54	316	-5.17	89
bSim _x ^{al}	alinS _{cidn} ^{rpos}	cSim ^{pos}	0.10	7.52	272	6.22	303	-5.36	90
bSim _p ^{al}	alinS _{cidn} ^{rpos}	cSim ^{pos}	0.08	6.28	266	5.23	288	-4.65	111
bSim _p ^{al}	alinS _{cidn} ^{rpos}	cSim ^{avg}	0.07	6.56	270	5.23	281	-4.87	114
bSim _x ^{al}	alinS _{cidn} ^{rpos}	cSim ^{min}	0.06	7.75	264	6.61	303	-5.62	98
bSim _p ^{al}	alinS _{cSim}	cSim ^{avg}	0.05	7.26	274	5.70	306	-4.99	85
bSim _p ^{cs}	-	cSim ^{pos}	0.02	7.70	265	5.95	305	-5.12	95
bSim _p ^{al}	alinS _{cSim} ^{rpos}	cSim ^{avg}	0.00	6.92	282	5.41	298	-5.12	85
bSim _p ^{al}	alinS _{cSim}	cSim ^{min}	0.00	7.66	280	6.01	302	-5.56	83
bSim _x ^{al}	alinS _{cidn}	cSim ^{min}	-0.00	5.97	258	4.92	281	-4.52	126
bSim _p ^{cs}	-	cSim ^{avg}	-0.01	7.74	258	6.19	320	-5.01	87
bSim _x ^{ci}	-	cSim ^{pos}	-0.03	6.99	275	5.87	314	-5.20	76
bSim _p ^{al}	alinS _{cSim} ^{rpos}	cSim ^{max}	-0.04	8.37	267	6.36	318	-5.42	80
bSim _x ^{ci}	-	cSim ^{avg}	-0.04	7.03	266	6.02	315	-5.16	84
bSim _p ^{cs}	-	cSim ^{min}	-0.07	7.69	270	5.87	297	-5.49	98
bSim _p ^{al}	alinS _{cidn}	cSim ^{max}	-0.20	5.89	246	4.78	290	-4.50	129
bSim _p ^{al}	alinS _{cidn}	cSim ^{pos}	-0.26	5.71	226	4.98	293	-4.43	146
bSim _p ^{al}	alinS _{cidn}	cSim ^{avg}	-0.30	6.11	230	4.99	294	-4.59	141
bSim _p ^{al}	alinS _{cidn}	cSim ^{min}	-0.35	5.74	228	4.90	294	-4.60	143

Continued on next page

Table S4 – continued from previous page

bSim	alinS	cSim	imprv(%)	imprv-std	#+	imprv(+%)	#-	imprv(-%)	#0
------	-------	------	----------	-----------	----	-----------	----	-----------	----

The columns corresponding to “bSim” and “cSim”, respectively, show the assistance bioassay selection method and the assistance compound selection method that result in the best improvement. The column corresponding to “alinS” has the ranking list alignment scoring schemes used in cSim, if applicable. The column corresponding to “imprv (%)”/“imprv-std” has the average improvement/improvement standard deviation of the best models over the baseline models in percentage. The columns corresponding to “#+”, “#-” and “#0” have the numbers of bioassays whose ranking models have been improved, declined and not changed, respectively. The columns corresponding to “imprv(#+%)” and “imprv(#-%)” have the average improvement and decline of the new models over the baseline models in percentage.

S8 Assistance Bioassay and Compound Selection Method Performance

Table S5-S44 present the performance of assistance bioassay selection and assistance compound selection methods on the top-10 bioassays, respectively.

Table S5: Top-10 Performance of $\text{bSim}^{\text{al}}(\text{alinS}_{\text{cldn}}^{\text{rpos}})\text{-cSim}^{\text{avg}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
349306	21	0.420	0.640	52.38	2	14
477863	19	0.267	0.367	37.50	1	14
459949	26	0.367	0.467	27.27	2	14
261405	16	0.533	0.667	25.00	1	5
1094446	24	0.440	0.547	24.24	2	23
480068	16	0.600	0.733	22.22	2	14
723193	24	0.480	0.580	20.83	2	24
197943	32	0.442	0.526	18.96	2	5
149865	22	0.533	0.633	18.75	1	5
749667	22	0.467	0.547	17.14	2	24

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S6: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{cidn}}^{\text{rpos}})\text{-cSim}^{\text{max}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
349306	21	0.420	0.660	57.14	2	13
459949	26	0.367	0.487	32.73	2	13
761444	19	0.433	0.567	30.77	2	24
149865	22	0.533	0.667	25.00	1	5
261405	16	0.533	0.667	25.00	1	5
480068	16	0.600	0.733	22.22	2	13
260896	22	0.553	0.660	19.28	2	5
197943	32	0.442	0.526	18.96	2	5
90687	20	0.533	0.633	18.75	1	5
491170	23	0.447	0.527	17.91	2	16

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S7: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{cldn}}^{\text{rpos}})\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.400	50.00	2	12
55843	22	0.460	0.607	31.88	2	5
459949	26	0.367	0.467	27.27	2	13
491170	23	0.447	0.560	25.37	2	13
261405	16	0.533	0.667	25.00	1	5
626142	26	0.513	0.640	24.68	1	22
1094446	24	0.440	0.547	24.24	2	24
480068	16	0.600	0.733	22.22	2	13
701506	20	0.467	0.567	21.43	2	23
723193	24	0.480	0.580	20.83	2	23

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|\mathcal{B}_i^+|$ " and " $|\mathcal{C}_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S8: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{cldn}}^{\text{rpos}})\text{-cSim}^{\text{pos}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
477863	19	0.267	0.367	37.50	1	14
761444	19	0.433	0.567	30.77	2	23
459949	26	0.367	0.467	27.27	2	14
261405	16	0.533	0.667	25.00	1	5
480068	16	0.600	0.733	22.22	2	14
701506	20	0.467	0.567	21.43	2	24
723193	24	0.480	0.580	20.83	2	24
626142	26	0.513	0.620	20.78	2	23
349306	21	0.420	0.507	20.64	2	14
197943	32	0.442	0.526	18.96	2	5

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S9: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{cldn}})\text{-cSim}^{\text{avg}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
349306	21	0.420	0.640	52.38	2	14
477863	19	0.267	0.367	37.50	1	14
261405	16	0.533	0.667	25.00	1	5
619831	15	0.267	0.333	25.00	2	3
1094446	24	0.440	0.547	24.24	2	23
480068	16	0.600	0.733	22.22	2	14
149865	22	0.533	0.633	18.75	1	5
389657	26	0.587	0.687	17.05	2	14
160735	19	0.667	0.767	15.00	1	4
652429	31	0.678	0.775	14.33	2	13

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|\mathcal{B}_i^+|$ " and " $|\mathcal{C}_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S10: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{cidn}})\text{-cSim}^{\text{max}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
349306	21	0.420	0.660	57.14	2	13
149865	22	0.533	0.667	25.00	1	5
261405	16	0.533	0.667	25.00	1	5
480068	16	0.600	0.733	22.22	2	13
722207	22	0.413	0.487	17.74	2	22
389657	26	0.587	0.687	17.05	2	13
761444	19	0.433	0.500	15.39	2	12
160735	19	0.667	0.767	15.00	2	5
417910	21	0.607	0.693	14.28	2	14
701506	20	0.467	0.533	14.28	2	23

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S11: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{cidn}})\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
261405	16	0.533	0.667	25.00	1	5
160735	19	0.667	0.833	25.00	2	4
1094446	24	0.440	0.547	24.24	2	24
480068	16	0.600	0.733	22.22	2	13
701506	20	0.467	0.567	21.43	2	23
149865	22	0.533	0.633	18.75	1	5
722207	22	0.413	0.487	17.74	2	22
595800	20	0.600	0.700	16.67	2	13
652429	31	0.678	0.789	16.29	2	13
727074	23	0.707	0.820	16.04	2	22

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|\mathcal{B}_i^+|$ " and " $|\mathcal{C}_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S12: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{cidn}})\text{-cSim}^{\text{pos}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
477863	19	0.267	0.367	37.50	1	14
261405	16	0.533	0.667	25.00	1	5
619831	15	0.267	0.333	25.00	2	3
480068	16	0.600	0.733	22.22	2	14
701506	20	0.467	0.567	21.43	2	24
349306	21	0.420	0.507	20.64	2	14
149865	22	0.533	0.633	18.75	1	5
389657	26	0.587	0.687	17.05	2	14
595800	20	0.600	0.700	16.67	2	14
160735	19	0.667	0.767	15.00	1	4

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S13: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{avg}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
619831	15	0.267	0.400	50.00	2	15
459949	26	0.367	0.487	32.73	6	24
162143	18	0.367	0.467	27.27	6	29
773016	25	0.540	0.680	25.93	5	25
749667	22	0.467	0.580	24.28	7	23
258960	23	0.527	0.640	21.52	5	20
149865	22	0.533	0.647	21.25	6	29
626142	26	0.513	0.620	20.78	5	22
241201	21	0.707	0.833	17.92	4	28
751624	20	0.567	0.667	17.65	6	23

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S14: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{max}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
619831	15	0.267	0.533	100.00	2	13
459949	26	0.367	0.500	36.36	6	24
625258	20	0.400	0.533	33.33	5	21
241231	21	0.633	0.840	32.63	5	28
498478	21	0.340	0.440	29.41	7	23
255080	20	0.633	0.800	26.31	3	30
261405	16	0.533	0.667	25.00	7	38
761444	19	0.433	0.533	23.08	7	21
749668	24	0.493	0.607	22.97	5	21
258960	23	0.527	0.633	20.25	5	25

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S15: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.467	75.00	2	14
261405	16	0.533	0.733	37.50	9	38
626142	26	0.513	0.660	28.57	7	18
477863	19	0.267	0.333	25.00	5	24
55843	22	0.460	0.567	23.19	6	25
1094446	24	0.440	0.533	21.21	7	27
241231	21	0.633	0.760	20.00	6	27
160735	19	0.667	0.800	20.00	5	33
773016	25	0.540	0.640	18.52	4	20
162143	18	0.367	0.433	18.18	8	28

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S16: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{pos}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
619831	15	0.267	0.400	50.00	2	15
149865	22	0.533	0.720	35.00	7	29
261405	16	0.533	0.667	25.00	6	38
258960	23	0.527	0.653	24.05	5	20
773016	25	0.540	0.660	22.22	5	25
459949	26	0.367	0.447	21.82	6	24
454488	20	0.467	0.567	21.43	6	18
255080	20	0.633	0.767	21.05	4	30
162143	18	0.367	0.433	18.18	7	29
241201	21	0.707	0.833	17.92	4	28

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S17: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{avg}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
619831	15	0.267	0.467	75.00	2	14
459949	26	0.367	0.487	32.73	6	24
162143	18	0.367	0.467	27.27	6	30
773016	25	0.540	0.680	25.93	5	25
349306	21	0.420	0.520	23.81	5	27
258960	23	0.527	0.640	21.52	5	20
749667	22	0.467	0.567	21.43	6	22
751624	20	0.567	0.667	17.65	6	24
262046	24	0.580	0.680	17.24	5	23
626142	26	0.513	0.600	16.88	5	21

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S18: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{max}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.533	100.00	2	13
625258	20	0.400	0.567	41.67	5	19
459949	26	0.367	0.500	36.36	6	24
241231	21	0.633	0.840	32.63	5	28
498478	21	0.340	0.440	29.41	7	23
255080	20	0.633	0.800	26.31	3	30
261405	16	0.533	0.667	25.00	6	36
477863	19	0.267	0.333	25.00	5	25
352215	24	0.400	0.493	23.33	4	27
761444	19	0.433	0.533	23.08	7	21

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S19: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.467	75.00	2	13
626142	26	0.513	0.680	32.47	7	18
261405	16	0.533	0.667	25.00	8	38
477863	19	0.267	0.333	25.00	5	27
55843	22	0.460	0.560	21.74	7	24
241231	21	0.633	0.760	20.00	6	27
160735	19	0.667	0.800	20.00	6	30
625249	38	0.505	0.600	18.87	6	32
241201	21	0.707	0.840	18.87	6	28
773016	25	0.540	0.640	18.52	4	20

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S20: Top-10 Performance of $\text{bSim}_p^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{pos}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.467	75.00	2	14
349306	21	0.420	0.553	31.75	5	27
149865	22	0.533	0.687	28.75	7	27
625258	20	0.400	0.500	25.00	4	21
261405	16	0.533	0.667	25.00	6	36
258960	23	0.527	0.653	24.05	5	20
773016	25	0.540	0.660	22.22	5	25
459949	26	0.367	0.447	21.82	6	24
454488	20	0.467	0.567	21.43	7	18
162143	18	0.367	0.433	18.18	7	30

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S21: Top-10 Performance of $\text{bSim}_p^{\text{cs}}\text{-cSim}^{\text{avg}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.467	75.00	7	26
261405	16	0.533	0.800	50.00	9	43
264807	18	0.600	0.833	38.89	3	24
459949	26	0.367	0.487	32.73	7	26
498478	21	0.340	0.440	29.41	11	28
258960	23	0.527	0.660	25.32	5	24
477863	19	0.267	0.333	25.00	9	30
749667	22	0.467	0.580	24.28	8	21
255080	20	0.633	0.767	21.05	5	34
240749	23	0.453	0.547	20.59	5	32

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S22: Top-10 Performance of $\text{bSim}_p^{\text{cs}}\text{-cSim}^{\text{max}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
619831	15	0.267	0.467	75.00	6	26
261405	16	0.533	0.800	50.00	9	33
498478	21	0.340	0.473	39.21	10	26
264807	18	0.600	0.833	38.89	5	28
625258	20	0.400	0.500	25.00	13	29
240749	23	0.453	0.560	23.53	5	34
761444	19	0.433	0.533	23.08	11	36
749668	24	0.493	0.607	22.97	8	20
459949	26	0.367	0.447	21.82	11	27
260896	22	0.553	0.673	21.69	10	25

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S23: Top-10 Performance of $\text{bSim}_p^{\text{cs}}\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.467	75.00	5	26
264807	18	0.600	0.833	38.89	5	28
459949	26	0.367	0.500	36.36	10	28
626142	26	0.513	0.660	28.57	6	23
261405	16	0.533	0.667	25.00	11	44
477863	19	0.267	0.333	25.00	7	33
721752	21	0.553	0.687	24.10	12	27
417910	21	0.607	0.740	21.98	5	41
454488	20	0.467	0.567	21.43	12	31
491170	23	0.447	0.540	20.89	13	23

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S24: Top-10 Performance of $\text{bSim}_p^{\text{cs}}\text{-cSim}^{\text{pos}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.467	75.00	7	26
261405	16	0.533	0.800	50.00	9	43
264807	18	0.600	0.833	38.89	3	24
459949	26	0.367	0.507	38.18	6	26
477863	19	0.267	0.367	37.50	9	30
498478	21	0.340	0.440	29.41	11	28
240749	23	0.453	0.560	23.53	5	32
454488	20	0.467	0.567	21.43	9	33
262046	24	0.580	0.700	20.69	5	32
241201	21	0.707	0.840	18.87	10	38

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S25: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cldn}}^{\text{rpos}})\text{-cSim}^{\text{avg}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
619831	15	0.267	0.400	50.00	7	29
162143	18	0.367	0.500	36.36	6	35
459949	26	0.367	0.480	30.91	5	22
149865	22	0.533	0.667	25.00	6	38
37667	17	0.400	0.500	25.00	4	29
723193	24	0.480	0.600	25.00	8	45
261405	16	0.533	0.667	25.00	7	39
47369	22	0.353	0.440	24.53	7	35
701506	20	0.467	0.567	21.43	4	23
749667	22	0.467	0.567	21.43	7	33

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S26: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cidn}}^{\text{rpos}})\text{-cSim}^{\text{max}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
162143	18	0.367	0.500	36.36	6	34
274062	22	0.653	0.867	32.65	7	42
626142	26	0.513	0.680	32.47	4	30
389657	26	0.587	0.773	31.82	8	35
149865	22	0.533	0.687	28.75	5	36
701506	20	0.467	0.600	28.57	4	23
459949	26	0.367	0.460	25.45	3	21
261405	16	0.533	0.667	25.00	6	33
619831	15	0.267	0.333	25.00	6	28
489476	23	0.453	0.560	23.53	9	24

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|B_i^+|$ " and " $|C_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S27: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cidn}}^{\text{rpos}})\text{-cSim}^{\text{min}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
149865	22	0.533	0.733	37.50	7	37
162143	18	0.367	0.500	36.36	5	33
459949	26	0.367	0.480	30.91	6	23
723193	24	0.480	0.627	30.56	6	42
274062	22	0.653	0.847	29.59	9	45
701506	20	0.467	0.600	28.57	4	23
389657	26	0.587	0.733	25.00	6	32
625245	24	0.553	0.680	22.89	7	35
47369	22	0.353	0.433	22.64	10	34
432545	22	0.567	0.687	21.18	4	35

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S28: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cldn}}^{\text{rpos}})\text{-cSim}^{\text{pos}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
619831	15	0.267	0.400	50.00	7	29
162143	18	0.367	0.500	36.36	6	35
459949	26	0.367	0.500	36.36	4	22
626142	26	0.513	0.680	32.47	5	30
349306	21	0.420	0.533	26.98	8	40
149865	22	0.533	0.667	25.00	6	38
37667	17	0.400	0.500	25.00	4	29
723193	24	0.480	0.600	25.00	7	45
261405	16	0.533	0.667	25.00	7	39
701506	20	0.467	0.567	21.43	4	23

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S29: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cidn}})\text{-cSim}^{\text{avg}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.400	50.00	1	6
701506	20	0.467	0.600	28.57	2	6
459949	26	0.367	0.460	25.45	1	4
261405	16	0.533	0.667	25.00	1	6
160735	19	0.667	0.833	25.00	2	6
441347	32	0.398	0.490	22.97	1	4
480068	16	0.600	0.733	22.22	2	6
197943	32	0.442	0.535	21.12	3	10
367273	15	0.667	0.800	20.00	1	14
491170	23	0.447	0.527	17.91	2	11

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|\mathcal{B}_i^+|$ " and " $|\mathcal{C}_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S30: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cidn}})\text{-cSim}^{\text{max}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.400	50.00	1	6
701506	20	0.467	0.600	28.57	2	6
261405	16	0.533	0.667	25.00	1	6
160735	19	0.667	0.833	25.00	2	6
477863	19	0.267	0.333	25.00	3	6
626142	26	0.513	0.640	24.68	1	8
480068	16	0.600	0.733	22.22	2	7
349306	21	0.420	0.507	20.64	1	5
367273	15	0.667	0.800	20.00	1	14
441347	32	0.398	0.476	19.62	1	4

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S31: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cidn}})\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
701506	20	0.467	0.600	28.57	1	6
149865	22	0.533	0.673	26.25	2	5
459949	26	0.367	0.460	25.45	1	4
197943	32	0.442	0.545	23.27	3	9
274062	22	0.653	0.793	21.43	1	4
626140	24	0.587	0.707	20.45	1	6
160735	19	0.667	0.800	20.00	1	5
491170	23	0.447	0.527	17.91	2	11
732855	31	0.678	0.794	17.13	1	3
241231	21	0.633	0.740	16.84	1	6

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S32: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{cidn}})\text{-cSim}^{\text{pos}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.400	50.00	1	6
701506	20	0.467	0.600	28.57	2	6
459949	26	0.367	0.460	25.45	1	4
626142	26	0.513	0.640	24.68	1	8
441347	32	0.398	0.490	22.97	1	4
197943	32	0.442	0.535	21.12	3	10
261405	16	0.533	0.633	18.75	1	6
732855	31	0.678	0.794	17.13	1	3
90682	22	0.593	0.693	16.85	2	7
241231	21	0.633	0.740	16.84	1	6

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S33: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{avg}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
349306	21	0.420	0.627	49.21	6	35
240749	23	0.453	0.660	45.59	3	34
625258	20	0.400	0.567	41.67	4	28
261405	16	0.533	0.733	37.50	5	34
55843	22	0.460	0.627	36.23	4	37
761444	19	0.433	0.567	30.77	8	34
264807	18	0.600	0.767	27.78	2	23
619831	15	0.267	0.333	25.00	4	21
626140	24	0.587	0.720	22.73	8	30
459949	26	0.367	0.447	21.82	5	30

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S34: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{max}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
749667	22	0.467	0.667	42.86	9	38
459949	26	0.367	0.500	36.36	5	28
312172	23	0.407	0.533	31.15	5	23
349306	21	0.420	0.547	30.16	7	35
240749	23	0.453	0.587	29.41	3	35
498478	21	0.340	0.440	29.41	4	32
274062	22	0.653	0.833	27.55	5	36
1094446	24	0.440	0.560	27.27	12	50
619831	15	0.267	0.333	25.00	3	21
761444	19	0.433	0.533	23.08	9	31

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|\mathcal{B}_i^+|$ ” and “ $|\mathcal{C}_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S35: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
349306	21	0.420	0.640	52.38	6	33
459949	26	0.367	0.500	36.36	4	29
723193	24	0.480	0.647	34.72	9	38
749667	22	0.467	0.620	32.86	6	38
55843	22	0.460	0.607	31.88	6	40
264807	18	0.600	0.767	27.78	2	22
491170	23	0.447	0.560	25.37	7	27
625258	20	0.400	0.500	25.00	5	35
619831	15	0.267	0.333	25.00	3	20
1094446	24	0.440	0.547	24.24	12	48

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|\mathcal{B}_i^+|$ " and " $|\mathcal{C}_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S36: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}}^{\text{rpos}})\text{-cSim}^{\text{pos}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
349306	21	0.420	0.680	61.90	6	35
625258	20	0.400	0.600	50.00	4	28
55843	22	0.460	0.680	47.83	4	37
240749	23	0.453	0.660	45.59	3	34
261405	16	0.533	0.733	37.50	4	34
749667	22	0.467	0.620	32.86	6	37
459949	26	0.367	0.467	27.27	6	30
319592	20	0.633	0.800	26.32	6	44
417910	21	0.607	0.760	25.27	7	27
619831	15	0.267	0.333	25.00	4	21

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|B_i^+|$ " and " $|C_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S37: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{avg}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
240749	23	0.453	0.660	45.59	3	35
625258	20	0.400	0.533	33.33	4	32
261405	16	0.533	0.700	31.25	4	47
459949	26	0.367	0.480	30.91	4	32
761444	19	0.433	0.567	30.77	9	36
312172	23	0.407	0.527	29.51	7	32
349306	21	0.420	0.540	28.57	6	38
389657	26	0.587	0.740	26.14	7	43
619831	15	0.267	0.333	25.00	6	30
626140	24	0.587	0.720	22.73	8	32

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S38: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{max}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.400	50.00	7	29
459949	26	0.367	0.500	36.36	5	30
261405	16	0.533	0.700	31.25	5	38
389657	26	0.587	0.767	30.68	9	40
349306	21	0.420	0.547	30.16	8	37
312172	23	0.407	0.527	29.51	7	32
240749	23	0.453	0.587	29.41	3	36
498478	21	0.340	0.440	29.41	6	34
241201	21	0.707	0.887	25.47	6	38
625150	23	0.533	0.667	25.00	6	41

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|\mathcal{B}_i^+|$ " and " $|\mathcal{C}_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S39: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{min}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
55843	22	0.460	0.640	39.13	6	40
459949	26	0.367	0.500	36.36	8	33
723193	24	0.480	0.647	34.72	10	40
625258	20	0.400	0.533	33.33	5	39
349306	21	0.420	0.540	28.57	6	33
749667	22	0.467	0.600	28.57	7	41
491170	23	0.447	0.560	25.37	8	31
261405	16	0.533	0.667	25.00	6	49
477863	19	0.267	0.333	25.00	4	36
619831	15	0.267	0.333	25.00	6	27

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S40: Top-10 Performance of $\text{bSim}_x^{\text{al}}(\text{alinS}_{\text{csim}})\text{-cSim}^{\text{pos}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
625258	20	0.400	0.600	50.00	4	32
240749	23	0.453	0.660	45.59	3	35
55843	22	0.460	0.660	43.48	4	38
261405	16	0.533	0.700	31.25	3	47
701506	20	0.467	0.600	28.57	4	32
749667	22	0.467	0.600	28.57	8	39
160735	19	0.667	0.833	25.00	4	29
619831	15	0.267	0.333	25.00	6	30
389657	26	0.587	0.727	23.86	7	43
761444	19	0.433	0.533	23.08	10	36

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S41: Top-10 Performance of $\text{bSim}_x^{\text{ci}}\text{-cSim}^{\text{avg}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
149865	22	0.533	0.733	37.50	6	30
459949	26	0.367	0.487	32.73	8	31
761444	19	0.433	0.567	30.77	3	29
349306	21	0.420	0.527	25.40	4	31
261405	16	0.533	0.667	25.00	9	28
619831	15	0.267	0.333	25.00	2	3
260896	22	0.553	0.680	22.89	6	31
701506	20	0.467	0.567	21.43	1	11
749667	22	0.467	0.567	21.43	7	37
162143	18	0.367	0.433	18.18	8	32

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S42: Top-10 Performance of $\text{bSim}_x^{\text{ci}}\text{-cSim}^{\text{max}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
149865	22	0.533	0.747	40.00	6	29
498478	21	0.340	0.440	29.41	8	32
260896	22	0.553	0.700	26.51	4	32
255080	20	0.633	0.800	26.31	3	17
459949	26	0.367	0.460	25.45	5	30
261405	16	0.533	0.667	25.00	8	27
143393	24	0.547	0.673	23.17	6	31
197943	32	0.442	0.543	22.84	4	26
241201	21	0.707	0.867	22.64	4	25
241231	21	0.633	0.753	18.95	7	38

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S43: Top-10 Performance of $\text{bSim}_x^{\text{ci}}\text{-cSim}^{\text{min}}$

B_i	$ \mathcal{C}_i $	RM_i	RM_i^+	imprv (%)	$ \mathcal{B}_i^+ $	$ \mathcal{C}_i^+ $
619831	15	0.267	0.400	50.00	2	3
149865	22	0.533	0.733	37.50	4	29
749667	22	0.467	0.613	31.43	6	36
349306	21	0.420	0.547	30.16	2	30
459949	26	0.367	0.467	27.27	7	33
260896	22	0.553	0.700	26.51	4	34
491170	23	0.447	0.560	25.37	5	38
723193	24	0.480	0.600	25.00	4	40
197943	32	0.442	0.543	22.84	4	28
241201	21	0.707	0.867	22.64	5	24

The column corresponding to " B_i " has the bioassay AIDs from PubChem. The column corresponding to " RM_i " shows the baseline model performance. The column corresponding to " RM_i^+ " has the best improved model performance. The column corresponding to "imprv (%)" has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to " $|\mathcal{B}_i^+|$ " and " $|\mathcal{C}_i^+|$ ", respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.

Table S44: Top-10 Performance of $\text{bSim}_x^{\text{ci}}\text{-cSim}^{\text{pos}}$

B_i	$ C_i $	RM_i	RM_i^+	imprv (%)	$ B_i^+ $	$ C_i^+ $
149865	22	0.533	0.733	37.50	5	30
459949	26	0.367	0.487	32.73	9	31
749667	22	0.467	0.613	31.43	6	37
761444	19	0.433	0.567	30.77	3	29
261405	16	0.533	0.667	25.00	8	28
619831	15	0.267	0.333	25.00	2	3
260896	22	0.553	0.687	24.10	5	31
197943	32	0.442	0.543	22.84	4	27
349306	21	0.420	0.513	22.22	2	31
701506	20	0.467	0.567	21.43	1	11

The column corresponding to “ B_i ” has the bioassay AIDs from PubChem. The column corresponding to “ RM_i ” shows the baseline model performance. The column corresponding to “ RM_i^+ ” has the best improved model performance. The column corresponding to “imprv (%)” has the improvement of the best model (i.e., RM_i^+) over the baseline model (i.e., RM_i) in percentage. The columns corresponding to “ $|B_i^+|$ ” and “ $|C_i^+|$ ”, respectively, show the number of assistance bioassays and the number of assistance compounds incorporated in the improved model.