Supporting Tables

Table S1 Overview of protein training and test sets

	Training Set	Test Set
PDBs	282	194
Residues	115,440	76,889
Helices	41,637	30,540
Strands	26,708	18,789
Loops	47,095	27,560

Table S2 Protein Training Set

1A05	1A28	1A2Z	1A34	1A3A	1A44	1A4M	1A4Y	1A58	1A6O
1A7W	1A80	1A8B	1A99	1AA7	1AAC	1AAP	1AAZ	1ABE	1ABR
1AD1	1AD6	1AE1	1AEW	1AG9	1AGI	1AGJ	1AH7	1AHO	1AIL
1AIR	1AJ8	1AKO	1AMK	1AMW	1AOE	1AOH	1APX	1AQB	1ARS
1AUN	1AVA	1AVB	1AW8	1AWC	1AWD	1AY7	1AYX	1AZP	1AZZ
1B00	1B0N	1B0W	1B0X	1B33	1B3A	1B3T	1B4B	1B5E	1B8O
1B93	1BB9	1BDO	1BHD	1BHP	1BJ1	1BJP	1BK5	1BLU	1BM8
1BM9	1BO4	1BOV	1BPL	1BS4	1BS9	1BTN	1BU7	1BUE	1BX4
1BX7	1BYF	1BZY	1C26	1C2T	1C3M	1C52	1C5E	1C76	1C9H
1C9O	1C9S	1CBJ	1CBK	1CC8	1CFY	1CFZ	1CIQ	1CJD	1CK4
1CKU	1CL8	1CLK	1CMB	1COZ	1CQ3	1CQD	1CQY	1CVW	1CXY
1D0B	1D0I	1D0Q	1D1J	1D1P	1D3V	1D4T	1D6R	1D8D	1D9T
1DAN	1DCE	1DCP	1DD3	1DD6	1DIF	1DK0	1DK7	1DLM	1DLW
1DM9	1DOZ	1DQI	1DQO	1DS7	1DSZ	1DTD	1DUN	1DWV	1DXE
1DXJ	1DZR	1E19	1E20	1E2A	1E6T	1EAY	1ED1	1EDY	1EI7
1EJ1	1EJF	1EKG	1EKJ	1EL6	1EM9	1EMV	1ENH	1EQ6	1ESC
1EUR	1EUV	1EYV	1EZ3	1F05	1F0C	1F3U	1F3V	1F3Z	1F41
1F5M	1F60	1F8Y	1F94	1FD3	1FLJ	1FQT	1FS1	1FUP	1FXO
1FZD	1G24	1G71	1GAK	1GCI	1GEN	1GNK	1GOT	1GYM	1HFE
1HOE	1HYP	1ICF	1IMB	1JDW	1JPC	1KNB	1KP6	1KPF	1KUH
1LAU	1LBE	1LDT	1LKI	1LUC	1MAT	1MKA	1MML	1MRO	1MSK
1NHK	1NOA	1OPC	1OUN	1OYC	1P35	1PBW	1PDO	1PHP	1PML
1POC	1POH	1PUD	1PYT	1QAU	1QB0	1QB2	1QC7	1QF9	1QHW
1QJB	1QJE	1QK2	1QSD	1QTN	1RAV	1RIS	1ROP	1RPX	1SFP
1TBG	1TFE	1TIG	1TOA	1TUP	1TYF	1UGI	1UTG	1VHH	1VHR
1VID	1WGJ	1WHI	1WHO	1YCQ	256B	2A0B	2ACY	2AFG	2AHJ
2BBK	2CTC	2E2C	2END	2KAU	2MHR	2NSY	2PTH	2SIC	2WRP
3CLA	3DAA	3EIP	3PYP	3TDT	4AAH	4ICB	4PAH	4UBP	6INS
6PRC	7CEI								

Table S3 Protein Test Set

1A17	1A1X	1A6J	1AH4	1AHS	1AJ2	1AL3	1AMX	1ANN	1APA
1ARB	1ARV	1AUG	1AX0	1AYF	1B16	1B4F	1B66	1B67	1B7D
1BBP	1BBZ	1BD8	1BEO	1BF4	1BIX	1BKP	1BKR	1BLI	1BLX
1BN8	1BOU	1BPI	1BQK	1BR9	1BU5	1BUO	1BV1	1BVY	1BXS
1BXY	1BYQ	1BYR	1C08	1CEX	1CHD	1CJB	1CJW	1CNU	1CO6
1CTF	1CV8	1CYN	1CYO	1D2Z	1D4O	1D7D	1DEA	1DFU	1DIO
1DJ8	1DJR	1DK8	1DLG	1DLY	1DP4	1DPJ	1DQE	1DV8	1DYN
1DZ3	1E6I	1E79	1EAI	1ECM	1ECP	1ECY	1EDM	1EFV	1EGP
1EGW	1EJ8	1EO9	1EP0	1EQO	1ERT	1EW4	1EYQ	1F21	1F2K
1F7D	1FAS	1FKJ	1FLE	1FLM	1FLT	1FOF	1FQ0	1FUA	1FXD
1G31	1G43	1G6G	1GDO	1GPR	1GUX	1HCB	1HCQ	1HLE	1HTP
1HXN	1IBR	1IDA	1IDO	1INP	1IRO	1JAC	1KPT	1LTS	1MJC
1MNM	1MOQ	1MWP	1MZM	1NBA	1NDD	1NEC	1NNC	1NPK	1OTF
1PBV	1PFK	1PHR	1PPF	1PTF	1PUC	1QD9	1QEX	1QGH	1QHV
1QIP	1QKJ	1QQ8	1QTW	1REG	1RFS	1RKD	1RVV	1SEI	1SML
1SPP	1SRV	1STM	1SUP	1TCD	1TIF	1TML	1UKZ	1UNK	1URO
1UTE	1VCC	1VMO	1WAP	1WBA	1XXA	1YCS	2ABK	2ARC	2ASR
2AY1	2BNH	2BOP	2CEV	2CPG	2CRO	2DOR	2ERL	2IZH	2LIS
2MTA	2PII	2PLC	2POO	2PRD	2PRG	2RN2	2TNF	2TRC	3FAP
3FIB	4FGF	4MON	4NOS						

Table S4 PCASSO Features

Feature	Type	Atoms Involved	Feature	Type	Atoms Involved
1	Distance	i, i-5	23	Distance	i, j+2
2	Distance	i, i-4	24	Distance	i+1, j-2
3	Distance	i, i-3	25	Distance	i+1, j-1
4	Distance	i, i-2	26	Distance	i+1, j
5	Distance	i, i-1	27	Distance	i+1, j+1
6	Distance	i, i+1	28	Distance	i+1, j+2
7	Distance	i, i+2	29	Distance	i-1, k-2
8	Distance	i, i+3	30	Distance	i-1, k-1
9	Distance	i, i+4	31	Distance	i-1, k
10	Distance	i, i+5	32	Distance	i-1, k+1
11	Angle	i-1, i, i+1	33	Distance	i-1, k+2
12	Dihedral	i, i+1, i+2, i+3	34	Distance	i, k-2
13	Angle	i-2, i, i+2	35	Distance	i, k-1
14	Distance	i-1, j-2	36	Distance	i, k
15	Distance	i-1, j-1	37	Distance	i, k+1
16	Distance	i-1, j	38	Distance	i, k+2
17	Distance	i-1, j+1	39	Distance	i+1, k-2
18	Distance	i-1, j+2	40	Distance	i+1, k-1
19	Distance	i, j-2	41	Distance	i+1, k
20	Distance	i, j-1	42	Distance	i+1, k+1
21	Distance	i, j	43	Distance	i+1, k+2
22	Distance	i, j+1			

Table S5 Data for the three lowest SSE accuracy structures

PDBID	Percent Accuracy				Correctly Predicted by PCASSO / DSSP			
	Helices	Strand	Loop	Total	Helices	Strand	Loop	Total
1CYO	96.6	75.0	79.5	84.1	28 / 29	15 / 20	31 / 39	74 / 88
1MWP	78.6	82.9	89.4	85.4	11 / 14	29 / 35	42 / 47	82 / 96
1G43	44.4	83.9	96.9	86.9	4 / 9	73 / 87	62 / 64	139 / 160

Table S6 PCASSO accuracy comparison

	Percent Accuracy[a]
SSE	PCASSO[b]
Helix	93.7 (97.9)
Strand	89.8 (95.9)
Loop	95.4 (87.8)
All	93.4

---

[a] STRIDE is used as the reference. The true positive rate (sensitivity) is shown and the positive prediction value (precision) is in parentheses.
[b] Trained on DSSP SSE assignments