Supplementary Document

DNCON2: Improved protein contact prediction using two-level deep convolutional neural networks

Badri Adhikari[1], Jie Hou[1], and Jianlin Cheng[1,2*]

[1]Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, Missouri, 65211, USA. [2]Informatics Institute, University of Missouri, Columbia, Missouri, 65211, USA.

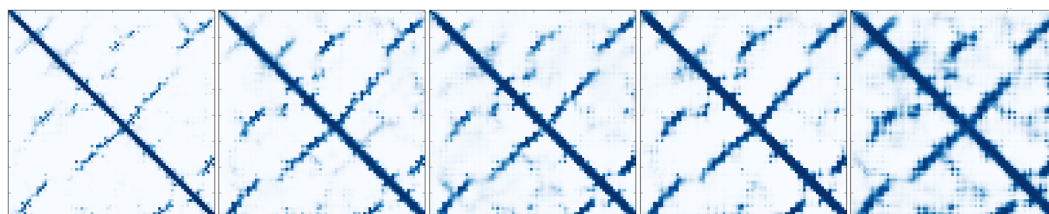*To whom correspondence should be addressed.

Figure S1. Contact maps predicted for chain A of the protein '1BDO' at distance thresholds 6, 7.5, 8, 8.5, and 10 Å (from left to right), showing the increase in the density/number of contacts as distance threshold is increased.

Table S2. List of features used in DNCON2. The first 20 features, excluding #18 and #19 (PSIPRED and PSISOLV predictions), are original DNCON features. Features numbered 21 through 27 (Shannon entropy through Mutual information) are coevolution-based features added in DNCON2.

| SN	Feature	Number of Dimensions	Number of Features	Number of 2D Features
1	Log of sequence length	Scalar	1	1
2	Log of number of sequences in the alignment	Scalar	1	1
3	Log of number of effective number of sequences in the alignment	Scalar	1	1
4	Ratio of number of 'buried' residues and length of the protein	1D	1	1
5	Ratio of number of 'strand' residues and length of the protein	1D	1	1
6	Ratio of number of 'helical' residues and length of the protein	1D	1	1
7	Atchley factors	1D	5	10
8	Binary predictions for helix, coild, and strand residues (SCRATCH)	1D	3	6
9	Solvent accessibility - 'buried' flag (SCRATCH)	1D	1	2
10	PSSM inf	1D	1	2
11	PSSM Sums (divided by 100)	1D	1	2
12	PSSM sum cosines	1D	1	2
13	Ratio of sequence separation and length of protein	2D	1	1
14	Flag for sequence separation between 23 and 28	2D	1	1
15	Flag for sequence separation between 28 and 38	2D	1	1
16	Flag for sequence separation between 38 and 48	2D	1	1
17	Flag for sequence separation 48+	2D	1	1
18	Probabilities of PSIPRED predictions for helix, coil, and strand residues	1D	3	6
19	Probabilities of PSISOLV predictions for solvent accessibility	1D	1	2
20	Pre-computed statistical potentials (Pref score, Scld-lu contact potential, Levitt contact potential, Braun contact potential, Joint entropy, and Pearson correlation)	2D	6	6
21	Shannon entropy sum of the alignment columns	2D	1	1
22	CCMpred prediction	2D	1	1
23	FreeContact prediction	2D	1	1
24	PSICOV prediction	2D	1	1
25	Mean contact potential (from alignment)	2D	1	1
26	Normalized mutual information (from alignment)	2D	1	1
27	Mutual information (from alignment)	2D	1	1
Total			**40**	**56**
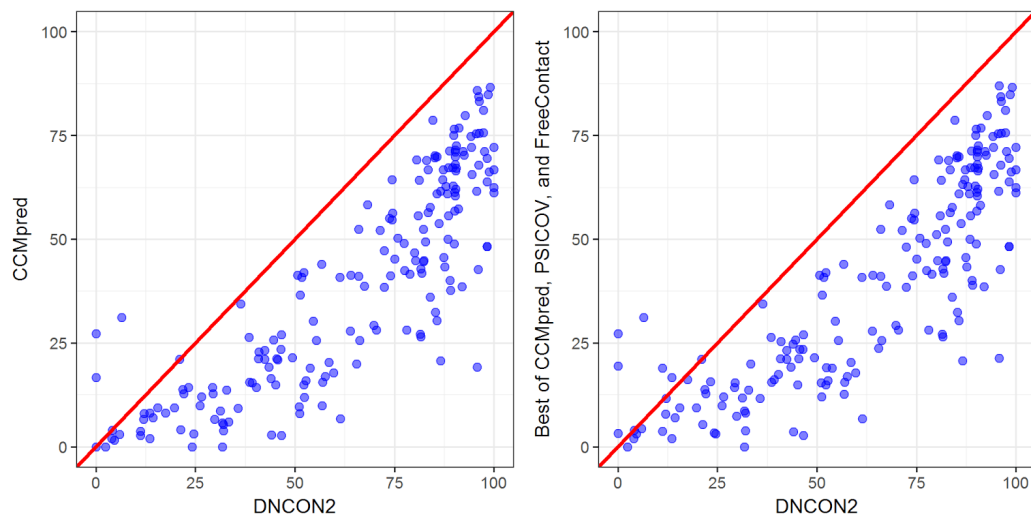
Figure S3. Performance of DNCON2 compared with the input coevolution-based features on the 195 proteins in the validation dataset, measured using the precision of top L/5 long-range contacts. The precisions of DNCON2's predictions are compared with the precisions of best coevolution-based method CCMpred (left) and with the best of the three methods CCMpred, FreeContact, and PSICOV (right).

Table S4. Performance of DNCON2 on CASP12 free-modeling domains, measured using top L/5 and L/2 long-range contacts, and comparison with the state-of-the-art methods Raptor-X, MetaPSICOV, iFold_1, and MULTICOM-NOVEL (DNCON). $P_{top-L/5}$ and $P_{top-L/2}$ stand for the precision of top L/5 and L/2 long-range contacts. The free-modeling domain T0907-D3 is excluded in the evaluation, because the native structure was not available for our evaluations.

Domain	Raptor-X		MetaPSICOV		iFold_1		DNCON		DNCON2	
	$P_{top-L/5}$	$P_{top-L/2}$	$P_{top-L/5}$	$P_{top-L/2}$	$P_{top-L/5}$	$P_{top-L/2}$	$P_{top-L/5}$	$P_{top-L/2}$	$P_{top-L/5}$	$P_{top-L/2}$
T0859-D1	0.0	3.5	4.35	8.77	0.0	0.0	0.0	8.8	0.0	0.0
T0862-D1	42.1	29.8	26.32	12.77	31.6	21.3	21.1	12.8	31.6	14.9
T0863-D1	28.2	17.5	12.82	13.4	10.3	9.3	5.1	8.3	12.8	11.3
T0863-D2	23.6	12.9	6.94	3.93	5.6	6.2	4.2	4.5	12.7	6.2
T0864-D1	80.0	62.6	64	53.66	84.0	77.2	32.0	20.3	77.6	56.1
T0866-D1	95.2	84.6	100	96.15	95.2	90.4	14.3	15.4	90.5	78.9
T0869-D1	57.1	36.5	52.38	32.69	4.8	1.9	47.6	42.3	52.4	32.7
T0870-D1	12.0	11.3	8	14.52	40.0	24.2	40.0	19.4	24.0	16.1
T0878-D1	24.6	23.8	43.48	31.98	55.1	45.4	26.1	16.9	59.4	38.4
T0880-D2	21.9	15.2	25	16.46	-	-	18.8	13.9	48.4	36.7
T0886-D1	100.0	88.6	100	91.43	100.0	85.7	7.1	5.7	100.0	82.9
T0886-D2	92.3	82.8	100	89.06	100.0	82.8	23.1	18.8	100.0	93.8
T0888-D1	24.0	18.0	4	8.2	4.0	8.2	0.0	3.3	54.2	29.5
T0890-D2	18.2	17.0	13.64	15.09	0.0	11.3	9.1	7.6	19.1	15.1
T0892-D2	72.7	52.7	63.64	49.09	63.6	38.2	63.6	41.8	68.2	49.1
T0894-D1	5.6	8.9	0	13.33	77.8	55.6	55.6	37.8	88.9	48.9
T0896-D3	0.0	4.9	3.03	2.47	15.2	16.1	9.1	12.4	12.5	6.2
T0897-D1	10.7	5.8	3.57	4.35	0.0	4.4	17.9	8.7	3.6	7.3
T0897-D2	12.0	8.1	16	9.68	16.0	17.7	20.0	9.7	48.0	27.4
T0898-D1	31.8	24.5	27.27	22.64	4.6	5.7	13.6	15.1	19.1	15.1
T0899-D1	84.6	56.9	86.54	57.69	75.0	51.5	40.4	20.8	90.4	56.9
T0899-D2	66.7	52.3	61.11	47.73	50.0	43.2	33.3	22.7	55.6	34.1
T0900-D1	33.3	31.4	95.24	74.51	23.8	19.6	71.4	52.9	95.0	82.4
T0901-D2	42.9	25.7	50	37.14	100.0	74.3	42.9	28.6	100.0	57.1
T0904-D1	27.5	23.0	25.49	19.84	66.7	44.4	29.4	22.2	84.0	63.5
T0905-D1	93.9	77.7	93.88	77.69	98.0	87.6	63.3	35.5	89.6	64.5
T0905-D2	71.4	51.5	71.43	45.45	92.9	66.7	42.9	33.3	84.6	54.6
T0912-D3	71.4	42.3	42.86	26.92	81.0	44.2	4.8	3.9	81.0	63.5
T0914-D1	34.4	20.3	3.12	5.06	21.9	12.7	31.3	-	18.8	10.1
T0914-D2	15.2	14.8	15.15	11.11	18.2	13.6	15.2	-	21.9	13.6
T0915-D1	41.9	31.2	38.71	23.38	19.4	16.9	29.0	22.1	32.3	23.4
T0918-D1	81.8	72.2	72.73	62.96	54.6	57.4	40.9	22.2	81.8	70.4
T0918-D2	100.0	90.3	84	69.35	88.0	72.6	20.0	27.4	100.0	72.6
T0918-D3	83.3	74.6	87.5	57.63	70.8	69.5	0.0	5.1	79.2	79.7
T0923-D1	19.0	17.5	15.52	12.59	8.6	9.8	22.4	12.6	0.0	7.0
T0941-D1	1.5	4.1	8.7	5.85	1.5	1.2	1.5	4.1	2.9	2.9
T0946-D1	93.8	67.5	62.5	57.5	68.8	37.5	6.3	10.0	37.5	30.0
Avg	**46.3**	**36.8**	**42.9**	**34.7**	**45.7**	**36.8**	**24.9**	**18.5**	**53.4**	**39.3**

2

Table S5. Performance of DNCON2 on all structural domains in CASP12, measured using the precision of long-range contacts.

Domain	Top-5	Top-L/10	Top-L/5	Top-L/2	Top-L	Top-2L
T0859-D1	0.0	0.0	0.0	0.0	0.0	1.8
T0860-D1	20.0	14.3	7.4	17.7	22.1	15.8
T0861-D1	100.0	100.0	98.4	88.5	78.2	56.1
T0862-D1	60.0	44.4	31.6	14.9	9.7	7.0
T0863-D1	40.0	21.1	12.8	11.3	7.8	5.4
T0863-D2	0.0	19.4	12.7	6.2	4.5	2.8
T0864-D1	100.0	84.0	77.6	56.1	43.9	30.9
T0866-D1	100.0	80.0	90.5	78.9	77.9	55.3
T0867-D1	0.0	10.0	19.1	13.5	13.5	10.6
T0868-D1	0.0	16.7	17.4	20.7	27.6	18.5
T0869-D1	60.0	50.0	52.4	32.7	24.0	15.9
T0870-D1	0.0	25.0	24.0	16.1	18.7	13.4
T0871-D1	100.0	100.0	100.0	95.0	75.9	49.5
T0872-D1	100.0	100.0	88.9	75.0	52.3	33.5
T0873-D1	80.0	97.8	95.7	85.7	65.2	42.6
T0877-D1	60.0	57.1	46.4	38.0	27.5	19.4
T0878-D1	100.0	85.3	59.4	38.4	28.8	21.8
T0879-D1	100.0	100.0	100.0	93.6	77.3	55.0
T0880-D2	60.0	56.3	48.4	36.7	26.8	18.8
T0881-D1	0.0	0.0	0.0	5.0	7.9	7.7
T0882-D1	0.0	12.5	18.8	12.5	15.2	12.0
T0884-D1	20.0	42.9	50.0	33.3	21.1	14.1
T0885-D1	100.0	81.8	78.3	61.4	38.6	25.9
T0886-D1	100.0	100.0	100.0	82.9	78.3	62.3
T0886-D2	100.0	100.0	100.0	93.8	74.8	53.5
T0888-D1	60.0	75.0	54.2	29.5	22.3	14.9
T0889-D1	100.0	100.0	97.9	94.2	83.3	58.2
T0890-D1	100.0	75.0	68.8	39.0	25.6	18.9
T0890-D2	40.0	18.2	19.1	15.1	15.1	13.2
T0891-D1	100.0	100.0	95.5	96.4	86.6	56.3
T0892-D1	80.0	71.4	50.0	40.0	21.7	15.2
T0892-D2	100.0	90.9	68.2	49.1	36.4	23.6
T0893-D1	0.0	28.6	13.3	16.2	13.7	11.6
T0893-D2	100.0	100.0	100.0	96.5	86.4	61.8
T0894-D1	100.0	100.0	88.9	48.9	27.0	16.3
T0894-D2	80.0	80.0	63.6	40.7	25.9	17.6
T0895-D1	40.0	50.0	29.2	21.7	16.7	11.7
T0896-D1	0.0	0.0	5.9	11.6	16.3	16.3
T0896-D2	20.0	10.0	25.0	20.0	17.5	12.8
T0896-D3	40.0	12.5	12.5	6.2	6.8	5.6
T0897-D1	0.0	0.0	3.6	7.3	8.0	5.4
T0897-D2	100.0	91.7	48.0	27.4	23.4	17.3
T0898-D1	20.0	27.3	19.1	15.1	10.4	9.9
T0898-D2	20.0	16.7	9.1	10.7	7.3	10.9
T0899-D1	100.0	92.3	90.4	56.9	36.7	22.2
T0899-D2	80.0	77.8	55.6	34.1	22.7	14.8
T0900-D1	100.0	100.0	95.0	82.4	59.8	41.7
T0901-D1	100.0	100.0	97.8	82.1	60.5	43.3
T0901-D2	100.0	100.0	100.0	57.1	35.7	22.1
T0902-D1	100.0	100.0	93.5	89.7	84.4	57.6
T0903-D1	100.0	96.9	93.9	87.7	65.4	34.4
T0904-D1	100.0	88.0	84.0	63.5	39.4	23.3
T0905-D1	100.0	95.8	89.6	64.5	49.6	34.1
T0905-D2	100.0	100.0	84.6	54.6	39.4	26.5
T0909-D1	80.0	65.6	50.8	37.8	28.5	19.8
T0911-D1	100.0	87.8	84.2	83.8	70.3	48.8
T0912-D1	100.0	100.0	98.8	88.9	81.4	59.5
T0912-D2	100.0	100.0	88.2	64.3	49.4	36.1
T0912-D3	100.0	90.0	81.0	63.5	42.7	28.6
T0913-D1	100.0	88.2	83.8	74.0	58.0	40.8
T0914-D1	0.0	18.8	18.8	10.1	7.6	6.0
T0914-D2	20.0	25.0	21.9	13.6	10.5	8.6
T0915-D1	0.0	20.0	32.3	23.4	16.9	11.4
T0917-D1	100.0	100.0	100.0	91.8	83.4	63.4
T0918-D1	100.0	90.9	81.8	70.4	52.8	31.9
T0918-D2	100.0	100.0	100.0	72.6	56.1	43.1
T0918-D3	100.0	75.0	79.2	79.7	66.1	37.3

T0920-D1	100.0	100.0	98.4	90.1	75.7	53.9
T0920-D2	0.0	13.6	13.6	12.7	8.7	5.5
T0921-D1	100.0	100.0	96.4	84.1	66.7	53.6
T0922-D1	80.0	71.4	46.7	46.0	36.5	25.0
T0923-D1	0.0	0.0	0.0	7.0	7.7	7.5
T0928-D1	100.0	94.1	86.8	65.5	50.4	34.9
T0941-D1	0.0	0.0	2.9	2.9	1.8	1.6
T0942-D1	80.0	88.2	74.3	49.4	30.6	19.4
T0942-D2	100.0	85.7	74.4	55.1	31.8	19.4
T0943-D1	0.0	0.0	0.0	29.0	32.3	18.6
T0943-D2	100.0	97.8	95.5	79.0	57.5	38.8
T0944-D1	100.0	96.0	88.2	84.3	74.7	50.2
T0945-D1	100.0	94.7	92.0	71.8	51.5	33.2
T0946-D1	60.0	50.0	37.5	30.0	20.0	16.3
T0946-D2	100.0	95.2	95.2	84.0	66.0	42.7
T0947-D1	100.0	94.4	88.6	79.6	66.3	44.3
T0948-D1	40.0	46.7	36.7	22.7	18.8	12.4
Average	**67.1**	**65.3**	**59.9**	**49.2**	**39.1**	**27.1**

Table S6. Performance of DNCON2 on CASP11 free-modeling domains, measured using top L/5 (upper table) and L/2 (lower table) long-range, medium-range, and short-range contacts, and comparison with MetaPSICOV (local run), and the best performing method in CASP11, CONSIP2.

FM Domain	Long-Range (Top L/5)			Medium-Range (Top L/5)			Short-Range (Top L/5)		
	CONSIP2	MetaPSICOV	DNCON2	CONSIP2	MetaPSICOV	DNCON2	CONSIP2	MetaPSICOV	DNCON2
T0761-D1	5.6	0.0	5.6	38.9	33.3	16.7	77.8	72.2	16.7
T0761-D2	8.7	13.0	4.4	69.6	65.2	65.2	60.9	69.6	56.5
T0763-D1	46.2	15.4	30.8	53.9	61.5	30.8	61.5	57.7	34.6
T0767-D2	58.3	66.7	50.0	22.2	33.3	52.8	16.7	19.4	19.4
T0771-D1	10.0	16.7	70.0	40.0	46.7	60.0	56.7	63.3	70.0
T0777-D1	23.2	18.8	18.8	29.0	27.5	18.8	49.3	40.6	42.0
T0781-D1	5.0	2.5	2.5	32.5	42.5	40.0	52.5	57.5	55.0
T0785-D1	18.2	4.6	18.2	27.3	27.3	72.7	50.0	59.1	50.0
T0789-D1	51.7	62.1	100.0	69.0	69.0	86.2	65.5	69.0	75.9
T0789-D2	28.0	60.0	92.0	80.0	80.0	100.0	92.0	92.0	100.0
T0790-D1	44.4	59.3	85.2	81.5	77.8	88.9	88.9	92.6	100.0
T0790-D2	26.9	69.2	96.2	69.2	80.8	100.0	84.6	96.2	100.0
T0791-D1	53.3	66.7	93.3	56.7	66.7	63.3	76.7	83.3	86.7
T0791-D2	42.9	75.0	92.9	75.0	96.4	89.3	57.1	75.0	85.7
T0794-D2	26.5	38.2	94.1	38.2	55.9	79.4	47.1	55.9	88.2
T0806-D1	84.3	70.6	98.0	70.6	47.1	84.3	64.7	64.7	74.5
T0808-D2	35.2	27.8	92.6	59.3	74.1	70.4	70.4	74.1	75.9
T0810-D1	17.4	21.7	21.7	4.4	0.0	8.7	0.0	13.0	17.4
T0814-D1	37.0	48.2	100.0	59.3	59.3	77.8	55.6	44.4	44.4
T0814-D2	82.6	73.9	100.0	73.9	69.6	100.0	73.9	65.2	91.3
T0820-D1	5.6	5.6	0.0	33.3	27.8	11.1	16.7	0.0	5.6
T0824-D1	45.5	72.7	81.8	36.4	68.2	77.3	45.5	22.7	40.9
T0827-D2	10.0	20.0	33.3	26.7	26.7	46.7	43.3	46.7	43.3
T0831-D2	7.7	7.7	10.3	28.2	25.6	25.6	20.5	20.5	23.1
T0832-D1	2.4	4.8	16.7	9.5	19.1	19.1	38.1	38.1	33.3
T0834-D1	5.0	0.0	0.0	40.0	40.0	50.0	85.0	80.0	85.0
T0834-D2	17.7	5.9	23.5	35.3	35.3	29.4	47.1	23.5	41.2
T0836-D1	43.9	68.3	43.9	19.5	24.4	24.4	14.6	14.6	12.2
T0837-D1	29.2	37.5	20.8	16.7	29.2	20.8	25.0	29.2	33.3
T0855-D1	17.4	0.0	4.4	69.6	69.6	47.8	78.3	69.6	73.9
Average	**29.7**	**34.4**	**50.0**	**45.5**	**49.3**	**55.3**	**53.9**	**53.7**	**55.9**

FM Domain	Long-Range (Top L/2)			Medium-Range (Top L/2)			Short-Range (Top L/2)		
	CONSIP2	MetaPSICOV	DNCON2	CONSIP2	MetaPSICOV	DNCON2	CONSIP2	MetaPSICOV	DNCON2
T0761-D1	4.6	2.3	11.4	27.3	25.0	25.0	45.5	47.7	18.2
T0761-D2	5.3	5.3	7.0	45.6	56.1	43.9	42.1	57.9	35.1
T0763-D1	26.2	12.3	16.9	33.9	47.7	33.9	41.5	38.5	24.6
T0767-D2	44.4	41.1	35.6	12.2	27.8	23.3	11.1	13.3	12.2
T0771-D1	6.6	14.5	44.7	31.6	32.9	43.4	39.5	42.1	42.1
T0777-D1	12.7	14.5	19.1	16.8	15.6	15.6	34.1	30.1	31.2
T0781-D1	6.0	1.0	3.0	21.0	25.0	32.0	31.0	33.0	31.0
T0785-D1	12.5	10.7	25.0	28.6	30.4	51.8	28.6	33.9	33.9
T0789-D1	43.1	48.6	80.6	43.1	52.8	65.3	38.9	43.1	41.7
T0789-D2	19.1	41.3	69.8	52.4	57.1	66.7	55.6	57.1	60.3
T0790-D1	39.7	44.1	64.7	52.9	58.8	64.7	67.7	73.5	73.5
T0790-D2	20.0	44.6	63.1	50.8	58.5	66.2	58.5	55.4	61.5
T0791-D1	41.3	56.0	78.7	36.0	38.7	32.0	49.3	44.0	52.0
T0791-D2	29.0	52.2	71.0	49.3	71.0	72.5	46.4	46.4	53.6
T0794-D2	24.4	38.4	77.9	33.7	32.6	46.5	30.2	32.6	52.3
T0806-D1	60.9	57.8	82.0	39.1	37.5	60.2	39.8	41.4	51.6
T0808-D2	25.2	32.6	78.5	44.4	50.4	50.4	48.2	51.1	54.8
T0810-D1	12.3	15.8	10.5	3.5	5.3	5.3	3.5	8.8	15.8
T0814-D1	37.7	33.3	84.1	33.3	33.3	49.3	33.3	33.3	40.6
T0814-D2	50.0	41.4	84.5	55.2	63.8	75.9	56.9	46.6	65.5
T0820-D1	4.4	2.2	4.4	22.2	22.2	17.8	6.7	4.4	4.4
T0824-D1	37.0	51.9	70.4	22.2	33.3	48.2	22.2	16.7	22.2
T0827-D2	12.0	16.0	26.7	18.7	18.7	24.0	26.7	28.0	33.3
T0831-D2	6.1	5.1	14.1	16.2	15.2	16.2	12.1	13.1	13.1
T0832-D1	2.9	11.4	10.5	12.4	17.1	13.3	21.9	24.8	29.5
T0834-D1	2.0	0.0	0.0	28.0	18.0	26.0	50.0	48.0	52.0
T0834-D2	16.3	9.3	20.9	27.9	23.3	32.6	23.3	11.6	23.3
T0836-D1	31.4	53.9	23.5	13.7	13.7	14.7	9.8	9.8	7.8
T0837-D1	18.0	21.3	11.5	16.4	18.0	19.7	19.7	18.0	26.2
T0855-D1	13.8	10.3	3.5	65.5	60.3	41.4	58.6	48.3	55.2
	22.2	**26.3**	**39.8**	**31.8**	**35.3**	**39.2**	**35.1**	**35.1**	**37.3**

Table S7. Performance of DNCON2 on CASP10 free-modeling domains, measured using top L/5 (upper table) and L/2 (lower table) long-range, medium-range, and short-range contacts, and comparison with MetaPSICOV (local run), and the best performing method in CASP10, DNCON.

FM Domain	Long-Range (Top L/5)			Medium-Range (Top L/5)			Short-Range (Top L/5)		
	DNCON	MetaPSICOV	DNCON2	DNCON	MetaPSICOV	DNCON2	DNCON	MetaPSICOV	DNCON2
T0658-D1	24.2	60.6	75.8	6.1	6.1	9.1	12.1	18.2	15.2
T0666-D1	11.1	27.8	41.7	5.6	13.9	27.8	11.1	11.1	19.4
T0684-D2	38.2	61.8	38.2	5.9	14.7	14.7	23.5	26.5	20.6
T0693-D1	10.0	5.0	5.0	20.0	40.0	35.0	40.0	50.0	70.0
T0695-D1	27.9	26.0	23.1	18.3	11.5	13.5	5.8	10.6	12.5
T0719-D6	33.3	21.2	0.0	39.4	0.0	21.2	39.4	30.3	39.4
T0726-D3	0.0	58.3	100.0	8.3	25.0	0.0	25.0	41.7	50.0
T0734-D1	7.1	16.7	76.2	26.2	31.0	38.1	35.7	47.6	64.3
T0735-D2	16.7	50.0	50.0	11.1	33.3	22.2	38.9	38.9	50.0
T0737-D1	13.0	30.4	13.0	0.0	0.0	4.4	26.1	21.7	26.1
T0739-D1	11.8	11.8	5.9	58.8	47.1	58.8	58.8	58.8	76.5
T0739-D2	58.3	66.7	91.7	41.7	41.7	33.3	50.0	66.7	83.3
T0740-D1	16.1	19.4	0.0	12.9	22.6	19.4	12.9	12.9	12.9
T0741-D1	4.0	4.0	4.0	24.0	8.0	4.0	24.0	40.0	0.0
T0756-D2	0.0	0.0	0.0	0.0	5.9	11.8	17.7	23.5	35.3
Average	**18.1**	**30.6**	**35.0**	**18.5**	**20.0**	**20.9**	**28.1**	**33.2**	**38.4**

| FM Domain	Long-Range (Top L/2)			Medium-Range (Top L/2)			Short-Range (Top L/2)		
	DNCON	MetaPSICOV	DNCON2	DNCON	MetaPSICOV	DNCON2	DNCON	MetaPSICOV	DNCON2
T0658-D1	19.3	36.1	62.7	9.6	6.0	13.3	15.7	12.1	14.5
T0666-D1	8.9	21.1	26.7	7.8	12.2	21.1	4.4	7.8	8.9
T0684-D2	26.2	44.1	33.3	7.1	9.5	17.9	15.5	15.5	13.1
T0693-D1	12.0	6.0	10.0	20.0	26.0	24.0	26.0	40.0	42.0
T0695-D1	18.0	18.0	14.6	10.3	10.3	10.0	5.8	6.9	6.1
T0719-D6	17.1	13.4	22.0	23.2	3.7	17.1	31.7	24.4	23.2
T0726-D3	0.0	40.0	83.3	16.7	23.3	3.3	10.0	23.3	26.7
T0734-D1	6.6	13.2	48.1	15.1	23.6	29.3	22.6	29.3	30.2
T0735-D2	6.8	34.1	31.8	6.8	29.6	15.9	18.2	18.2	25.0
T0737-D1	17.0	13.6	22.0	1.7	0.0	5.1	20.3	22.0	22.0
T0739-D1	9.3	7.0	4.7	39.5	34.9	44.2	48.8	39.5	67.4
T0739-D2	51.7	41.4	51.7	24.1	24.1	27.6	27.6	27.6	51.7
T0740-D1	11.5	18.0	7.7	6.4	12.8	11.5	7.7	9.0	6.4
T0741-D1	7.9	7.9	4.8	19.1	19.1	6.4	22.2	34.9	1.6
T0756-D2	0.0	2.3	0.0	4.7	2.3	4.7	9.3	11.6	16.3
Average	**14.2**	**21.1**	**28.2**	**14.1**	**15.8**	**16.7**	**19.1**	**21.5**	**23.7**

6
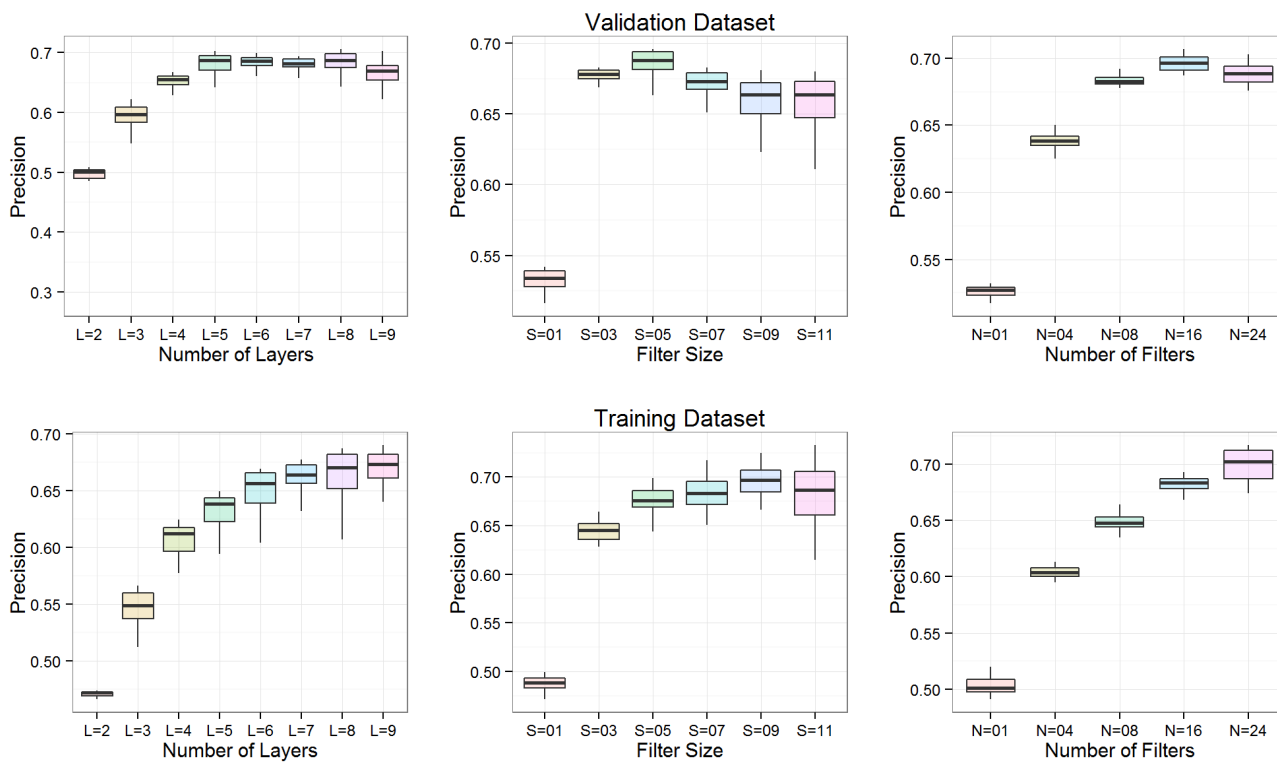
Figure S8. Effect of number of layers (first column), filter size (second column), and number of filters (last column) on the precision of top L/5 long-range predictions for the 195 proteins in the validation dataset (top row) and a sample of training data set (bottom row). The increasing performance on training dataset (plots in bottom row) shows that increasing the filter size, number of filters, and depth of the network increase the learning capacity of the network. Data collected for these plots are from the 30 best models obtained over 40 epochs of training.
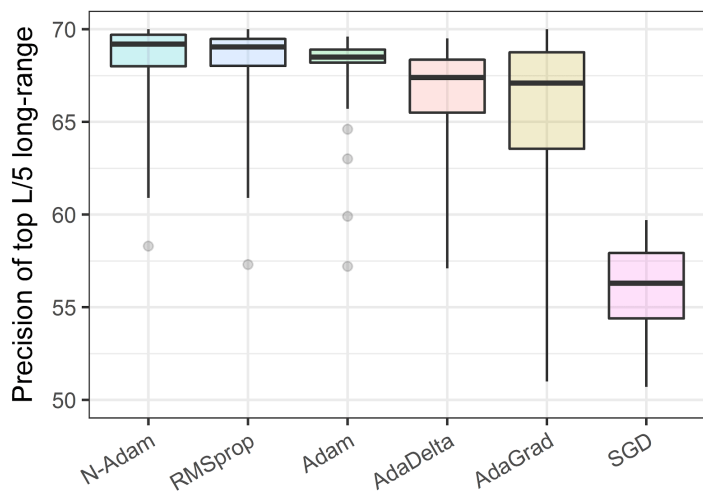
Figure S9. Comparison of the performance of various optimization functions (ADADELTA, Adagrad, Adam, Nesterov Adam, RMSprop, and Stochastic gradient descent) on the 195 proteins in the validation dataset.
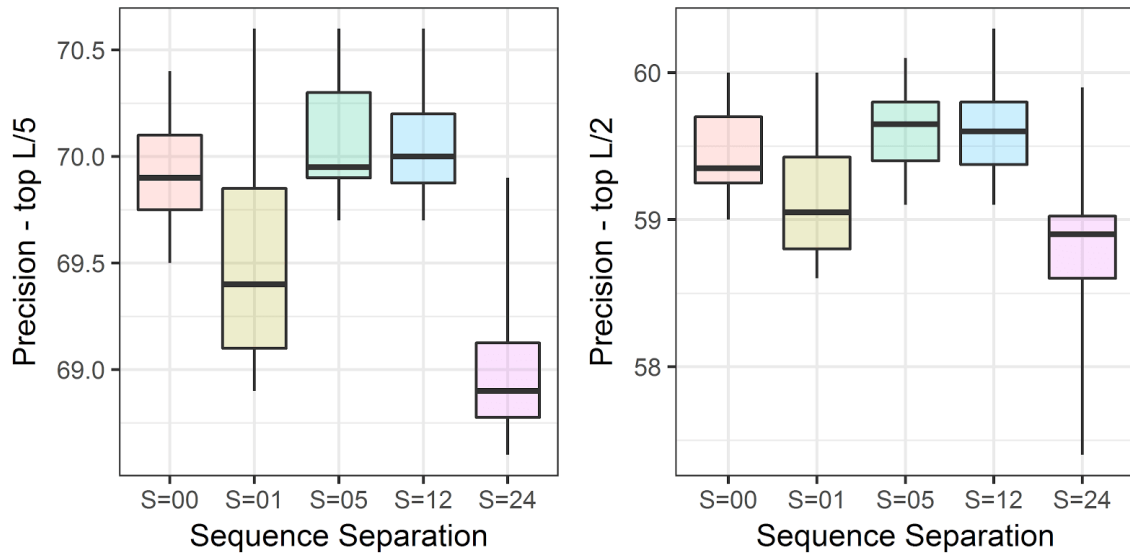
Figure S10. Precision of top L/5 (left) and L/2 (right) long-range contacts on the 195 proteins in validation dataset using models that are trained with contacts at minimum sequence separation 0, 1, 5, 12, and 24. The sequence separation at 12 corresponds to training using medium-range and long-range contacts, and the sequence separation at 24 corresponds to training using long-range contacts only.