Prediction of helix capping and β-turn motifs from NMR chemical shifts

Yang Shen and Ad Bax

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.

Supplementary Material

Table S1. Positional sequence preferences for each amino acid at each of the six positions of Ncap and Ccap motifs in the structure database.

	Ncap [a]						***Ccap*** [b]					
	i-1	*i*	*i+1*	*i+2*	*i+3*	*i+4*	*i*	*i+1*	*i+2*	*i+3*	*i+4*	*i+5*
	N'	*Ncap*	*N1*	*N2*	*N3*	*N4*	*C3*	*C2*	*C1*	*Ccap*	*C'*	*C''*
A	0.74	*0.08*	1.06	1.36	1.14	**1.64**	**2.36**	1.24	**1.67**	**1.41**	*0.15*	1.04
C	1.02	1.18	0.41	*0.24*	0.86	1.11	2.15	0.56	0.48	1.26	*0.16*	1.37
D	0.68	**3.95**	1.00	**1.63**	**2.21**	*0.25*	0.32	0.67	0.97	0.47	0.80	0.64
E	0.61	*0.20*	1.49	**3.23**	**3.05**	0.42	0.60	**1.63**	**2.08**	0.95	0.38	0.76
F	1.32	*0.04*	0.93	0.57	0.53	**1.50**	1.27	0.65	0.43	1.33	*0.08*	1.35
G	1.00	*0.25*	0.43	0.60	0.60	*0.25*	0.33	*0.22*	0.28	0.27	**8.94**	0.87
H	0.73	0.86	0.70	0.79	1.03	0.55	0.54	1.13	0.79	**1.83**	0.80	0.86
I	**1.36**	*0.02*	0.84	0.38	0.37	**1.82**	0.97	1.17	0.57	0.47	*0.01*	**1.76**
K	0.78	*0.20*	1.01	1.21	0.71	0.86	0.64	**2.12**	**2.08**	1.10	1.02	1.20
L	**1.60**	*0.06*	1.04	0.51	0.59	**1.97**	**2.62**	1.10	0.93	**1.55**	*0.06*	**1.40**
M	**2.03**	*0.06*	0.86	0.50	0.92	**1.76**	**2.02**	1.09	0.99	1.39	*0.11*	1.22
N	0.72	**3.04**	0.52	0.83	0.59	0.35	0.38	0.61	0.84	**1.61**	**2.02**	0.63
P	1.34	*0.00*	**2.95**	0.52	*0.00*	*0.04*	*0.09*	*0.02*	*0.00*	*0.00*	*0.00*	0.55
Q	0.70	*0.15*	0.96	**1.56**	**2.41**	1.05	0.73	**1.40**	**1.74**	1.36	0.73	0.77
R	0.74	*0.19*	1.01	0.91	0.64	1.25	0.87	**1.83**	**1.72**	1.26	0.64	0.86
S	0.70	**5.24**	0.70	0.99	0.88	0.41	0.57	0.64	0.98	0.95	0.34	0.50
T	0.88	**4.29**	0.72	0.83	1.21	0.54	0.36	0.58	0.57	1.23	*0.06*	0.52
V	1.21	*0.02*	0.94	0.49	0.58	**1.47**	0.74	0.96	0.50	0.38	*0.02*	**1.38**
W	1.18	*0.06*	1.22	0.64	0.58	1.31	1.06	0.80	0.45	0.65	*0.04*	0.98
Y	1.03	*0.08*	0.84	0.63	0.48	1.09	1.00	0.83	0.58	1.52	*0.09*	1.24
c[c]	0.49	*0.06*	0.42	0.44	0.39	1.27	1.28	0.55	*0.08*	0.46	*0.11*	0.59

The sequence preferences for each of the amino acids are calculated by $P^i(a) = f^i(a) / < f >$, where $f^i(a)$ is the ratio of the number of a particular amino acid a in position i of an Ncap/Ccap over the number of that amino acid in the entire database, and $< f >$ is the ratio of the total number of amino acids in Ncap/Ccap motifs over the total number of amino acids in the database. Significantly high values are bold faced, and very low values are italicized.

[a] Calculated over all hexapeptides in the structure database forming an Ncap box, a *sc*(*i*)←*bb*(*i*+3) Ncap or a *bb*(*i*)→*sc*(*i*+3) Ncap motif.

[b] Calculated over all hexapeptides in the structure database exhibiting an ideal Schellman Ccap or αL Ccap motif.

[c] Lower case c refers to oxidized Cys.

Table S2. Positional β-turn sequence preferences for each amino acid at each of the four positions (*i* to *i*+3) in β-turns observed in the structure database.

	Type I				***Type II***				***Type I'***				***Type II'***				***Type VIII***			
	i	*i+1*	*i+2*	*i+3*	*i*	*i+1*	*i+2*	*i+3*	*i*	*i+1*	*i+2*	*i+3*	*i*	*i+1*	*i+2*	*i+3*	*i*	*i+1*	*i+2*	*i+3*
A	0.54	1.02	0.63	0.91	0.86	1.08	*0.21*	1.08	0.60	0.59	*0.16*	0.66	0.70	0.30	0.63	0.76	1.44	0.92	0.63	0.71
C	1.45	0.43	0.76	1.29	0.64	0.34	*0.19*	1.73	0.99	0.77	*0.22*	0.64	1.72	0.30	0.62	0.90	1.48	0.49	0.92	0.88
D	**3.03**	1.25	**2.95**	1.01	0.54	1.05	1.02	1.28	0.90	**3.39**	0.84	0.35	0.97	1.01	**3.54**	1.02	0.73	**2.21**	1.20	0.73
E	0.44	1.43	1.16	0.63	0.89	1.47	0.28	1.05	0.97	0.99	0.33	1.27	1.02	0.39	1.39	1.05	0.52	1.27	1.07	0.59
F	0.78	0.50	0.94	1.03	0.97	0.61	0.41	0.88	1.44	0.58	0.31	0.63	1.28	*0.20*	0.46	1.06	1.16	0.55	1.46	0.80
G	1.29	0.46	0.64	**3.03**	1.01	0.25	**8.92**	1.18	0.68	1.84	**9.71**	0.51	0.69	**9.64**	0.70	1.31	0.99	0.37	0.17	0.71
H	1.54	0.75	1.54	1.13	0.87	0.77	0.91	0.94	1.47	1.40	0.42	1.23	1.41	0.56	0.90	1.50	0.84	0.95	1.69	0.79
I	0.28	0.42	0.23	0.75	1.06	0.55	*0.01*	0.69	1.58	0.04	*0.00*	0.91	1.13	*0.05*	0.30	0.62	0.73	0.68	1.26	1.23
K	0.53	1.23	1.00	0.74	1.34	1.44	0.30	0.96	1.31	0.88	0.34	2.83	1.01	0.50	1.04	1.31	0.62	1.41	1.12	0.88
L	0.55	0.52	0.52	0.80	1.03	0.45	*0.13*	0.64	0.73	0.24	0.08	0.62	0.77	*0.14*	0.38	0.69	0.78	0.81	0.76	0.85
M	0.56	0.51	0.53	0.86	0.83	0.51	*0.24*	1.18	0.75	0.52	*0.19*	0.81	0.81	*0.23*	0.33	0.89	0.67	0.59	0.70	0.67
N	**2.48**	0.74	**2.64**	1.13	0.73	0.72	**2.32**	0.76	0.70	**5.86**	1.85	0.89	0.87	1.02	**3.19**	1.23	0.86	0.92	1.60	0.73
P	1.28	**4.69**	*0.10*	*0.02*	**2.20**	**4.82**	*0.00*	0.06	0.32	*0.00*	*0.00*	*0.01*	0.57	*0.02*	0.73	0.01	**3.23**	1.60	0.04	3.68
Q	0.54	0.85	1.01	0.83	1.08	0.91	0.35	1.40	0.83	1.01	0.44	2.06	1.25	0.40	0.67	1.18	0.66	0.99	1.02	0.86
R	0.65	0.88	0.92	0.77	1.18	0.92	0.38	0.75	1.01	0.84	0.29	1.93	1.05	0.36	1.00	1.22	0.64	1.05	1.17	0.83
S	1.91	1.40	1.49	1.02	0.68	0.94	0.32	1.50	0.87	0.65	0.38	0.81	1.04	0.32	1.51	1.19	1.32	1.22	1.02	0.87
T	1.48	0.82	1.40	0.87	0.92	0.76	*0.11*	1.47	0.66	*0.04*	*0.07*	0.96	0.83	*0.07*	1.11	1.50	0.86	1.17	1.09	0.95
V	0.28	0.46	*0.25*	0.85	1.04	0.75	*0.01*	1.13	1.75	*0.06*	*0.03*	1.29	1.45	*0.04*	*0.23*	0.80	0.94	0.74	1.42	1.49
W	0.63	0.76	1.24	1.03	0.70	0.51	0.31	0.95	1.29	0.42	0.31	0.70	1.00	*0.20*	0.40	1.20	0.84	0.68	1.04	0.99
Y	0.74	0.56	1.10	1.03	0.94	0.73	0.51	0.93	1.84	0.71	0.27	0.89	1.66	*0.19*	0.40	1.21	0.97	0.63	1.56	0.80
c	0.96	1.07	0.88	1.69	1.82	0.62	0.66	1.44	1.90	0.84	0.74	0.37	1.23	*0.16*	1.07	1.23	1.97	0.58	1.17	1.65

The sequence preference for each of amino acid is calculated as $P_k^i(a) = f_k^i(a) / < f >$, where $f_k^i(a)$ is the ratio of the number of amino acid type a in position i of a type k β-turn over the number of that amino acid type in the entire database, $< f >$ is the ratio of the total number of amino acids in type k β-turns over the total number of amino acids in the database. Only isolated β-turns are considered for the calculations in this Table. Significantly high values are highlighted in bold and low values are italicized. Lower case c refers to oxidized Cys.

Table S3. Average secondary chemical shifts for residues in helix capping motifs and five types of β-turns.

	$<\Delta\delta^{1}H^{N}>$	$<\Delta\delta^{15}N>$	$<\Delta\delta^{1}H^{\alpha}>$	$<\Delta\delta^{13}C'>$	$<\Delta\delta^{13}C^{\alpha}>$	$<\Delta\delta^{13}C^{\beta}>$
helix capping						
Ncap box						
N'	0.01±0.80	-0.65±3.65	0.36±0.39	-0.10±1.41	-0.60±1.51	1.91±1.94
Ncap	***0.54±0.66***	***-0.71±4.06***	***0.38±0.35***	***0.67±1.16***	***-1.58±1.02***	***2.12±1.40***
N1	0.64±0.58	-1.59±2.22	-0.45±0.44	1.71±1.16	3.66±1.42	-0.70±1.15
N2	0.05±0.49	-3.23±2.26	-0.31±0.24	2.54±1.10	3.31±0.79	-0.56±0.85
N3	-0.46±0.38	-0.91±1.85	-0.32±0.21	2.76±0.96	3.08±0.96	0.30±0.95
N4	0.04±0.51	-1.44±2.07	-0.46±0.30	1.62±1.21	3.55±1.11	-0.25±0.99
Ccap (Schellman) [a]						
C3	-0.05±0.55	-3.06±2.55	-0.46±0.31	2.28±1.38	3.36±1.35	-1.00±1.44
C2	-0.08±0.55	-1.73±2.43	-0.20±0.33	3.99±1.03	3.06±1.13	-0.33±0.89
C1	-0.07±0.48	-0.71±2.59	-0.20±0.27	1.28±1.14	2.78±1.02	-0.42±0.79
Ccap	**-0.71±0.45**	**-5.54±2.94**	**0.00±0.28**	**-0.93±1.11**	**-0.21±1.06**	**0.19±1.10**
C'	-0.47±0.30	-3.92±2.62	-0.17±0.31	-0.36±1.18	0.82±0.74	-2.51±1.37
C''	-0.25±0.46	-2.71±3.30	0.05±0.41	-0.48±1.35	-0.61±1.65	0.69±1.72
Ccap (αL) [a]						
C3	0.07±0.46	-1.89±2.78	-0.36±0.38	2.51±0.99	3.44±1.12	-0.17±1.18
C2	-0.20±0.44	-2.89±2.35	-0.29±0.27	2.02±1.39	2.76±1.07	-0.12±0.76
C1	-0.30±0.43	-3.80±2.85	-0.21±0.30	1.52±0.96	2.26±1.04	0.27±0.95
Ccap	**-0.28±0.52**	**-6.41±3.09**	**-0.01±0.27**	**0.14±1.05**	**0.37±1.23**	**1.23±1.39**
C'	-0.41±0.44	-2.51±3.36	-0.11±0.35	-0.92±1.35	0.61±1.22	-1.33±2.04
C''	-0.53±0.68	-2.97±3.70	0.13±0.48	-1.13±1.59	-1.04±1.92	1.27±1.85
Ccap (Proline)						
C3	-0.26±0.47	-3.99±2.49	-0.39±0.29	1.52±0.97	2.52±1.00	-0.30±0.87
C2	-0.67±0.40	-5.34±3.09	-0.28±0.21	1.55±0.92	2.06±1.01	0.18±1.07
C1	-0.71±0.47	-4.92±2.33	-0.17±0.30	0.14±0.93	1.44±1.30	0.98±1.09
Ccap	**-0.35±0.52**	**-3.05±3.40**	**0.26±0.47**	**-2.15±1.59**	**-1.41±3.27**	**-0.26±1.73**
C'	-	-	0.05±0.26	0.60±1.09	1.53±1.24	-0.02±0.76
C''	0.06±0.58	-3.72±3.69	0.04±0.22	0.10±1.29	0.20±1.13	-0.11±1.03
β-turns [b]						
Type I						
i	0.01±0.79	0.11±3.88	0.12±0.43	0.68±1.63	-0.80±1.77	0.85±1.82
i+1	0.33±0.69	-0.40±4.27	-0.18±0.35	0.35±1.23	2.06±1.22	-0.10±0.96
i+2	-0.12±0.78	-4.86±3.26	0.03±0.26	-0.20±1.23	-0.16±1.20	0.09±1.49
i+3	-0.55±0.55	-1.69±3.32	0.06±0.44	-0.89±1.55	-0.08±1.65	0.61±2.22
Type II						
i	0.05±0.81	-0.10±4.30	0.15±0.32	-0.71±1.49	-1.14±1.45	1.07±1.57
i+1	0.16±0.52	0.92±3.30	-0.31±0.36	0.41±0.94	1.59±1.25	-0.16±1.07
i+2	0.57±0.63	1.30±4.03	-0.12±0.33	-0.68±1.10	0.51±1.05	-1.45±1.58
i+3	-0.42±0.48	-1.23±3.08	0.20±0.40	-0.92±1.36	-0.07±1.61	1.46±1.61
Type I'						
i	0.47±0.66	1.82±3.95	0.30±0.32	-0.59±1.23	-0.70±1.31	2.27±1.39
i+1	0.94±0.67	4.18±4.40	-0.36±0.26	-0.55±0.98	1.12±0.80	-1.49±1.24
i+2	0.25±0.56	-6.36±4.17	-0.18±0.30	-1.01±0.96	0.67±0.96	-0.94±1.47
i+3	-0.36±0.38	-0.39±2.24	0.29±0.33	-1.21±1.20	-1.35±1.26	1.82±1.31
Type II'						
i	0.28±0.55	2.02±4.78	0.15±0.30	-0.48±0.98	-0.82±1.27	1.14±1.10
i+1	0.50±0.82	3.90±3.91	-0.19±0.25	-0.38±1.18	1.91±0.87	-1.50±2.16
i+2	0.24±0.49	3.12±3.55	0.08±0.26	-0.71±1.06	-0.07±1.11	0.01±0.92
i+3	-0.27±0.36	-0.88±2.22	0.27±0.39	-1.16±1.23	-0.45±1.44	1.72±1.51
Type VIII						
i	-0.05±0.76	0.07±4.11	0.07±0.42	-0.51±1.40	-0.34±1.68	0.59±1.55
i+1	0.24±0.61	0.63±3.85	-0.11±0.26	-0.16±1.26	1.76±1.53	0.58±1.02
i+2	-0.43±0.54	-4.16±3.46	0.33±0.39	-1.72±1.57	-1.27±1.45	2.35±1.79
i+3	0.28±0.65	0.87±4.08	0.25±0.51	-1.29±1.55	-0.73±1.42	1.62±1.83

[a] only ideal Schellman and αL helix capping motifs are considered (see Methods)
[b] only isolated β-turns are considered (see Methods)

Table S4. Parameters $c_{a,j,k}$ used for β-turn score calculation (Eq. 6).[a]

a_j	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}	$(\phi+\psi)_{i+1}$	$(\phi+\psi)_{i+2}$	$\psi_{i+1}+\phi_{i+2}$
k = *turn I*	3.8	3.3	2.6	2.8	-[b]	-	3
k = *turn II*	3.3	3.5	3	2.5	-	3	-
k = *turn I'*	3	3	3	3	3	3	-
k = *turn II'*	4	4	3	3	-	4	-
k = *turn VIII*	3	3	3	3	-	-	-

[a] The weight factor $w_{a,j,k}$=2 was used for all analyses of turns
[b] These angles are not used to calculate the β-turn score

Table S5. ANN networks for predicting N-terminal helix capping (Ncap), C-terminal helix capping (Ccap) and β-turns.

	Ncap	*Ccap*	*β-turn I*	*β-turn II*	*β-turn I'*	*β-turn II'*	*β-turn VIII*
ANN							
level1							
Input	(6-mer) size: 6×32 [AA_{20}[a],CS_{12}[b]]	(6-mer) 6×32 [AA_{20}[a],CS_{12}[b]]	(4-mer) 4×32 [AA_{20}[a],CS_{12}[b]]	(4-mer) 4×32 [AA_{20}[a],CS_{12}[b]]	(4-mer) 4×32 [AA_{20}[a],CS_{12}[b]]	(4-mer) 4×32 [AA_{20}[a],CS_{12}[b]]	(4-mer) 4×32 [AA_{20}[a],CS_{12}[b]]
Output	[c][H,E,C,S_{Ncap}]	[c][H,E,C,S_{Ccap}]	[c][H,E,C,S_{I}]	[c][H,E,C,S_{II}]	[c][H,E,C,$S_{I'}$]	[c][H,E,C,$S_{II'}$]	[c][H,E,C,S_{VIII}]
level2							
Input	(6-mer) 6×4 [$OUT1_{1-4}$][d]	(6-mer) 6×4 [$OUT1_{1-4}$][d]	(10-mer) 10×4 [$OUT1_{1-4}$][d]	(10-mer) 10×4 [$OUT1_{1-4}$][d]	(10-mer) 10×4 [$OUT1_{1-4}$][d]	(10-mer) 10×4 [$OUT1_{1-4}$][d]	(10-mer) 10×4 [$OUT1_{1-4}$][d]
Output	[c][H,E,C,S_{Ncap}]	[c][H,E,C,S_{Ccap}]	[c][H,E,C,S_{I}]	[c][H,E,C,S_{II}]	[c][H,E,C,$S_{I'}$]	[c][H,E,C,$S_{II'}$]	[c][H,E,C,S_{VIII}]
T/P ratio[e]							
0.2-0.3	0.32	0.15	0.24	0.16	0.21	0.06	0.20
0.3-0.4	0.46	0.25	0.41	0.44	0.22	0.48	0.37
0.4-0.5	0.53	0.45	0.50	0.59	0.73	0.56	0.48
0.5-0.6	0.63	0.55	0.73	0.74	0.86	0.70	0.63
0.6-0.7	0.75	0.69	0.78	0.87	0.75	0.67	0.70
0.7-0.8	0.78	0.84	0.88	0.94	1.00	0.80	0.84
0.8-0.9	0.95	0.95	0.95	0.98	0.94	1.00	0.93
0.9-1.0	0.98	1.00	0.98	1.00	0.97	1.00	0.99
Score cutoff[f]	0.30	0.30	0.31	0.30	0.30	0.22	0.21
Q_{pred}[g]	0.86	0.94	0.78	0.82	0.93	0.78	0.72
Q_{obs}[g]	0.84	0.88	0.70	0.70	0.75	0.73	0.64
MCC[g]	0.85	0.92	0.73	0.76	0.83	0.75	0.67

[a] Amino acid type represented by its values (n=20 for each residue) in the BLOSUM62 sequence homology matrix
[b] $^{1}H^{N}/^{15}N/^{1}H^{\alpha}/^{13}C^{\alpha}/^{13}C^{\beta}/^{13}C'$ secondary chemical shifts (n=6) and Boolean number indicators for missing chemical shifts (n=6)
[c] Four-state classification (n=6) of the query hexpeptides or tetrapeptides; H(Helix)/E(Strand)/C(Coil) are the binary secondary structure classification from DSSP; S_{Ncap}, S_{Ccap} and S_k (k=I, II, I', II' and VIII) are the Ncap/Ccap/β-turn scores developed in this work (see Methods)
[d] 4-state prediction output (n=4) from the first level ANN
[e] true positive ratio distribution with respect to the binned predicted Ncap/Ccap/β-turn scores, P_4^k (shown also in Figs. 6 and 8, main text); a positive prediction is defined if P_4^k falls above the threshold (as listed in the "Score cutoff" row)
[f] selected threshold of the predicted Ncap/Ccap/β-turn score to define a positive MICS Ncap/Ccap/β-turn prediction
[g] Q_{pred}, Q_{obs} and MCC scores used to evaluate the performance of the Ncap/Ccap/β-turn predictions (See Eqs 12, 13, 14, main text)

Table S6. Comparison table of the number of β-turns assigned by using their original definitions and by the method used in this work

	β-turn I	β-turn II	β-turn I'	β-turn II'	β-turn VIII
N_{orig} [a]	65361	24159	7301	4954	27626
N_{new} [b]	66313	25568	7447	4966	76843[d]
N_{both} [c]	64779	23700	6705	4802	24569

[a] the number of the observed β-turns in the protein structure database if the original definition is used to define a β-turn (see Methods; their Cα(i,i+3) distance distributions are shown with the gray bars in Fig. S2). Note that those β-turns include both overlapped and isolated β-turns (as listed in Table 2).

[b] the number of observed β-turns in the protein structure database when Eq. 5 is used to define a β-turn (with the S_k cut-off level listed in Table S5; and their Cα(i,i+3) distance distributions are shown with blue lines in Fig S2).

[c] the number of observed β-turns in the protein structure database as identified by both methods.

[d] Type VIII b-turns in our study include the more "open" turns; 98.4% of these have a Cα(i,i+3) ≤8.5 Å.

Table S7. H-bond[a] preference for residues in Ncap and Ccap motifs.[b]

Ncap

	Ncap	%	N1	%	N2	%	N3	%	N4	%
1	**sc←bb(i+3)**	55.1	bb→bb(i-2)	8.6	bb→sc(i-2)	24.5	**bb→sc(i-3)**	53.9	bb→bb(i-3)	5.9
2	bb←bb(i+3)	23.3	bb←bb(i+3)	5.4	bb←bb(i+3)	6.2	bb→bb(i-3)	23.9	bb←sc(i+4)	5.0
3	**bb→sc(i+3)**	22.6	bb←sc(i+4)	4.6	bb←sc(i+4)	5.3	**sc←bb(i-3)**	22.2	bb←bb(i+3)	3.9
4	sc←bb(i+2)	21.4	bb←sc(i+3)	0.6	sc←sc(i-2)	2.8	bb←sc(i+4)	4.4	sc→bb(i-4)	3.3
5	bb→bb(i-2)	9.1	sc←sc(i+4)	0.5	sc→sc(i-2)	1.8	sc←bb(i-4)	4.3	bb←bb(i+5)	1.8
6	bb→bb(i-3)	8.8	sc→sc(i+4)	0.5	bb←bb(i+5)	1.4	bb←bb(i+3)	3.9	sc→bb(i-5)	1.4
7	sc→sc(i+3)	4.0	bb←bb(i+5)	0.4	bb←sc(i+3)	1.4	sc←sc(i-3)	3.9	sc→bb(i-6)	0.9
8	sc←sc(i+3)	3.4	sc←sc(i+3)	0.3	bb→bb(i-3)	0.5	sc→sc(i-3)	3.4	bb←sc(i+3)	0.7
9	bb→sc(i-2)	3.3			sc←sc(i+4)	0.5	bb→bb(i-4)	1.4	sc←bb(i-5)	0.7
10	bb←sc(i+4)	3.2			sc←sc(i+3)	0.5	bb←bb(i+5)	1.2	sc←bb(i-6)	0.6

Ccap

	C3	%	C2	%	C1		Ccap	%	C'	%
1	**bb←bb(i+5)**	85.8	**bb←bb(i+3)**	74.8	sc←bb(i-4)	3.6	**bb←bb(i-3)**	18.6	**bb←bb(i-3)**	74.8
2	bb←bb(i+4)	39.7	sc←bb(i-4)	3.6	sc←bb(i-3)	3.3	sc←bb(i+2)	9.4	bb←bb(i-4)	39.7
3	**bb←bb(i+3)**	18.6	bb←bb(i-3)	1.2	bb←bb(i-3)	2.1	sc←bb(i-4)	8.7	bb←bb(i+2)	12.0
4	bb←bb(i+6)	10.7	sc←bb(i-3)	0.8	sc←sc(i-4)	0.7	sc←bb(i-3)	2.7	bb←bb(i-5)	2.4
5	sc←bb(i-4)	4.2	sc←bb(i+6)	0.6	sc←sc(i-4)	0.6	bb←bb(i+3)	2.1	bb←bb(i+3)	2.0
6	bb←sc(i+3)	2.7	sc←sc(i-3)	0.6	sc←sc(i-3)	0.5	bb←bb(i-5)	1.5	bb←bb(i+4)	1.3
7	sc←bb(i+7)	2.1	sc←sc(i-4)	0.4	bb←bb(i-5)	0.4	sc←sc(i+2)	1.2		
8	bb←bb(i-3)	1.4	sc←sc(i-4)	0.4	bb←sc(i+2)	0.2	sc←sc(i-4)	0.9		
9	sc←bb(i+5)	1.3	sc←sc(i+6)	0.4	sc←sc(i-3)	0.1	sc←sc(i+2)	0.7		
10	sc←bb(i-3)	0.8	sc←bb(i+6)	0.4			bb←sc(i+2)	0.5		

[a] For each H-bond listed in the table, the first letter (bb or sc) represents the H-bond partner (backbone or side-chain) from present residue in the applicable column, the second letter (bb or sc) with the number j in parentheses represents the H-bond partner from residue *j*, and the arrow denotes the donor to acceptor direction. For residues in an α-helix, such as the N1-N3 and C3-C1 residues, the characteristic bb←bb(i+4) H-bonds are not listed. The classic H-bonds in Ncap and Ccap motifs are bold faced.

[b] Tabulated H-bond statistics include all true Ncap and C-cap motifs, as well as Ncap- and Ccap-like motifs with S_k cutoff scores ≥ 0.3.

Table S8. β-turn prediction performance of various methods for eleven test proteins which are not present in the training database (listed in Table 3, main text)

	β-turn I	*β-turn II*	*β-turn I'*	*β-turn II'*	*β-turn VIII*
MICS					
Q_{obs}	0.86	0.54	0.60	1.00	0.80
Q_{pred}	0.76	1.00	1.00	0.67	0.89
MCC	0.80	0.73	0.77	0.82	0.84
COUDES [a]					
Q_{obs}	0.36	0.28	0.25	0.17	0.11
Q_{pred}	0.22	0.15	0.11	0.05	0.05
MCC	0.25	0.19	0.16	0.09	0.06
DEBT [b]					
Q_{obs}	0.25	0.33	-	-	0.28
Q_{pred}	0.12	0.12	-	-	0.11
MCC	0.13	0.18	-	-	0.16
NetTurnP [c]					
Q_{obs}	0.62	0.39	0.50	0.17	0.56
Q_{pred}	0.21	0.16	0.12	0.02	0.04
MCC	0.33	0.24	0.23	0.05	0.13

[a] Fuchs, PFJ and Alix, AJP (2005) High accuracy prediction of beta-turns and their types using propensities and multiple alignments. Proteins 59: 828–839.

[b] Kountouris P and Hirst JD (2010) Predicting beta-turns and their types using predicted backbone dihedral angles and secondary structures. BMC Bioinformatics 11:407.

[c] Petersen, B, Lundegaard, C and Petersen, TN (2010) NetTurnP - Neural network prediction of beta-turns by use of evolutionary information and predicted protein sequence features. PLoS One 5(11):e15079.

Fitting parameters between the prediction accuracy Q_{pred} (y) and predicted score P_4 (x) in Figures 7&8:

Ncap motif (Figure 7a'): $y = -0.755\,x^2 + 1.864\,x - 0.124$ [S1]

Ccap motif (Figure 7b'): $y = -1.076x^2 + 2.315\,x - 0.226$ [S2]

Type I β-turn (Figure 8a'): $y = -0.954\,x^2 + 2.169\,x - 0.221$ [S3]

Type II β-turn (Figure 8b'): $y = -1.321x^2 + 2.679\,x - 0.343$ [S4]

Type I' β-turn (Figure 8c'): $y = 0.620x^3 - 2.370x^2 + 3.008\,x - 0.256$ [S5]

Type II' β-turn (Figure 8d'): $y = 0.695x^3 - 2.512x^2 + 3.083\,x - 0.267$ [S6]

Type VIII β-turn (Figure 8e'): $y = -0.936x^2 + 2.040\,x - 0.098$ [S7]

α-helix: $y = -0.105x^2 + 1.033\,x + 0.070$ [S8]

β-sheet: $y = -0.175x^2 + 0.934\,x + 0.214$ [S9]

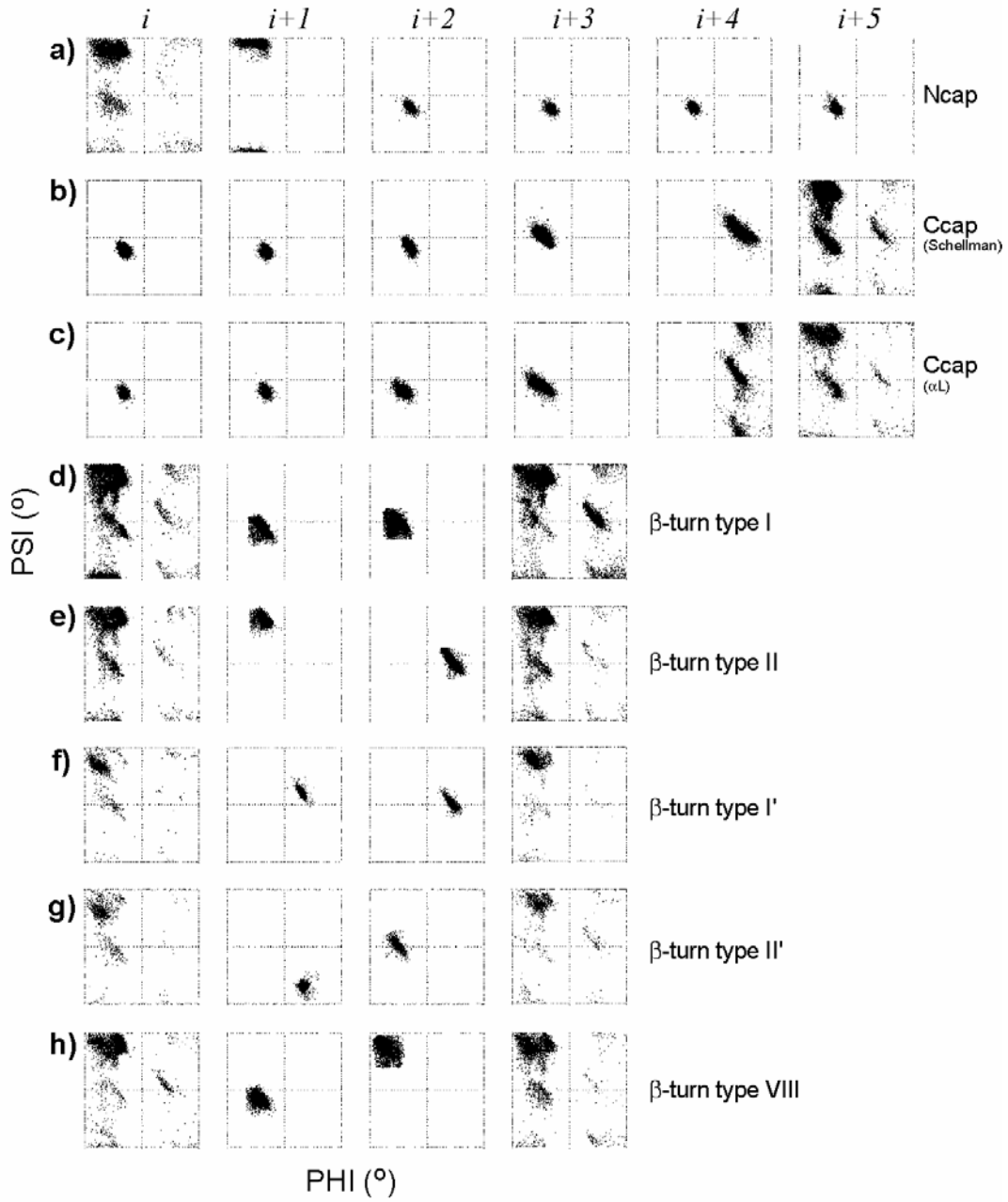

Fig. S1 Positional ϕ/ψ torsion angle distributions for residues in three helix capping motifs and five types of β-turns. (A) Plots for all hexapeptides (N′-Ncap-N1-N2-N3-N4 for residues from position *i-1* to *i*+4, respectively forming an N-terminal helix capping box motif in the structure database. (B-C) Plots for all hexapeptides (C3-C2-C1-Ccap-C′-C″ for residues from position *i* to *i*+5, respectively) forming an ideal Schellman (B) or αL (C) Ccap motif (see Methods). (D-H) Plots for all tetrapeptides (for residues from position *i* to *i*+3, respectively) forming an isolated type I (D), type II (E), type I′ (F), type II′ (G) or type VIII (H) β-turn. All plots cover a range of -180° to 180° for both ϕ and ψ angles, and definitions used for selecting helix capping and turn motifs follow the original literature.

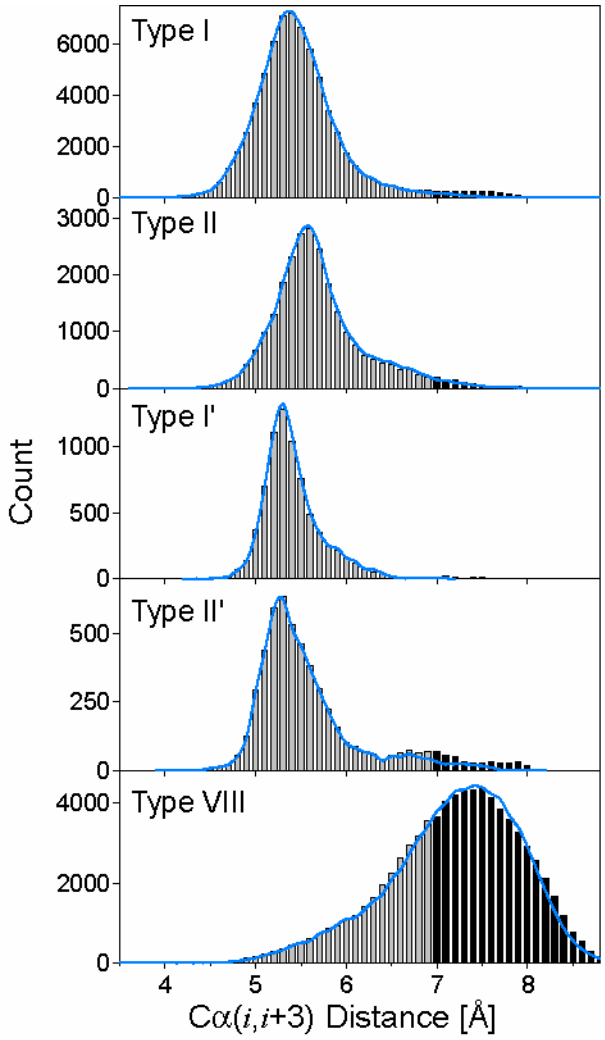

Fig. S2 Histogram of $C_\alpha(i,i+3)$ distances in five types of β-turns. The distribution of distances between the C^α atoms of the first (i) and the last (i+3) residue is plotted (gray bars) for five types of β-turns (type I, II, I′, II′, and VIII respectively) observed in the sequence database. If a relaxed distance cutoff criterion, $C_\alpha(i,i+3) \leq 9$ Å, is used to define a β-turn, the $C_\alpha(i,i+3)$ distance distributions of the additional "open" β-turns (or 4-residue fragments) are shown as black bars. Moreover, the distribution of $C_\alpha(i,i+3)$ distances in the β-turns defined by using Eq 5 (with a score cutoff listed in Table S5) is plotted as blue lines.

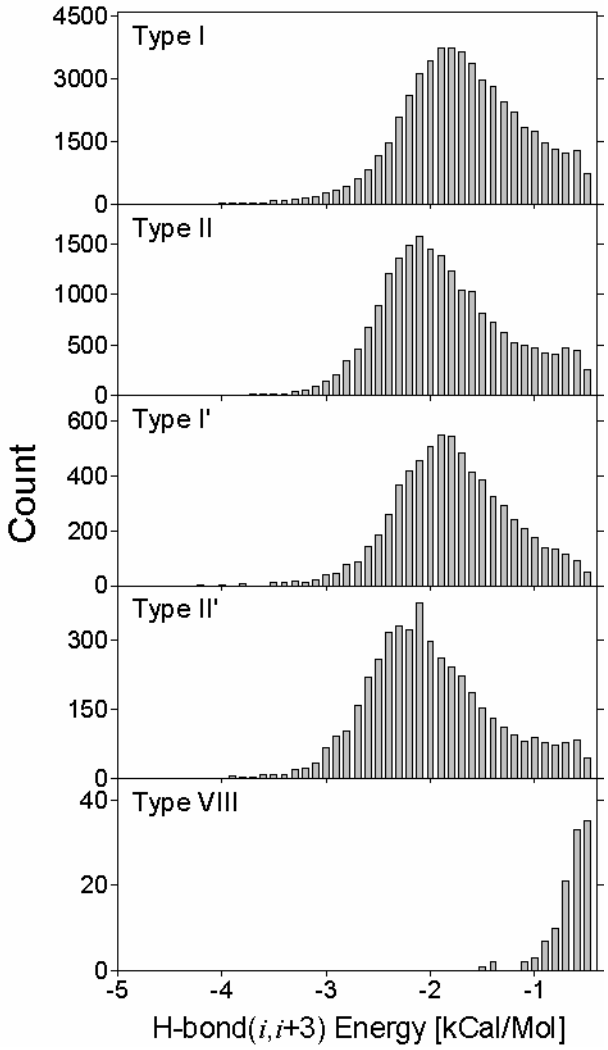

Fig. S3 Histogram of the electrostatic interaction energy (Kabsch and Sander, 1983) of the intra-turn *bb*(*i*)←*bb*(*i*+3) hydrogen bond in five types of β-turns (Type I, II, I′, II′ and VIII).

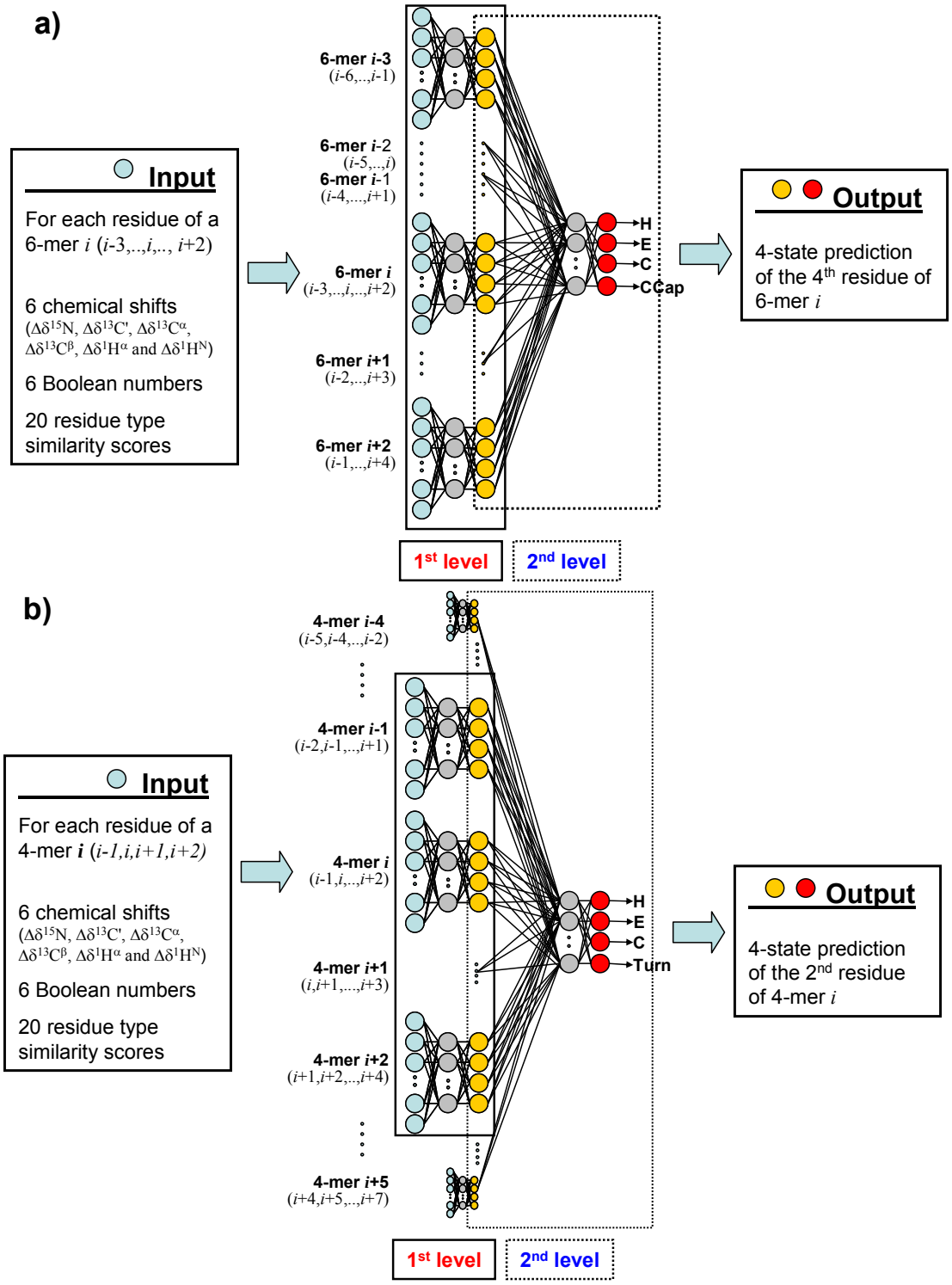

Fig. S4 Architecture of (**a**) the Ccap ANN and (**b**) the β-turn ANN. See Fig. 6 and Methods (main text) for a detailed description.

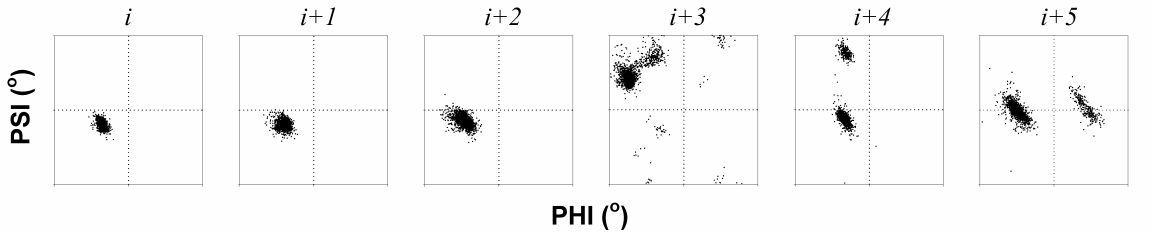

Fig. S5 Positional ϕ/ψ torsion angle distributions for residues in Proline Ccap motifs. In the structure database, 1704 helices are terminated by a Proline Ccap motif, i.e., a Ccap motif "-C3-C2-C1-Ccap-C′-C″-C‴-C⁗-" containing a Proline at the C′ position and either a *bb*(*Ccap*)←*bb*(C″) H-bond or a *bb*(*Ccap*)←*bb*(*C⁗*) H-bond. The ϕ/ψ torsion angles of each of the residues from C3 (*i*) to C″ (*i*+5) are plotted. Each panel is plotted with a range of -180° to 180° for both ϕ and ψ angles.

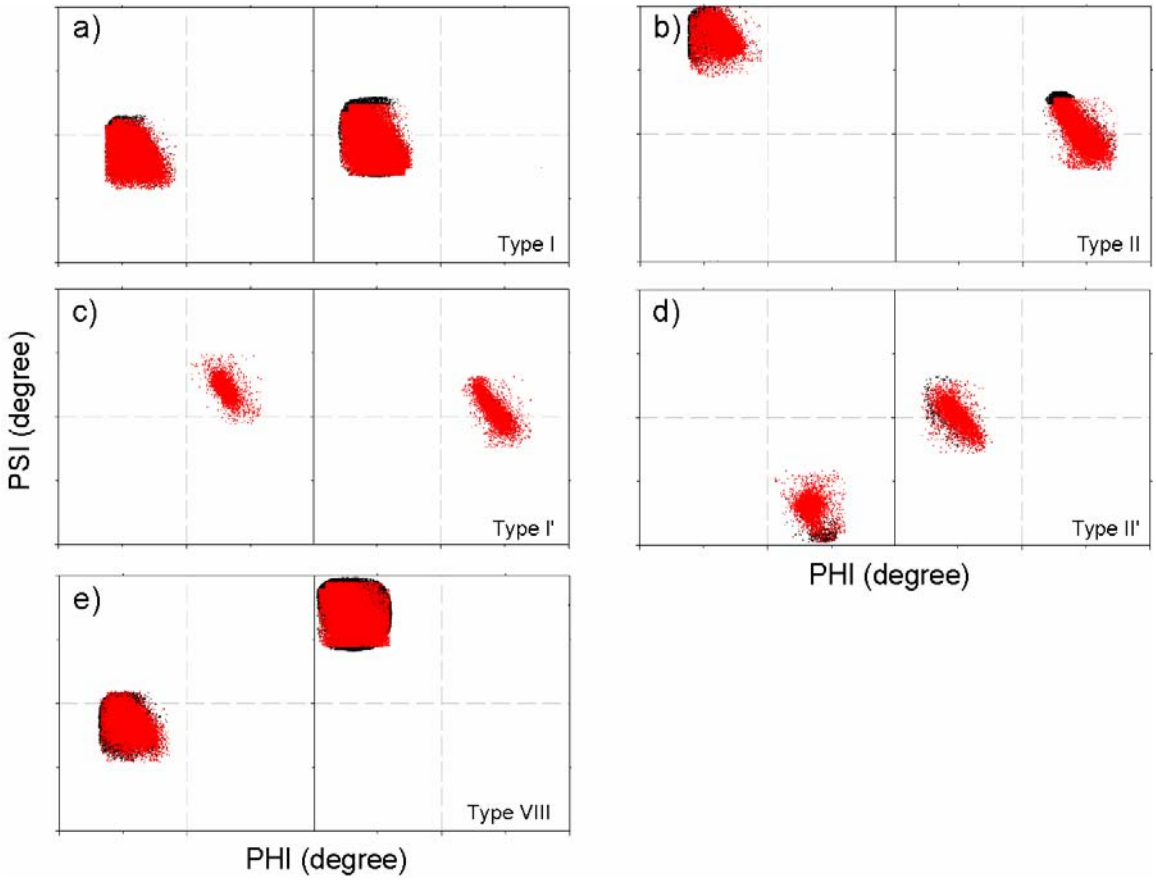

Fig. S6 Comparisons of the ϕ/ψ torsion angle distributions for the two center residues (residue *i*+1 in the left half of each panel; residue *i*+2 in the right half) for five types of β-turns when using Eq. 5 (black dots), with a S_k cutoff value listed in Table S5, and the original definition (red dots). The number of each type of β-turn observed in the structure database is listed in Table S6.

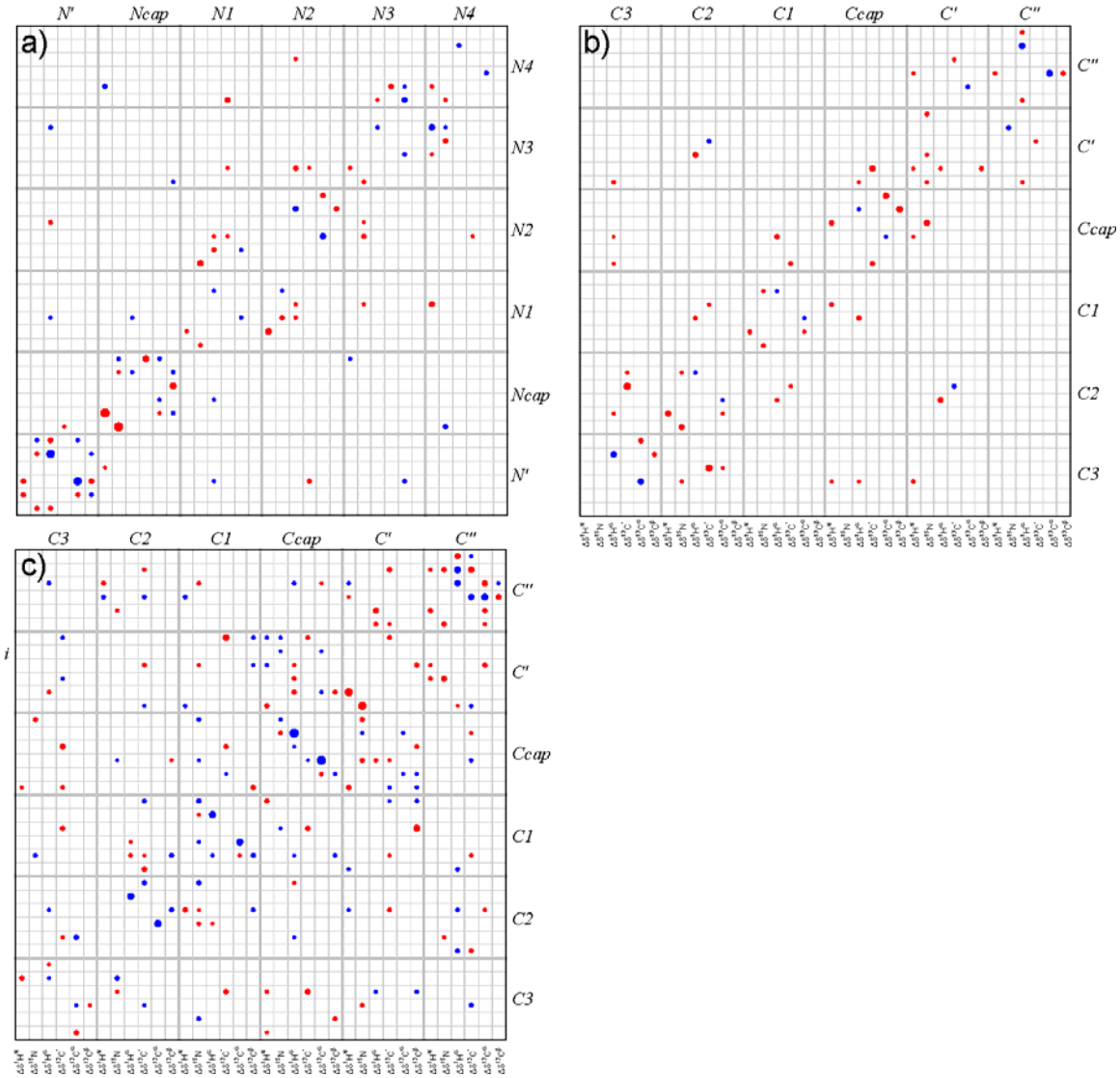

Fig. S7 Plots of correlation coefficients observed between the secondary chemical shifts of backbone and $^{13}C^{\beta}$ nuclei for the six residues in Ncap and Ccap motifs. For each of six residues of (**a**) Ncap box motifs, and (**b**) ideal Schellman Ccap and (**c**) αL Ccap motifs, the Pearson's correlation coefficient, R_P, between the secondary $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C'$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts is presented for each combination that has a significant degree of correlation ($|R_P| \geq 0.2$ for (**a**,**b**), and ≥ 0.25 for (**c**)), with the absolute value of R_P represented by the size of the filled circle (ranging from > 0.6 to > 0.2). Positive R_P values are shown in red, negative values in blue.

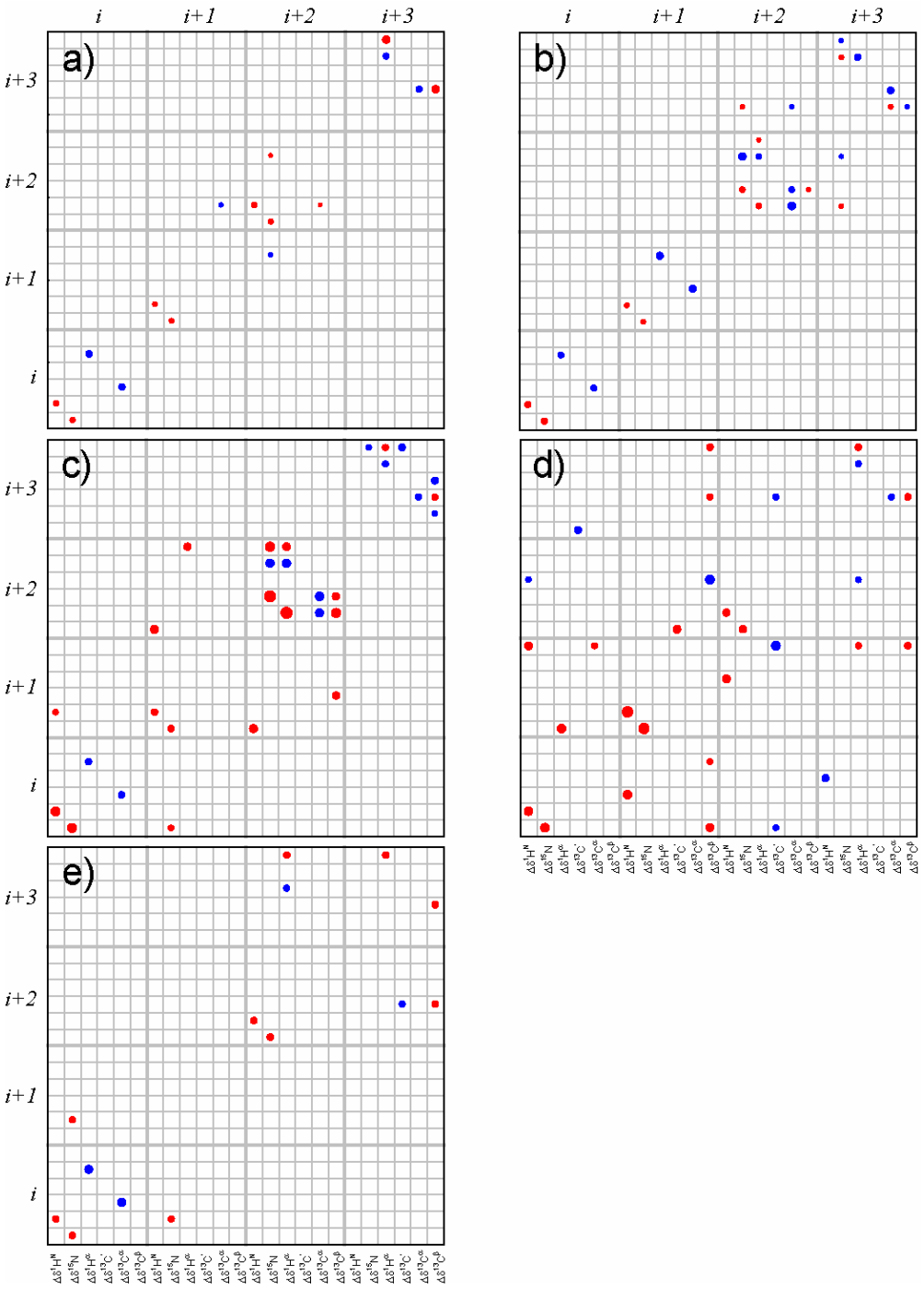

Fig. S8 Plots of correlation coefficients observed between the secondary chemical shifts of backbone and $^{13}C^{\beta}$ nuclei for the four residues in five types of β-turns. For each of four residues (positions *i* to *i*+3) of five types of β-turns [(**a**) I; (**b**) II; (**c**) I′; (**d**) II′ and (**e**) VIII] the Pearson's correlation coefficient, R_P, between the secondary $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$, $^{13}C'$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts is presented for each combination that has a significant degree of correlation ($|R_P| \geq 0.2$), with the absolute value of R_P represented by the size of the filled circle (ranging from > 0.6 to > 0.2). Positive R_P values are shown in red, negative values in blue.