

## Supplementary Materials

**Table S1.** Summary of the ligand CoA and its 62 derivatives. The ‘Original’ column represents the number of protein chains that bind the corresponding ligand in the rows. The ‘Transfer’ column provides the number of representative sequences after transferring the binding annotations from sequences sharing 100% sequence identity with the representative sequences. The ‘NR’ column is the number of protein chains after removing the redundant sequences at  $\geq 25\%$  sequence identity.

ID	Name	Original	Transfer	NR
COA	Coenzyme A	1059	415	122
ACO	Acetyl Coenzyme *A	373	157	44
CO8	Octanoyl-Coenzyme A	24	8	4
CAA	Acetoacetyl-Coenzyme A	101	27	1
MLC	Malonyl-Coenzyme A	29	12	3
HXC	Hexanoyl-Coenzyme A	28	12	3
1VU	Propionyl Coenzyme A	22	5	3
SCA	Succinyl-Coenzyme A	18	8	3
BCO	Butyryl Coenzyme A	14	12	2
BYC	Benzoyl Coenzyme A	10	4	3
COD	Dephospho Coenzyme A	31	9	2
HMG	3-Hydroxy-3-Methylglutaryl-Coenzyme A	22	8	2
ST9	Stearoyl-Coenzyme A	15	5	2
CAO	Oxidized Coenzyme A	14	4	2
COO	Crotonyl Coenzyme A	11	6	2
2CP	2-Carboxypropyl-Coenzyme A	5	2	1
4BN	3-Oxo-4-Pregnene-20-Carboxyl-Coenzyme A	2	2	2
4CO	4-Hydroxyphenacyl Coenzyme A	19	7	1
CMC	Carboxymethyl Coenzyme *A	17	8	2
0FQ	Phenacyl Coenzyme A	12	2	0
DCA	Desulfo-Coenzyme A	11	3	1
3HC	3-Hydroxybutanoyl-Coenzyme A	9	5	1
S0N	O-Succinylbenzoyl-N-Coenzyme A	8	3	1
BCA	4-Hydroxybenzoyl Coenzyme A	7	6	1
CMX	Carboxymethyldethia Coenzyme *A	7	4	0
COZ	Coenzyme A	6	3	1
GRA	Glutaryl-Coenzyme A	6	3	1
IVC	Isovaleryl-Coenzyme A	4	3	1
SFC	(S)-Ibuprofenoyl-Coenzyme A	4	1	1
HDC	3r-Hydroxydecanoyl-Coenzyme A	3	1	1
01A	4-Chlorophenacyl-Coenzyme A	2	1	0
OXK	Oxalyl-Coenzyme A	2	1	1
FRE	Feruloyl Coenzyme A	1	1	1
MDE	5-Mercaptoethanol-2-Decenoyl-Coenzyme A	1	1	1
TH3	Trans Delta2 Palmitenoyl-Coenzyme A	1	1	1
MCA	Methylmalonyl-Coenzyme A	18	7	1

COS	Coenzyme A Persulfide	15	7	0
CO6	Isobutyryl-Coenzyme A	10	2	0
UOQ	Undeca-2-One Coenzyme A	8	0	0
HFQ	2,4-Dihydroxyphenacyl Coenzyme A	6	1	1
3CP	3-Carboxypropyl-Coenzyme A	4	0	0
CS8	3-Thiooctanoyl-Coenzyme A	4	1	0
MCD	Methylmalonyl(Carbadethia)-Coenzyme A	4	0	0
MRR	(R)-2-Methylmyristoyl-Coenzyme A	4	0	0
MRS	(S)-2-Methylmyristoyl-Coenzyme A	4	0	0
RFC	(R)-Ibuprofenoyl-Coenzyme A	4	0	0
SCD	Succinyl(Carbadethia)-Coenzyme A	4	1	0
SDX	S-Citryldethia Coenzyme A	4	0	0
4CA	4-Hydroxybenzyl Coenzyme A	3	0	0
2MC	Methacrylyl-Coenzyme A	2	1	0
52O	Pivalyl-Coenzyme A	2	1	0
COF	Trifluoroacetyl Coenzyme A	2	1	0
COW	Anthraniloyl-Coenzyme A	2	2	0
FAQ	Phenylacetyl Coenzyme A	2	1	0
AMX	Amidocarboxymethylidethia Coenzyme *A	1	0	0
CIC	Citryl-Thioether-Coenzyme *A	1	0	0
ETB	Ethyl Coenzyme A	1	1	0
FAM	Alpha-Fluoro-Amidocarboxymethylidethia Coenzyme A	1	0	0
FCX	Alpha-Fluoro-Carboxymethylidethia Coenzyme A	1	0	0
HAX	N-Hydroxyamidocarboxymethylidethia Coenzyme *A	1	0	0
JBT	GDP-N-Acetylperosamine-Coenzyme A	1	0	0
NMX	Nitromethylidethia Coenzyme A	1	0	0
SPF	Sinapoyl Coenzyme A	1	0	0

**Table S2.** The  $p$ -values of the statistical tests for the MCC differences among different methods on the test set TE73. The '+'/'-' sign in the lower triangle indicates that the average MCC for the method from the corresponding row is higher/lower than the one from the corresponding column. The values on the upper triangle are the  $p$ -values. TemPred<sup>P</sup> represents the template-based method based on PSI-BLAST. TemPred\* represents the TemPred built with the original training data that do not transfer binding annotations from identical proteins.

Method	SVMpred	TemPred	CoABind	TemPred <sup>P</sup>	TemPred*	S-SITE
SVMpred	NA	$3.13 \times 10^{-5}$	$1.07 \times 10^{-6}$	$5.55 \times 10^{-7}$	$4.34 \times 10^{-5}$	$5.70 \times 10^{-5}$
TemPred	+	NA	$3.31 \times 10^{-3}$	$6.45 \times 10^{-8}$	$5.46 \times 10^{-2}$	$1.74 \times 10^{-5}$
CoABind	+	+	NA	$2.52 \times 10^{-8}$	$1.27 \times 10^{-5}$	$3.14 \times 10^{-6}$
TemPred <sup>P</sup>	-	-	-	NA	$1.03 \times 10^{-7}$	$1.06 \times 10^{-03}$
TemPred*	+	-	-	+	NA	$1.42 \times 10^{-5}$
S-SITE	-	-	-	+	-	NA

**Table S3.** The predictive quality of the proposed methods on the two subsets of independent test set.

Dataset	Method	MCC	Pre	Rec	AUC	AUC <sub>L</sub>	Ratio
TE41	SVMpred	0.555	0.62	0.577	0.871	0.058	11.03
	TemPred	0.605	<b>0.642</b>	0.660	0.873	0.05	9.598
	CoABind	<b>0.609</b>	0.617	<b>0.711</b>	<b>0.906</b>	<b>0.067</b>	<b>12.956</b>
TE32	SVMpred	0.287	<b>0.376</b>	0.293	0.749	0.034	6.439
	TemPred	0.32	0.313	0.376	0.692	0.026	4.979
	CoABind	<b>0.335</b>	0.373	<b>0.393</b>	<b>0.775</b>	<b>0.039</b>	<b>7.567</b>

**Table S4.** Comparison of the proposed methods with the template-based method S-SITE.

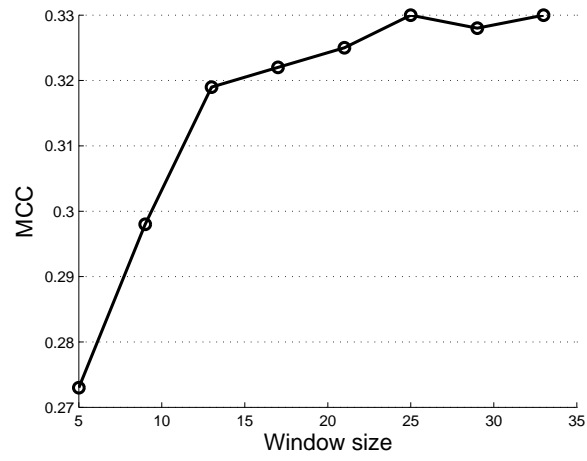
Method	MCC	Pre	Rec	AUC	AUC <sub>L</sub>	R
S-SITE	0.396	0.378	0.476	0.752	0.027	5.360
SVMpred	0.438	<b>0.513</b>	0.453	0.817	0.047	9.067
TemPred	0.481	0.498	0.535	0.794	0.039	7.573
CoABind	<b>0.489</b>	0.510	<b>0.571</b>	<b>0.849</b>	<b>0.055</b>	<b>10.516</b>

**Table S5.** Influence of the alignment algorithm and data design to TemPred. TemPred<sup>P</sup> represents the template-based method based on PSI-BLAST. TemPred\* represents the TemPred built with the original training data that do not transfer binding annotations from identical proteins. TemPred<sup>T</sup> represents the TemPred for dataset with CoA only data. The last column indicates the percentage (P) of proteins with homologous templates detected by PSI-BLAST and HHsearch (at  $e$ -value < 0.001).

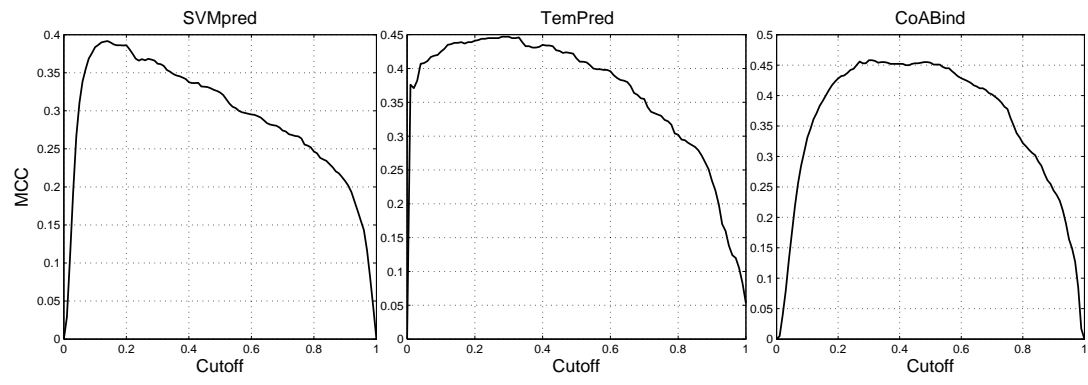
Method	MCC	Pre	Rec	AUC	AUC <sub>L</sub>	R	P(%)
TemPred	<b>0.481</b>	<b>0.498</b>	<b>0.535</b>	0.794	<b>0.039</b>	<b>7.573</b>	<b>82.2</b>
TemPred <sup>P</sup>	0.357	0.384	0.425	0.704	0.016	3.081	56.2
TemPred*	0.474	0.491	0.522	<b>0.795</b>	0.038	7.365	82.2
TemPred <sup>T</sup>	0.419	0.422	0.472	0.765	0.031	5.983	73.8

**Table S6.** The average accuracy of TemPred on the training and test sets from 10 different divisions of dataset.

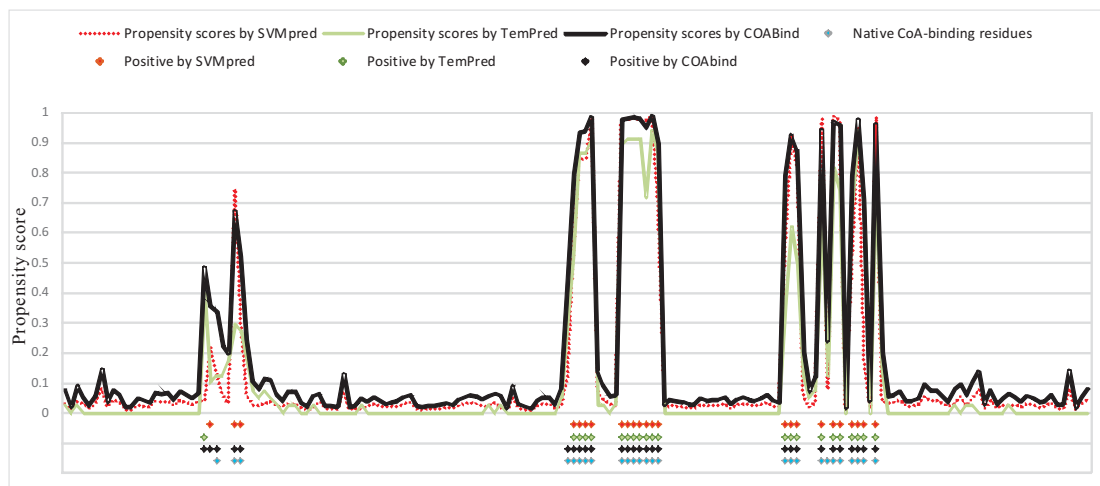
Dataset	MCC	Pre	Rec	AUC	AUC <sub>L</sub>	R
Training set	0.423	0.434	0.474	0.758	0.032	6.072
Test set	0.485	0.501	0.533	0.792	0.036	7.014



**Fig. S1.** The selection of the window size. The MCC values are obtained based on 5-fold cross validation on the training set.



**Fig. S2.** The selection of the probability cutoffs for the proposed methods based on 5-fold cross validation on the training set.



**Fig. S3.** An example illustrating the predictions by SVMpred, TemPred, and CoABind.