Supplementary Data

Table S1. Summary of methods for residue-residue contact prediction sorted according to the year of publication. The citation data are taken from Google Scholar (https://scholar.google.com).

Method	Reference (year)	Citation	Type	Input	Prediction model/Remark	URL of Web server	URL of standalone package
Gobel et al	[1] (1994)	662	ML	SEQ	Linear correlation between CM and contacts	NA	NA
Shindyalov et al	[2] (1994)	239	ML	SEQ	Phylogenetic tree	NA	NA
Olmea et al	[3] (1997)	227	ML	SEQ	Linear combination of CM and other features	NA	NA
CORNET	[4] (1999)	176	ML	SEQ	NN	NA	NA
Fariselli	[5] (2001)	207	ML	SEQ	NN and CM	NA	NA
CMAPpro	[6] (2002)	164	ML	SEQ	HMM and 2D-Recursive NN	http://scratch.proteomics.ics.uci.edu/	NA
HMMSTR-CM	[7](2003)	79	ML	SEQ	HMM	NA	NA
GPCPRED	[8](2004)	86	ML	SEQ	Genetic programming	http://www.sbc.su.se/~maccallr/contactmaps/	NA
Karypis et al	[9] (2005)	50	ML	SEQ	SVM	NA	NA
BETAcon	[10] (2005)	106	ML	SEQ	2D-Recursive NN	NA	http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html
PROFcon	[11] (2005)	145	ML	SEQ	NN	NA	NA
XX-STOUT	[12] (2006)	88	ML	SEQ	bidirectional recurrent NN	http://distill.ucd.ie/xxstout/	NA
SVMcon	[13] (2007)	167	ML	SEQ	SVM	http://scratch.proteomics.ics.uci.edu/	http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html
SVMSEQ	[14] (2008)	91	ML	SEQ	SVM	http://zhanglab.ccmb.med.umich.edu/SVMSEQ/	http://zhanglab.ccmb.med.umich.edu/SVMSEQ/download.html
NNcon	[15] (2009)	96	ML	SEQ	2D-Recursive NN	NA	http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html
mpDCA	[16] (2009)	379	DC	SEQ	Message-passing algorithm	NA	NA
SPINE-2D	[17] (2009)	35	ML	SEQ	NN	http://sparks-lab.org/html_old/SPINE-2D/spine-2d.html	NA

PSICOV	[18] (2011)	224	DC	MSA	Graphical Lasso for sparse inverse covariance estimation	NA	http://bioinfadmin.cs.ucl.ac.uk/downloads/PSICOV/
ProC_S3	[19] (2011)	24	ML	SEQ	Random forest	NA	NA
mfDCA	[20] (2011)	304	DC	MSA	Mean field, HMM	http://dca.rice.edu/portal/dca/	http://dca.rice.edu/portal/dca/download
DNcon	[21] (2012)	41	ML	SEQ	Restricted Boltzmann machine and deep belief network	http://iris.rnet.missouri.edu/dncon/	http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html
PSpro	[22] (2012)	21	ML	SEQ	1D-recursive NN	NA	http://sysbio.rnet.missouri.edu/multicom_toolbox/tools.html
GREMLIN	[23] (2013)	82	DC	MSA	Pseudo likelihood estimation	http://gremlin.bakerlab.org/submit.php	http://gremlin.bakerlab.org/gremlin.php
plmDCA	[24] (2013)	116	DC	MSA	Pseudo likelihood estimation	NA	http://plmdca.csc.kth.se/
PhyCMAP	[25] (2013)	37	ML	SEQ	Random forest and integer linear programming	NA	NA
EPC-map	[26] (2014)	10	CB	SEQ	GREMLIN and SVM based training	http://compbio.robotics.tu-berlin.de/epc-map/	NA
CCMpred	[27] (2014)	23	DC	MSA	Markov random field pseudo-likelihood maximization	NA	https://github.com/soedinglab/CCMpred
FreeContact	[28] (2014)	25	DC	MSA	Implement of mfDCA and PSICOV	NA	https://rostlab.org/owiki/index.php/FreeContact
MetaPSICOV	[29] (2014)	25	CB	SEQ	Two stage feed-forward NN	http://bioinf.cs.ucl.ac.uk/MetaPSICOV/	http://bioinfadmin.cs.ucl.ac.uk/downloads/MetaPSICOV/
PconsC2	[30] (2014)	24	CB	SEQ	Deep Learning, a multilayer feed-forward stack of random forest learners	http://pconsc3.bioinfo.se/pred/	https://github.com/ElofssonLab/PconsC2
CoinDCA	[31] (2015)	6	CB	SEQ	Evolutionary coupling analysis and random forest for combination	http://raptorx.uchicago.edu/ContactMap/	NA
bbcontacts	[32] (2015)	3	ML	SM	HMM	NA	https://github.com/soedinglab/bbcontacts
BND	[33] (2015)	5	DC	SM	Balanced network deconvolution	http://www.csbio.sjtu.edu.cn/bioinf/BND/	http://www.csbio.sjtu.edu.cn/bioinf/BND/
R_2C	[34] (2016)	0	CB	SEQ	Query-driven dynamic fusion and Gaussian Noise filter	http://www.csbio.sjtu.edu.cn/bioinf/R2C/	NA

ML: machine learning based method; DC: direct coupling based method. CB: consensus based method; SEQ: amino acid sequence; CM: correlated mutation; SM: symmetric matrix; MSA: Multiple sequence alignment; NN: Neural network; SVM: support vector machine; HMM: hidden Markov model; NA: not available.

Table S2. The negative-to-positive ratio (NPR) of residue pairs for proteins in the benchmark dataset.

	Ranges			
Structural classes	Short-range	Medium-range	Long-range	All
α	29.4	73.6	105.7	65.4
β	15.6	22.0	50.9	31.7
α+β	18.7	30.0	57.3	47.3

Table S3. The precision (%) of the top *L* **(A)**, *L*/2 **(B)** and *L*/10 **(C)** predicted contacts by 15 predictors evaluated for different ranges, structural classes and target types.

(A)

	Ranges			Structural classes			Target types		
Methods	short-range	medium-range	long-range	α	β	α+β	easy	medium	hard
DNcon	9.01	13.92	13.45	17.44	30.04	29.73	29.70	26.38	24.36
bbcontacts	2.50	2.84	6.08	0.75	18.56	12.33	13.79	10.01	8.54
PSpro.beta	18.23	10.38	13.52	13.98	26.77	25.41	24.57	22.99	22.02
PSpro	10.48	7.67	7.37	7.45	31.44	23.21	22.16	22.12	21.48
NNcon	19.40	15.94	13.42	15.98	32.90	27.13	25.64	26.41	26.94
SVMcon	21.26	19.39	16.92	21.15	34.89	32.43	30.73	31.46	30.45
BETAcon	21.03	17.87	16.52	17.30	38.60	33.26	31.67	32.14	30.35
SVMSEQ	20.72	18.97	18.54	20.58	35.81	34.71	33.93	31.60	29.82
FreeContact	7.57	6.74	8.09	6.92	10.69	9.84	11.27	8.92	6.99
PSICOV	12.44	13.33	22.42	19.70	29.67	32.83	39.73	23.88	16.96
plmDCA	12.76	13.95	23.98	22.03	32.32	35.02	40.94	28.09	20.06
GREMLIN	13.54	15.10	27.73	23.58	35.26	40.57	48.00	29.60	21.19
CCMpred	13.58	15.09	27.65	23.62	35.19	40.38	47.68	29.73	21.26
PconsC2	20.41	24.68	42.85	33.98	55.29	57.99	66.16	47.50	34.91
MetaPSICOV	24.18	26.23	41.02	37.03	61.52	61.01	65.70	52.66	44.24

(B)

	Ranges			Structural classes			Target types		
Methods	short-range	medium-range	long-range	α	β	α+β	easy	medium	hard
DNcon	10.22	25.38	23.43	22.34	36.81	38.27	37.91	33.37	30.83
bbcontacts	5.02	5.71	12.19	1.50	37.18	24.75	27.63	20.13	17.14
PSpro.beta	26.22	13.51	16.77	19.44	34.78	33.70	32.48	30.63	29.34
PSpro	20.86	15.02	11.18	13.69	41.51	34.13	31.74	33.13	31.09
NNcon	27.72	21.78	16.64	20.85	41.83	35.39	33.43	35.04	34.21
SVMcon	30.64	26.28	21.16	27.03	41.17	40.57	37.79	39.51	37.72
BETAcon	31.23	25.10	21.50	23.53	48.14	43.60	41.05	42.51	39.21
SVMSEQ	32.89	26.78	23.55	27.21	43.98	45.42	43.70	41.28	38.37
FreeContact	8.43	8.23	10.12	8.12	12.39	11.56	12.79	10.90	8.58
PSICOV	17.40	19.64	31.65	28.29	39.52	43.88	53.21	33.55	22.79
plmDCA	17.96	20.67	34.98	31.29	43.84	47.82	56.26	39.40	26.91
GREMLIN	19.75	23.07	38.86	32.92	47.23	52.03	61.83	40.61	28.43
CCMpred	19.77	23.06	38.77	32.84	47.32	52.16	61.74	40.73	28.75
PconsC2	31.57	37.42	53.95	43.40	64.14	67.25	76.26	57.09	42.60
MetaPSICOV	38.48	39.87	53.76	47.94	73.08	73.85	78.39	64.94	55.61

(C)

	Ranges			Structural classes			Target types		
Methods	short-range	medium-range	long-range	α	β	α+β	easy	medium	hard
DNcon	12.84	44.95	37.23	30.97	50.91	55.61	53.79	49.76	43.63
bbcontacts	24.48	26.42	47.29	7.23	73.81	75.81	74.73	56.08	45.31
PSpro.beta	46.21	22.30	23.80	37.63	53.24	51.06	50.61	48.08	46.71
PSpro	47.54	33.48	20.83	36.06	56.88	51.39	49.49	50.84	49.15
NNcon	47.80	35.45	24.87	37.45	56.95	51.48	49.72	51.29	49.60
SVMcon	49.55	39.98	29.23	41.26	53.73	54.55	50.85	53.59	52.25
BETAcon	55.33	42.07	31.60	44.53	66.29	61.39	59.94	61.71	56.80
SVMSEQ	58.36	44.58	33.61	46.57	58.60	63.79	61.77	58.20	55.96
FreeContact	11.35	12.59	16.66	11.62	14.90	16.32	17.58	15.40	11.23
PSICOV	36.85	41.00	52.42	44.29	59.30	62.53	72.76	53.10	38.90
plmDCA	39.57	44.75	56.40	47.37	62.93	65.24	74.67	57.34	42.91
GREMLIN	43.71	48.43	58.02	49.06	65.61	67.19	76.98	58.09	45.22
CCMpred	43.36	48.46	58.28	48.76	65.76	67.24	76.93	58.45	45.13
PconsC2	58.45	62.35	69.55	58.92	76.29	79.47	88.27	71.84	55.85
MetaPSICOV	70.20	68.08	72.66	67.40	85.18	87.83	91.35	79.54	72.78

Table S4. Wilcoxon signed-rank test between the 15 considered contact predictors on the top *L* (A), *L/2* (B) and *L/10* (C) predicted contacts. In this test, the short- and medium-range contacts are combined and the results are presented at the upper triangle. The data for long-range contacts are listed at the lower triangle. The plus/minus means that a predictor at the row is significantly better/worse (at *p*-value<0.05). The equal sign indicates there is no significant differences (at *p*-value=0.05).

(A)

Methods	DNcon	bbcontacts	PSpro.beta	PSpro	NNcon	SVMcon	BETAcon	SVMSEQ	FreeContact	PSICOV	plmDCA	GREMLIN	CCMpred	PconsC2	MetaPSICOV
DNcon		+	-	+	-	-	-	-	+	=	-	-	-	-	-
bbcontacts	-		-	-	-	-	-	-	-	-	-	-	-	-	-
PSpro.beta	=	+		+	-	-	-	-	+	+	+	-	-	-	-
PSpro	-	+	-		-	-	-	-	+	-	-	-	-	-	-
NNcon	=	+	=	+		-	-	-	+	+	+	+	+	-	-
SVMcon	+	+	+	+	+		+	-	+	+	+	+	+	-	-
BETAcon	+	+	+	+	+	-		-	+	+	+	+	+	-	-
SVMSEQ	+	+	+	+	+	+	+		+	+	+	+	+	-	-
FreeContact	-	+	-	+	-	-	-	-		-	-	-	-	-	-
PSICOV	+	+	+	+	+	+	+	+	+		-	-	-	-	-
plmDCA	+	+	+	+	+	+	+	+	+	+		-	-	-	-
GREMLIN	+	+	+	+	+	+	+	+	+	+	+		=	-	-
CCMpred	+	+	+	+	+	+	+	+	+	+	+	-		-	-
PconsC2	+	+	+	+	+	+	+	+	+	+	+	+	+		-
MetaPSICOV	+	+	+	+	+	+	+	+	+	+	+	+	+	-	

(B)

Methods	DNcon	bbcontacts	PSpro.beta	PSpro	NNcon	SVMcon	BETAcon	SVMSEQ	FreeContact	PSICOV	plmDCA	GREMLIN	CCMpred	PconsC2	MetaPSICOV
DNcon		+	-	-	-	-	-	-	+	=	-	-	-	-	-
bbcontacts	-		-	-	-	-	-	-	=	-	-	-	-	-	-
PSpro.beta	-	+		-	-	-	-	-	+	+	=	-	-	-	-
PSpro	-	-	-		-	-	-	-	+	+	=	-	-	-	-
NNcon	-	+	=	+		-	-	-	+	+	+	=	=	-	-
SVMcon	-	+	+	+	+		-	-	+	+	+	+	+	-	-
BETAcon	-	+	+	+	+	=		-	+	+	+	+	+	-	-
SVMSEQ	=	+	+	+	+	+	+		+	+	+	+	+	-	-
FreeContact	-	-	-	-	-	-	-	-		-	-	-	-	-	-
PSICOV	+	+	+	+	+	+	+	+	+		-	-	-	-	-
plmDCA	+	+	+	+	+	+	+	+	+	+		-	-	-	-
GREMLIN	+	+	+	+	+	+	+	+	+	+	+		=	-	-
CCMpred	+	+	+	+	+	+	+	+	+	+	+	-		-	-
PconsC2	+	+	+	+	+	+	+	+	+	+	+	+	+		-
MetaPSICOV	+	+	+	+	+	+	+	+	+	+	+	+	+	=	

(C)

Methods	DNcon	bbcontacts	PSpro.beta	PSpro	NNcon	SVMcon	BETAcon	SVMSEQ	FreeContact	PSICOV	plmDCA	GREMLIN	CCMpred	PconsC2	MetaPSICOV
DNcon		+	-	-	-	-	-	-	+	-	-	-	-	-	-
bbcontacts	+		-	-	-	-	-	-	+	-	-	-	-	-	-
PSpro.beta	-	-		-	-	-	-	-	+	-	-	-	-	-	-
PSpro	-	-	-		-	-	-	-	+	=	-	-	-	-	-
NNcon	-	-	=	+		-	-	-	+	=	-	-	-	-	-
SVMcon	-	-	+	+	+		-	-	+	=	=	-	-	-	-
BETAcon	-	-	+	+	+	+		-	+	+	+	=	=	-	-
SVMSEQ	-	-	+	+	+	+	+		+	+	+	+	+	-	-
FreeContact	-	-	-	-	-	-	-	-		-	-	-	-	-	-
PSICOV	+	+	+	+	+	+	+	+	+		-	-	-	-	-
plmDCA	+	+	+	+	+	+	+	+	+	+		-	-	-	-
GREMLIN	+	+	+	+	+	+	+	+	+	+	+		=	-	-
CCMpred	+	+	+	+	+	+	+	+	+	+	+	+		-	-
PconsC2	+	+	+	+	+	+	+	+	+	+	+	+	+		-
MetaPSICOV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	

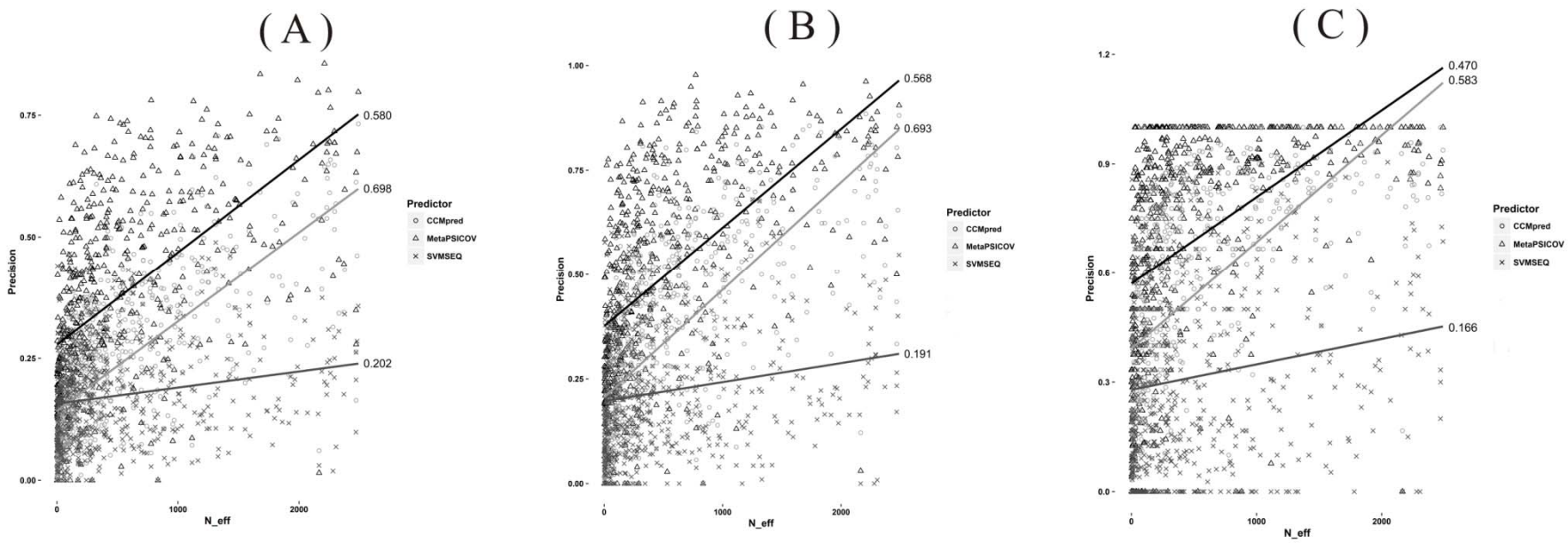

Figure S1. Precision of the top *L* (A), *L/2* (B) and *L/10* (C) long-range contacts as a function of the alignment depth. Three representative methods are used, SVMSEQ for machine learning based methods, CCMpred for direct coupling based methods, and MetaPSICOV for consensus based methods. The lines are the linear fits of the corresponding data.

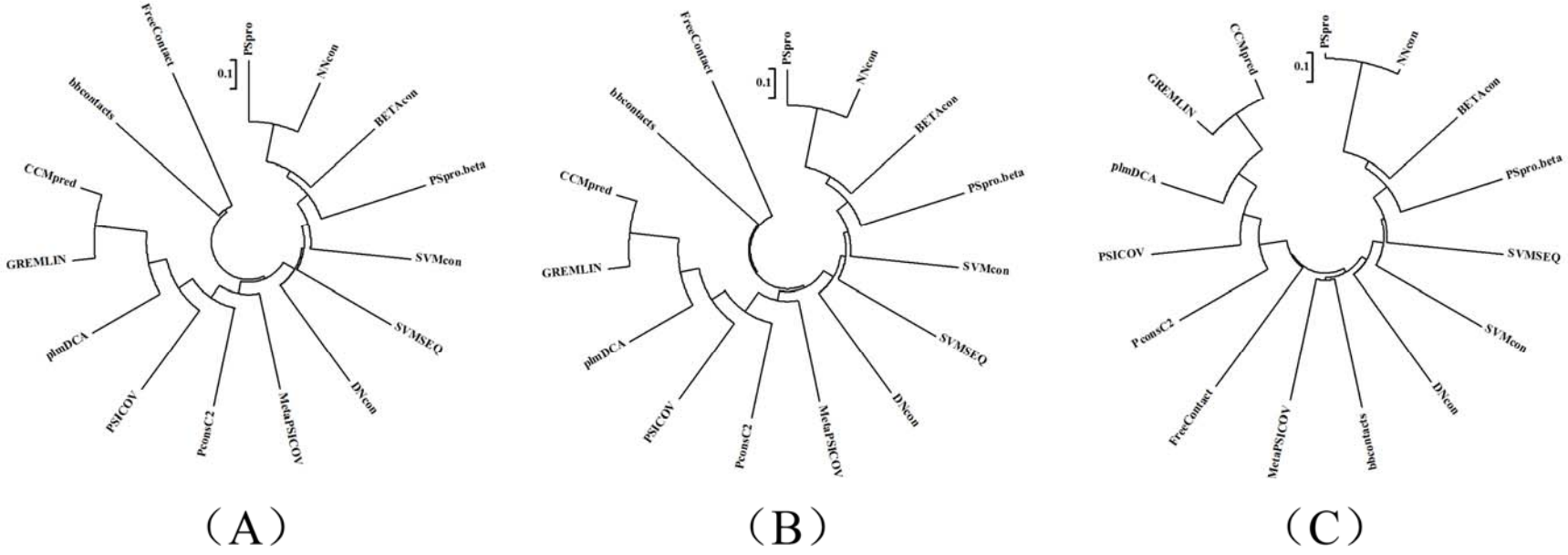

Figure S2. Neighbor joining dendrogram illustrating the relationship between different predictors using the top *L* (A), *L/2* (B) and *L/10* (C) predicted contacts.

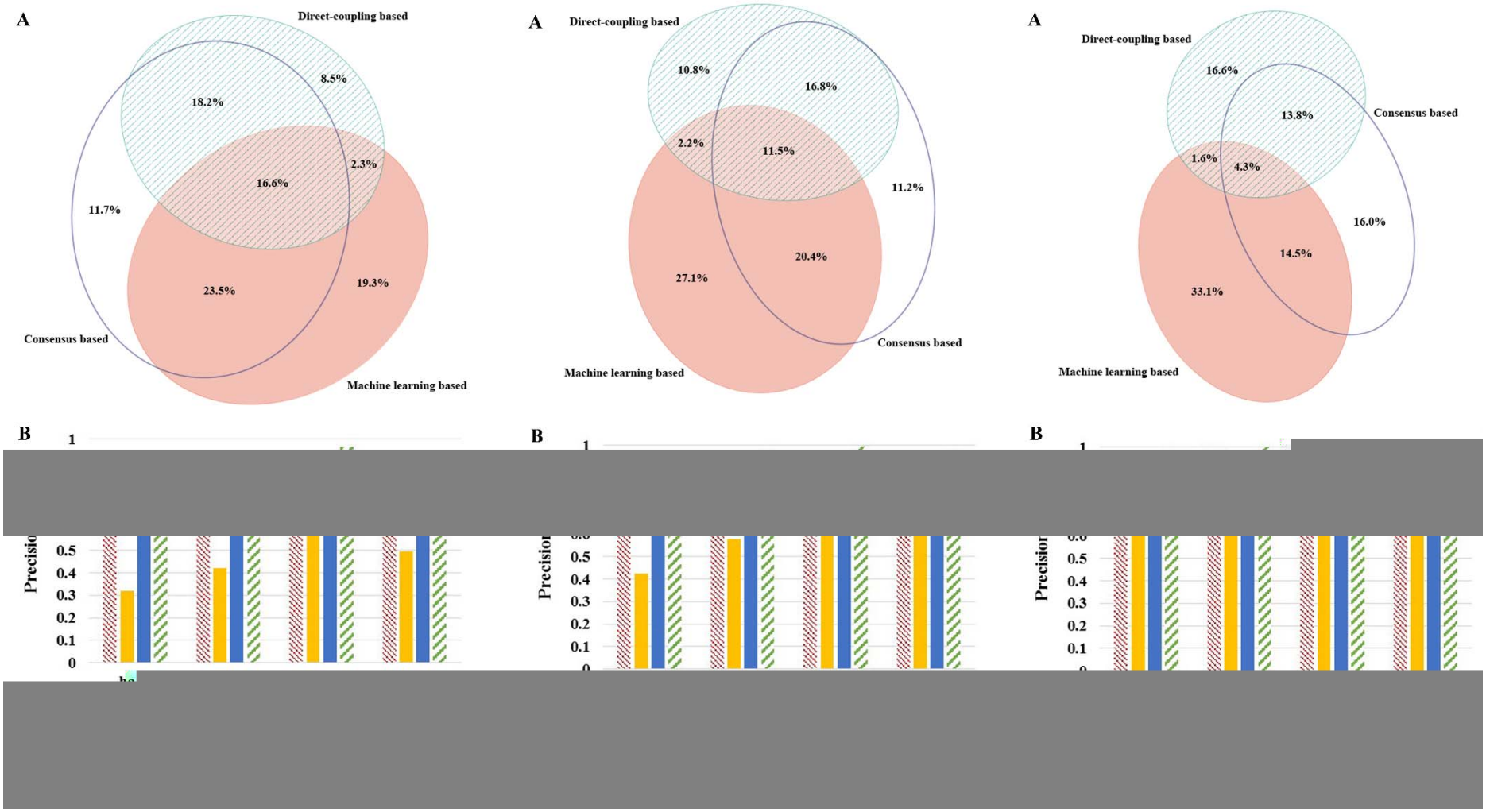

Figure S3. The upper limit of contact prediction by combing the top *L* (I), *L/2* (II) and *L/10* (III) predicted contacts. (A) The Venn

diagram of machine learning based, direct-coupling based and consensus based methods. (B) The upper limit of contact prediction for easy, medium and hard targets.

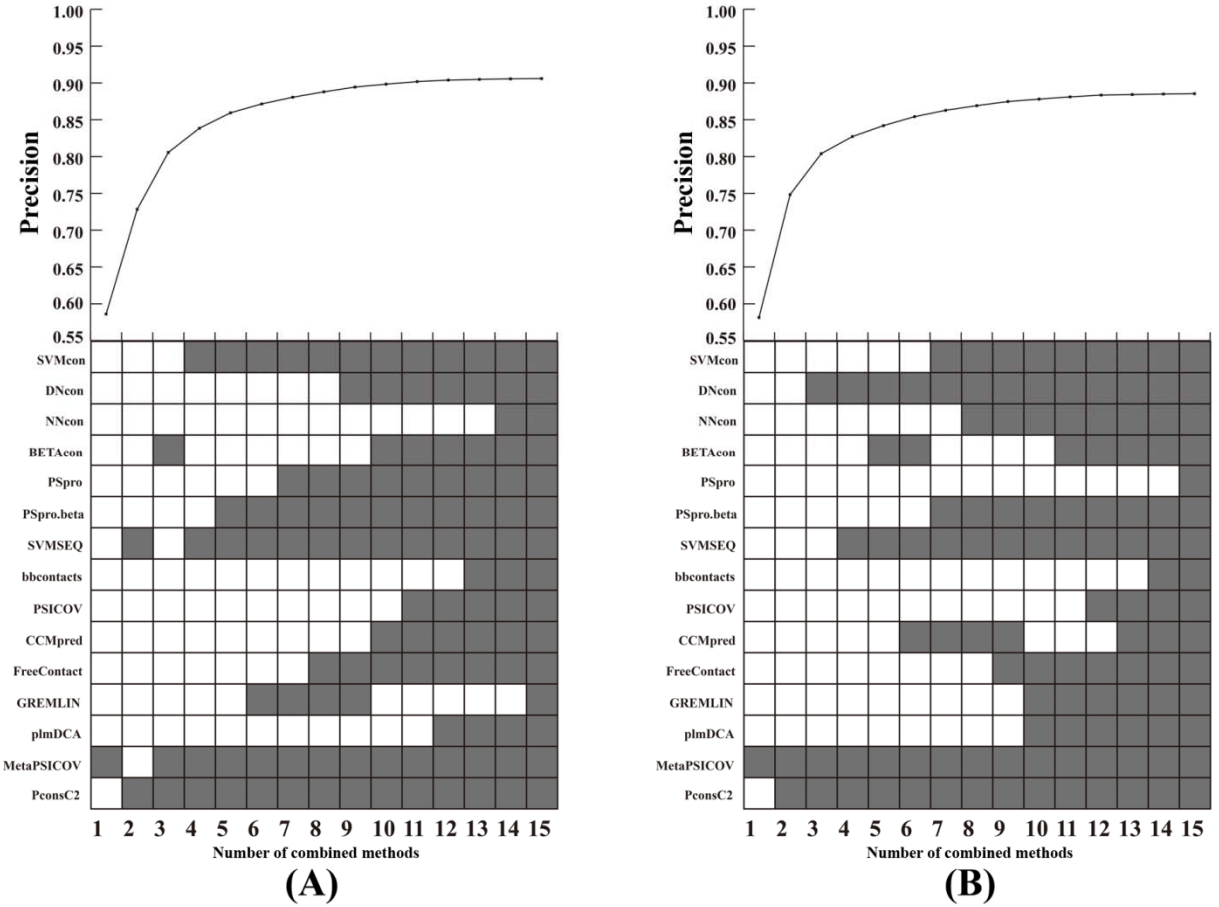

Figure S4. The upper limit of precision for meta-predictors on the top L/5 (A) short-range and (B) long-range contacts. For each number, selected predictors are indicated by filled squares.

References

1. Göbel U, Sander C, Schneider R et al. Correlated mutations and residue contacts in proteins, Proteins: Structure, Function, and Bioinformatics 1994;18:309-317.
2. Shindyalov IN, Kolchanov NA, Sander C. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations?, Protein Engineering 1994;7:349-358.
3. Olmea O, Valencia A. Improving contact predictions by the combination of correlated mutations and other sources of sequence information, Folding and Design 1997;2, Supplement 1:S25-S32.
4. Fariselli P, Casadio R. A neural network based predictor of residue contacts in proteins, Protein Engineering 1999;12:15-21.
5. Fariselli P, Olmea O, Valencia A et al. Prediction of contact maps with neural networks and correlated mutations, Protein Engineering 2001;14:835-843.
6. Pollastri G, Baldi P. Prediction of contact maps by GIOHMMs and recurrent neural networks using lateral propagation from all four cardinal corners, Bioinformatics 2002;18 Suppl 1:S62-70.
7. Shao Y, Bystroff C. Predicting interresidue contacts using templates and pathways, Proteins: Structure, Function, and Bioinformatics 2003;53:497-502.
8. MacCallum RM. Striped sheets and protein contact prediction, Bioinformatics 2004;20:i224-i231.
9. Zhao Y, Karypis G. PREDICTION OF CONTACT MAPS USING SUPPORT VECTOR MACHINES, International Journal on Artificial Intelligence Tools 2005;14:849-865.
10. Cheng J, Baldi P. Three-stage prediction of protein β-sheets by neural networks, alignments and graph algorithms, Bioinformatics 2005;21:i75-i84.
11. Punta M, Rost B. PROFcon: novel prediction of long-range contacts, Bioinformatics 2005;21:2960-2968.
12. Vullo A, Walsh I, Pollastri G. A two-stage approach for improved prediction of residue contact maps, BMC Bioinformatics 2006;7:1-12.
13. Cheng J, Baldi P. Improved residue contact prediction using support vector machines and a large feature set, BMC Bioinformatics 2007;8:1-9.
14. Wu S, Zhang Y. A comprehensive assessment of sequence-based and template-based methods for protein contact prediction, Bioinformatics 2008;24:924-931.
15. Tegge AN, Wang Z, Eickholt J et al. NNcon: improved protein contact map prediction using 2D-recursive neural networks, Nucleic Acids Research 2009;37:W515-W518.

16. Martin Weigt RAW, Hendrik Szurmant, James A. Hoch, and Terence Hwa,. Identification of direct residue contacts in protein–protein interaction by message passing, PNAS 2009;106:67-72.
17. Xue B, Faraggi E, Zhou Y. Predicting residue–residue contact maps by a two-layer, integrated neural-network method, Proteins: Structure, Function, and Bioinformatics 2009;76:176-183.
18. Jones DT, Buchan DWA, Cozzetto D et al. PSICOV: Precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments, Bioinformatics 2011.
19. Li Y, Fang Y, Fang J. Predicting residue–residue contacts using random forest models, Bioinformatics 2011;27:3379-3384.
20. F. Morcos AP, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, . Direct-coupling analysis of residue coevolution captures native contacts across many protein families, Proc Natl Acad Sci U S A 2011;108:1293-1301.
21. Eickholt J, Cheng J. Predicting protein residue-residue contacts using deep networks and boosting, Bioinformatics 2012.
22. Cheng J, Li J, Wang Z et al. The MULTICOM toolbox for protein structure prediction, BMC Bioinformatics 2012;13:1-12.
23. Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era, Proc Natl Acad Sci U S A 2013;110:15674-15679.
24. Magnus E, Cecilia L, Yueheng L et al. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models, Phys Rev E Stat Nonlin Soft Matter Phys 2013;87.
25. Wang Z, Xu J. Predicting protein contact map using evolutionary and physical constraints by integer programming, Bioinformatics 2013;29:i266-i273.
26. Schneider M, Brock O. Combining Physicochemical and Evolutionary Information for Protein Contact Prediction, PLoS ONE 2014;9:e108438.
27. Seemayer S, Gruber M, Söding J. CCMpred—fast and precise prediction of protein residue–residue contacts from correlated mutations, Bioinformatics 2014.
28. Kaján L, Hopf TA, Kalaš M et al. FreeContact: fast and free software for protein contact prediction from residue co-evolution, BMC Bioinformatics 2014;15:1-6.
29. Jones DT, Singh T, Kosciolek T et al. MetaPSICOV: Combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins, Bioinformatics 2014.
30. Skwark MJ, Raimondi D, Michel M et al. Improved Contact Predictions Using the Recognition of Protein Like Contact Patterns, PLoS

Comput Biol 2014;10:e1003889.
31. Ma J, Wang S, Wang Z et al. Protein Contact Prediction by Integrating Joint Evolutionary Coupling Analysis and Supervised Learning, Bioinformatics 2015.
32. Andreani J, Söding J. bbcontacts: prediction of β-strand pairing from direct coupling patterns, Bioinformatics 2015.
33. Sun H-P, Huang Y, Wang X-F et al. Improving accuracy of protein contact prediction using balanced network deconvolution, Proteins: Structure, Function, and Bioinformatics 2015;83:485-496.
34. Yang J, Jin Q-Y, Zhang B et al. R2C: improving ab initio residue contact map prediction using dynamic fusion strategy and Gaussian noise filter, Bioinformatics 2016.