SUPPLEMENTARY DATA

Supplementary Table S1. The parameters for the detecting method of the first-level search of S2L-based methods. The number of iteration is set as 1 according to the requirements of SMI-BLAST framework.

Basic methods	**commands**
PSI-BLAST	ncbi-blast-2.7.1+/bin/psiblast -query $protein.fasta -db scope-95-2.06 -out ./$protein.txt -outfmt 5 -num_iterations 1
DELTA-BLAST	Version: ncbi-blast-2.10.1+ 1.1 Produce the first iteration results of DELTA-BLAST on benchmark dataset : deltablast -query $query_name.fasta -db scope-95-2.06 -rpsdb cdd_delta (default parameters) -outfmt 5 -num_iterations 1 -out $query_name.txt
PSI-BLASTexB based	Version: ncbi-blast-2.5.0+psiblastexb.1.0_linux64 1.1 Produce PSSM: (default parameters) psiblast -query $query_name.fasta -db uniref50 -num_alignments 1000000 -out_pssm $query_name.pssm -save_pssm_after_last_round -out $query_name.txt -outfmt 6 1.2 Produce the first iteration results of PSIBLASTexB on benchmark dataset: psiblast -in_pssm $query_name.pssm -db scope-95-2.06 –out $query_name.txt -outfmt 5 -evalue 10 -num_iterations 1

Supplementary Table S2. The parameters for the Filter 1 (SMI-BLAST) of the first-level search of S2L-based methods.

Basic methods	**commands**
PSI-BLAST	SMI-BLAST (default parameters) –iteration 5 -db scope-95-2.06 -cutoff 0
DELTA-BLAST	SMI-DELTABLAST (default parameters) –iteration 5 -db scope-95-2.06 -cutoff 0
PSI-BLASTexB based	SMI-PSIBLASTexB (default parameters) –iteration 5 -db scope-95-2.06 -cutoff 0

Supplementary Table S3. The parameters for the Filter 2 (double link strategy) of the first-level search of S2L-based methods.

Basic methods	**commands**
PSI-BLAST	Double-link of PSI-BLAST (default parameters of PL-search) -e-value of in-link: 0.001 -e-value of out-link: 0.001
DELTA-BLAST	Double-link of DELTA-BLAST (default parameters of PL-search) -e-value of in-link: 0.001 -e-value of out-link: 0.001
PSI-BLASTexB based	Double-link of PSI-BLASTexB (default parameters of PL-search) -e-value of in-link: 0.001 -e-value of out-link: 0.001

Supplementary Table S4. The parameters for the detecting method (PL-search) of the second-level search of S2L-based methods.

Basic methods	**Commands**
PSI-BLAST	PL-BLAST(default parameters of PL-search) $-\beta_1$ 0.95 $-\beta_2$ 1.50 -threshold 0.99 -db scope-95-2.06
DELTA-BLAST	PL-DELTABLAST(default parameters of PL-search)

	$-\beta_1$ 0.95 $-\beta_2$ 1.50 -threshold 0.99 -db scope-95-2.06
PSI-BLASTexB based	PL-PSIBLASTexB(default parameters of PL-search) $-\beta_1$ 0.95 $-\beta_2$ 1.50 -threshold 0.99 -db scope-95-2.06

Supplementary Table S5. The parameters for the similarity matrix of the second-level search of S2L-based methods.

1. Alignment score	
PSI-BLAST based	Version:ncbi-blast-2.7.1+ psiblast -query $protein.fasta -db scope-95-2.06 -out ./$protein.txt -outfmt 5 -num_iterations 5
DELTA-BLAST based	Version: ncbi-blast-2.10.1+ 1.2 Produce the first iteration results of DELTA-BLAST on benchmark dataset: deltablast -query $query_name.fasta -db scope-95-2.06 -rpsdb cdd_delta -outfmt 5 -num_iterations 5 -out $query_name.txt
PSI-BLASTexB based	Version: ncbi-blast-2.5.0+psiblastexb.1.0_linux64 1.3 Produce PSSM (default parameters): psiblast -query $query_name.fasta -db uniref50 -num_alignments 1000000 -out_pssm $query_name.pssm -save_pssm_after_last_round -out $query_name.txt -outfmt 6 1.4 Produce the first iteration results of PSIBLASTexB on benchmark dataset: psiblast -in_pssm $query_name.pssm -db scope-95-2.06 –out $query_name.txt -outfmt 5 -evalue 10 -num_iterations 5

2. First-level link similarity score	
Fscore	No paramters
3. Feature similarity score: Profile construction	
HMM profile construction	(default parameters of hhblits) hhblits -i $query_sequence -d uniprot20_2013_03 -ohhm $result_HMMprofile
PSSM profile construction	(default parameters of Pse-in-one) psiblast -query $query.fasta -db nrdb90 -out $xml_file -evalue 0.001 -num_iterations 10 -num_threads 5 -out_ascii_pssm $pssm_file -outfmt 5 (Those parameters are the default parameters of the source code of Pse-in-one2.0 (Liu, et al., 2017))
3. Feature similarity score: Feature extraction	
ACCPSSM	(default parameters of Pse-in-one) Pse-in-one2.0/acc.pyc $xxx.fasta Protein ACC -out $xxx.txt
DR	(default parameters of Pse-in-one) Pse-in-one2.0/nac.pyc $xxx.fasta Protein DR -out $xxx.txt
DTPSSM	(default parameters of Pse-in-one) Pse-in-one2.0/profile.pyc $xxx.fasta DT -out $xxx.txt
ACCHMM feature	Source code of ACCHMM feature can be accessed at http://bliulab.net/S2L-PSIBLAST/download/ (default parameters of original source code from Dong, et al) ACCHMM feature is extracted by autocross-covariance (ACC) transformation from the HMM profile. This code is implemented by re-writing the source code of ACCPSSM (http://www.iipl.fudan.edu.cn/demo/accpkg.html) (Dong, et al., 2009).
DTHMM feature	Source code of DTHMM feature can be accessed at http://bliulab.net/S2L-PSIBLAST/download/ (default parameters of original source code from Liu, et al) DTHMM feature is extracted by Distance-based Top-n-gram (DT) (Liu, et al., 2014) from HMM profile. This code is implemented by re-writing the source code of DTHMM obtained from Pse-in-one 2.0 (Liu, et al., 2017).

Supplementary Table S6. The parameters for the learning to rank model of the

second-level search of S2L-based methods.

Training command	(default parameters of RankLib (https://sourceforge.net/p/lemur/wiki/RankLib/)) **java -jar RankLib-2.10.jar -train $train_file -test $test_file -ranker 6 -metric2t NDCG@50 -norm linear -save**
Testing command	(default parameters of RankLib (https://sourceforge.net/p/lemur/wiki/RankLib/)) **java -jar RankLib-2.10.jar -load $model_file -rank $test_file -score $score_file -norm linear**

Supplementary Table S7. The parameters for PSI-BLAST, DELTA-BLAST and PSI-BLASTexB.

Basic methods	**commands**
PSI-BLAST	ncbi-blast-2.7.1+/bin/psiblast -query $protein.fasta -db scope-95-2.06 -out ./$protein.txt -outfmt 5 -num_iterations 5
DELTA-BLAST	Version: ncbi-blast-2.10.1+ 1.3 Produce the fifth iteration results of DELTA-BLAST on benchmark dataset : deltablast -query $query_name.fasta -db scope-95-2.06 -rpsdb cdd_delta (default parameters) -outfmt 5 -num_iterations 5 -out $query_name.txt
PSI-BLASTexB	Version: ncbi-blast-2.5.0+psiblastexb.1.0_linux64 1.5 Produce PSSM: (default parameters) psiblast -query $query_name.fasta -db uniref50 -num_alignments 1000000 -out_pssm $query_name.pssm -save_pssm_after_last_round -out $query_name.txt -outfmt 6 1.6 Produce the fifth iteration results of PSIBLASTexB on benchmark dataset: psiblast -in_pssm $query_name.pssm

	-db scope-95-2.06 –out $query_name.txt -outfmt 5 -evalue 10 -num_iterations 5

Supplementary Table S8. The relationship of first-level search results and second-level search results of S2L-PSIBLAST, S2L-DELTABLAST and S2L-PSIBLASTexB on benchmark dataset

Methods	1L [a]	2L [a]	1L & 2L [b]	1L-2L[c]	2L-1L[c]	1L and 2L[d]
S2L-PSIBLAST	25510	16965	14493	11017	2472	27982
SPL-DELTABLAST	26380	20267	18656	7724	1611	27991
SPL-PSIBLASTexB	26097	19039	17152	8945	1887	27984

[a] represents the number of query sequences from benchmark dataset can obtain result list by the first-level search or second-level search;
[b] represents the number of query sequences from benchmark dataset can obtain first-level results list and second-level result list;
c represents the number of query sequences from benchmark dataset can obtain first-level result list as final results or second-level result list as final result;
d represents the total number of query sequences from benchmark dataset can obtain result list by first-level search and second-level search.

Supplementary Table S9. Comparison of the performance of the detecting methods and filtered results in the first-level search of three S2L-PSIBLAST-based methods on benchmark dataset

Methods	ROC1	ROC50
PSI-BLAST [a]	0.8318	0.8896
DELTA-BLAST [a]	0.8906	0.9243
PSI-BLASTexB [a]	0.8311	0.9030
S2L-PSIBLAST [b]	0.9083	0.9087
S2L-DELTABLAST [b]	0.9349	0.9361
S2L-PSIBLASTexB [b]	0.9281	0.9288

[a] represents the performance of the detecting search methods in the first-level search.
[b] represents the performance of the first-level search of three S2L-PSIBLAST –based search methods

Supplementary Table S10. The usage frequencies of sequence similarity features in the LambdaMART model of three S2L-PSIBLAST-framework-based methods.

Usage frequency	S2L-PSIBLAST	S2L-DELTABLAST	S2L-PSIBLASTexB
E-value	995	189	395
Align length	265	343	290
Order information	471	494	334

J_score	1135	1057	1127
SPL-similarity	804	820	626
ACCPSSM-PC	843	1057	983
DT-ED	660	599	636
DT-MD	796	907	1305
DR-ED	571	764	844
ACCHMM-PC	925	834	765
HMMDT-ED	1002	1085	898
HMMDT-MD	533	851	797

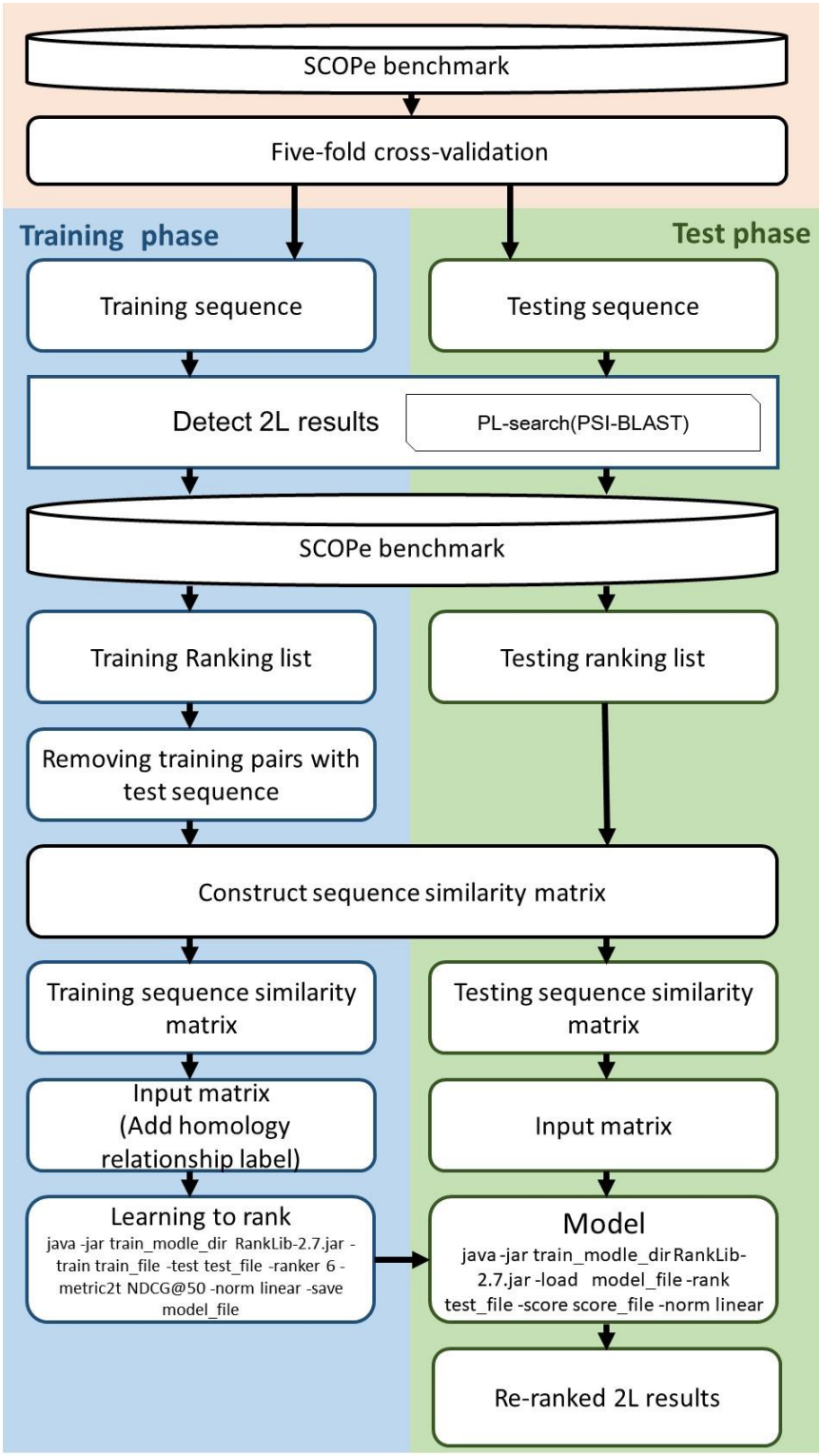
Supplementary Table S11. The components' settings of S2L-PSIBLAST on the new training set.

Components	**On new cross-validation set**
Detecte 1L results: PSI-BLAST	PSI-BLAST searches the new training set with 25209 protein sequences
Filter 1: SMI-BLAST	Re-trained on the new training set. Performance on new training set with Five-fold cross-validation: ROC1: 0.9145 ROC50: 0.9169
Filter 2: Double-link strategy	Re-constructed on the new training set.
Detected 2L results: PL-search	Detecting the protein sequences on the new cross-validation set.
Add similarity scores: 2.First-level link similarity score	The extend-link of the first level results and the Fscore are re-calculated according to the first level results on the new training set.

Supplementary Table S12. Comparison of the performance of JackHMMER, SMI-HMMER and S2L-JackHMMER on SCOPe benchmark dataset.

Method	**ROC1**	**ROC50**
JackHMMER	0.8919	0.9059
SMI-HMMER	0.8975	0.9138
S2L-JackHMMER[a]	0.9111	0.9217

[a] All parameters of S2L-JackHMMER are default value. The parameters of SMI-HMMER and PL-HMMER in S2L-JackHMMER are consistent with previous studies (Jin, et al., 2021; Jin, et al., 2021).



Supplementary Fig. S1. The flowchart of training and test processes of learning-to-rank models.

References

Dong, Q., Zhou, S. and Guan, J. A new taxonomy-based protein fold recognition

approach based on autocross-covariance transformation. *Bioinformatics* 2009;25(20):2655-2662.

Jin, X., Liao, Q. and Liu, B. PL-search: a profile-link-based search method for protein remote homology detection. *Brief Bioinform* 2021;22(3).

Jin, X.*, et al.* SMI-BLAST: a novel supervised search framework based on PSI-BLAST for protein remote homology detection. *Bioinformatics* 2021;37(7):913-920.

Liu, B., Wu, H. and Chou, K. Pse-in-One 2.0: An Improved Package of Web Servers for Generating Various Modes of Pseudo Components of DNA, RNA, and Protein Sequences. *Natural Science* 2017;09(04):67-91.

Liu, B.*, et al.* Using distances between Top-n-gram and residue pairs for protein remote homology detection. *Bmc Bioinformatics* 2014;15.