Supplementary material

Protein backbone chemical shifts from searching a database for torsion angles and sequence homology

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.

Table S1. Proteins contained in the SPARTA database

Protein Name (BMRB #)	*RCSB code*	*Resolution(Å)*	*Protein Name (BMRB #)*	*RCSB code*	*Resolution(Å)*
4060	1JRL	1.95	5258	1M9F	1.73
4082	1FIL	2.00	5350	5PNT	2.20
4084	1L3K	1.10	5404	1XEO	1.30
4091	1BFG	1.60	5462	1G6H	1.60
4102	1A7T	1.85	5484	1EB0	1.85
4115	1EMV	1.70	5485	1H7M	1.96
4132	1U9B	2.00	5571	1F4P	1.30
4152	4AKE	2.20	5579	1CRB	2.10
4161	2IRF	2.20	5623	1MN8	1.00
4162	1EPF	1.85	5688	1KWI	2.19
4193	1E4V	1.85	5712	1IPB	2.00
4215	1C76	2.25	5721	1ND4	2.10
4236	1NM5	2.40	5738	1NCX	1.80
4267	1L6M	2.40	5741	1RX4	2.20
4272	1SFC	2.40	5756	1N0S	2.00
4297	1B79	2.30	5761	1IU1	1.80
4340	1JV4	1.75	5789	1A6J	2.35
4354	1FQA	1.90	5792	1EW4	1.40
4378	1G8I	1.90	5799	1J97	1.50
4388	1I5Z	1.90	5824	1V2Z	1.80
4397	1F2L	2.00	5843	1Q4R	1.90
4417	1QMR	2.15	5892	MOLM	1.85
4421	1GXQ	2.00	5898	1UOH	2.00
4438	1C44	1.80	5899	1PEY	2.25
4553	1XUO	1.80	5921	1VC1	2.00
4558	1K46	2.20	5935	1TIP	2.20
4567	1FZY	1.90	5958	1UGM	2.05
4568	1DTI	1.70	6057	1NG2	1.70
4573	1DHN	1.65	6074	1H70	1.80
4636	1F3V	2.00	6075	1JL3	1.60
4637	1LE8	2.30	6090	1FF3	1.90
4679	830C	1.60	6122	1SMX	1.80
4718	1BM9	2.00	6150	1VYF	1.85
4735	1F35	2.30	6184	1J54	1.70
4765	1LOP	1.80	6317	1UTX	1.90
4786	1BYL	2.30	6332	1M45	1.65
4885	1A1V	2.20	6334	1FD9	2.41
4897	1J8R	1.80	6363	1TTZ	2.11
4936	1QJ8	1.90	6375	1U07	1.13
4940	1R5R	1.60	6399	1HH8	1.80
4983	1CMX	2.25	6416	1ZLQ	1.80
4987	1MPD	2.30	6418	1UHI	1.80
4989	1F80	2.30	6468	1ZYJ	2.00
4998	1F9F	1.90	6494	1DBF	1.30
5008	1SCE	2.20	6503	1F2F	1.70
5071	1NCX	1.80	6585	1VJH	2.10
5130	1IWT	1.40	6695	1GAW	2.20
5182	1M5E	1.46	6722	1U2P	1.90
5194	1TJM	1.18	6744	2BX6	2.10
5209	1JN3	2.35	6754	1QAV	1.90
5226	2BKY	1.70	6760	3FAP	1.85

Table S1. (Continued)

Protein Name (BMRB #)	*RCSB ID*	*Resolution (Å)*	*Protein Name (BMRB #)*	*RCSB ID*	*Resolution (Å)*
6779	1ZE3	1.84	*dehydrase*	1MKA	2.00
6824	1XAK	1.80	*dna_polym_B(5208)*	1BPB	2.30
6877	1BY9	2.20	*endonucleaseV(5244)*	2END	1.45
6922	2D3D	1.60	*epsin_NTH(4959)*	1EYH	1.56
6923	2AWG	1.60	*epsps_complexed(4848)*	1G6S	1.50
6932	1KBL	1.94	*fkbp(4077)*	1BKF	1.60
6940	1IHO	1.70	*frataxin(4342)*	1EKG	1.80
6956	2F8K	2.00	*fyve(4898)*	1JOC	2.20
6970	1MH1	1.38	*gabarap(5058)*	1GNU	1.75
6980	2D58	1.90	*gag(5316)*	1HIW	2.30
6983	1SK7	1.60	*gyraseB(5218)*	1KZN	2.30
6986	1WVH	1.50	*hav(4836)*	1QA7	1.90
7003	1BYQ	1.50	*hca_I(4022)*	1HCB	1.60
7107	1B56	2.05	*hpr*	-	2.00
7115	1L8B	1.80	*hpt(4857)*	2A0B	1.57
7130	1E0C	1.80	*hsp90(5355)*	1AM1	2.00
7138	1QFJ	2.20	*interleukin_1beta(1061)*	4I1B	2.00
7216	1NEY	1.20	*interleukin_4(4094)*	1RCB	2.25
7234	1O08	1.20	*interleukin5(5373)*	1HUL	2.40
7250	2GOL	2.20	*lactamase(4102)*	1ZNB	1.85
HIVprotease	1Z6E	1.80	*lectin_TC14(4782)*	1BYF	2.00
Hmet(4336)	1HTP	2.20	*mad2(5299)*	1GO4	2.05
IIB_chitobiose(4955)	1IIB	1.80	*malate(5471)*	1D8C	2.00
IIIglc	1F3G	2.10	*maxacal*	1SVN	1.40
L25(4395)	1DFU	1.80	*mbp*	1DMB	1.80
MotA_NTD(4957)	1BJA	2.19	*mia(5220)*	1I1J	1.39
RNAse_A(4031)	3RN3	1.45	*mmp1(4064)*	1HFC	2.20
RNAse_SA(4259)	1RGE	1.15	*mmp*	-	1.56
Vesl_1S(4766)	1DDW	1.70	*ns1(4317)*	1AIL	1.90
YopH(4961)	1HUF	2.00	*omtky3(4864)*	2OVO	1.50
adrenodoxin(4566)	1AYF	1.85	*ospA(4076)*	1OSP	1.95
alpha_LP	2ALP	1.70	*pin1(5305)*	1PIN	1.35
apo_HPPK(4299)	1HKA	1.50	*plastocyanin(4019)*	1PLC	1.33
apo_lfabp(4098)	1LFO	2.30	*profilin*	1ACF	2.00
app_d49(4078)	1VAP	1.60	*ptp1b(5474)*	1G7F	1.80
arsenate_reduct(4944)	1JF8	1.12	*recoverin(5332)*	1REC	1.90
barnase(4964)	1BNJ	2.10	*rOnc(4371)*	-	1.70
bpti	5PTI	1.00	*rubredoxin(4050)*	1FHM	1.50
cadherin(4380)	1EDH	2.00	*s100b(5206)*	1MHO	2.00
calbindin(390)	4ICB	1.60	*sap_sh2d1a(5211)*	1D4T	1.10
calmodulin_complexed(547)	1CLL	1.70	*snase*	1SNC	1.65
calmodulin(1634)	1CDL	2.00	*sNTnC(4401)*	1AVS	1.75
ci2(4974)	2CI2	2.00	*suc1(5009)*	1PUC	1.95
crabp_II(4186)	1CBS	1.80	*syntaxin_1A(4198)*	1EZ3	1.90
crd_galectin_3(4909)	1A3K	2.10	*sytI_C2A(4039)*	1RSY	1.90
cspA(4296)	1MJC	2.00	*thioredoxin_red*	1ERT	1.70
cutinase(4101)	1CEX	1.00	*ubiquitin*	1UBQ	1.80
cvn	3EZM	1.50	*vegf(5185)*	2VPF	1.93
cyclophilin	2CPL	1.63	*xylanase(5352)*	1H4H	1.90

78 proteins (in italic font) are also contained in the TALOS database, extended in 2005 by G. Cornilescu.

Table S2. Residue-type-specific distribution of secondary chemical shifts contained in the SPARTA database.

Res. Name	^{15}N	$^{1}H^{N}$	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$
A	1883	1629	1292	1895	1860	1579
C	210	197	141	208	194	163
D	1493	1355	1066	1511	1464	1246
E	1756	1579	1269	1763	1700	1467
F	1007	883	729	1015	977	818
G	1659	1414	1147	1697	-	1408
H	488	439	357	501	474	410
I	1503	1353	1054	1512	1457	1245
K	1607	1435	1193	1626	1575	1340
L	2222	1995	1561	2240	2153	1855
M	483	430	348	488	458	391
N	977	835	695	1002	967	839
P	-	-	540	846	810	691
Q	911	816	659	921	880	767
R	1093	1012	746	1104	1031	892
S	1287	1119	977	1300	1223	1059
T	1379	1197	1056	1386	1316	1137
V	1763	1571	1307	1780	1716	1471
W	320	293	215	324	302	278
Y	809	735	538	810	774	670
c	103	88	98	100	73	79
Total	22953	20375	16988	24029	21404	19805

Table S3. Residue similarity factors, $\Delta_{ResType}$, used by SPARTA in Eq 1, main text.

Res	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	0	4	2	2	2	3	3	8	4	4	4	4	3	4	8	2	3	4	4	3
R	4	0	2	2	3	1	1	8	4	3	3	1	3	4	8	3	3	4	4	3
N	2	2	0	0	2	1	1	6	4	2	2	2	2	4	8	2	3	4	4	3
D	2	2	0	0	2	1	1	6	4	2	2	2	2	4	8	2	3	4	4	3
C	2	3	2	2	0	3	3	6	4	4	4	4	1	4	8	2	3	4	4	3
Q	3	1	1	1	3	0	1	6	3	3	3	1	3	3	8	3	3	3	3	3
E	3	1	1	1	3	1	0	6	3	3	3	1	3	3	8	3	3	3	3	3
G	8	8	6	6	6	6	6	0	6	6	6	6	6	6	12	6	6	6	6	6
H	4	4	4	4	4	3	3	6	0	3	3	3	3	1	8	2	3	1	1	3
I	4	3	2	2	4	3	3	6	3	0	1	2	3	4	8	3	2	4	4	1
L	4	3	2	2	4	3	3	6	3	1	0	2	3	4	8	3	2	4	4	2
K	4	1	2	2	4	1	1	6	3	2	2	0	3	4	8	4	3	4	4	3
M	3	3	2	2	1	3	3	6	3	3	3	3	0	4	8	2	3	4	4	3
F	4	4	4	4	4	3	3	6	1	4	4	4	4	0	8	3	3	1	0	3
P	8	8	8	8	8	8	8	12	8	8	8	8	8	8	0	8	8	8	8	8
S	2	2	2	2	2	3	3	6	2	3	3	4	2	3	8	0	1	4	4	3
T	3	3	3	3	3	3	3	6	3	2	2	3	3	3	8	1	0	1	1	1
W	4	4	4	4	4	3	3	6	1	4	4	4	4	1	8	2	4	0	1	3
Y	4	4	4	4	4	3	3	6	1	4	4	4	4	0	8	2	4	1	0	3
V	3	3	3	3	2	3	3	6	3	1	2	3	3	3	8	2	3	3	3	0

Table S4. Original random coil shifts used for calculating secondary shifts.[a]

Res. Name	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	^{15}N	$^{1}H^{N}$
A	4.32	52.30	19.00	177.80	123.80	8.15
R	4.34	56.10	30.30	176.30	120.50	8.27
N	4.74	52.80	37.90	175.20	118.70	8.38
D	4.64	54.00	40.80	176.30	120.40	8.37
C	4.55	56.90	28.90	174.60	118.80	8.23
c	4.71	55.40	43.70	174.60	118.60	8.23
Q	4.34	56.10	28.40	176.00	119.80	8.27
E	4.35	56.40	29.70	176.60	120.20	8.36
G	3.96	45.10	-	174.90	108.80	8.29
H	4.73	54.50	27.90	173.30	118.20	8.28
I	4.17	61.30	38.00	176.40	119.90	8.21
L	4.34	55.10	42.30	177.60	121.80	8.23
K	4.32	56.50	32.50	176.60	120.40	8.25
M	4.48	55.30	32.60	176.30	119.60	8.29
F	4.62	58.00	39.00	175.80	120.30	8.30
P	4.42	63.10	31.70	177.30	-	-
S	4.47	58.20	63.20	174.60	115.70	8.31
T	4.35	62.10	69.20	174.70	113.60	8.24
W	4.66	57.70	30.30	176.10	121.30	8.18
Y	4.55	58.10	38.80	175.90	120.30	8.28
V	4.12	62.30	32.10	176.30	119.20	8.19

[a] $^{13}C^{\alpha}$, $^{13}C^{\beta}$ random coil values are from Spera S, Bax A (1991) J Am Chem Soc, 113:5491–5492. Others are from Wishart et al. (1995) J Biomol NMR, 5:67–81. These values are prior to modification by the adjustment values of Table S5.

Table S5. Random coil chemical shift adjustment values used by SPARTA when deriving secondary chemical shifts.

Res. Name	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	^{15}N	$^{1}H^{N}$
A	0	-0.26	0	-0.22	-0.98	0
R	0	0	0	0	0	0
N	0	0.25	0.59	0.31	0	0
D	0	0	0	0	0	0
C	-0.09	2.12	-1.16	0.39	0.50	0
c	-0.14	0.93	-1.19	-0.51	0	0
Q	0	-0.30	0.34	0	0	0
E	0	0	0	0	0	0
G	0	0	0	0	0	0
H	-0.17	1.55	1.86	2.03	0.87	0
I	0	0	0	-0.28	0.67	0
L	0	-0.26	-0.24	-0.73	-0.58	0
K	0	-0.21	0	0	0	0
M	0	0	0	0	0	0
F	-0.09	0	0	0	-0.35	0
P	0	0	0	0	0	0
S	0	0	0	0	0	0
T	0	0	-0.33	0	0.95	0
W	0	-0.36	-0.84	0.42	0	0
Y	0	0	-0.34	0	0	0
V	0	0	-0.28	-0.21	0.71	0

Table S6. SPARTA random coil shift adjustments to account for the preceding residue.

Res. Name	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	^{15}N	$^{1}H^{N}$
A	0	0	0	0	-0.20	0
R	0	0	0	0	1.45	0
N	0	0	0	0	1.15	0
D	0	0	0	0	1.18	0
C	0	0	0	0	2.43	0
c	0	0	0	0	2.55	0
Q	0	0	0	0	1.57	0
E	0	0	0	0	1.80	0
G	0	0	0	0	1.07	0
H	-0.09	0	0	0	1.20	0
I	0	0	0	0	4.14	0
L	0	0	0	0	1.50	0
K	0	0	0	0	1.55	0
M	0	0	0	0	1.33	0
F	0	0	0	0	1.25	0
P	0	0	0	0	0.60	0
S	0	0	0	0	2.59	0
T	0	0	0	0	3.20	0
W	0	0	0	0	1.64	0
Y	0	0	0	0	1.75	0
V	0	0	0	0	3.83	0

Table S7. SPARTA random coil shift adjustments to account for the following residue.

Res. Name	$^{1}H^{\alpha}$	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	$^{13}C'$	^{15}N	$^{1}H^{N}$
A	0	0	0	0	0	0
R	0	0	0	0	0	0
N	0	0	0	0	0	0
D	0	0	0	-0.32	0	0
C	0	0	0	0	-0.40	0
c	-0.08	0	-0.31	0	-0.86	0
Q	0	0	0	0	0	0
E	0	0	0	0	0	0
G	0	0	0	0.39	0	0
H	0	0	0	0	0	0
I	0	0	0	0	-0.40	0
L	0	0	0	0	0	0
K	0	0	0	0	0	0
M	0	0	0	0	0	0
F	0	0	0	0	0	0
P	0	0	0	0	0	0
S	0	0	0	0	0.48	0
T	0	0	0	0	0	0
W	0	0	0	0	-0.50	0
Y	0	0	0	-0.33	0	0
V	0	0	0	0	0	0

Fitting parameters between the prediction accuracy σ and precision σ_{20} in Figure 6, main text:

$$\sigma(^{15}N) = 0.46 \times \sigma_{20} + 1.34 \quad [s1]$$

$$\sigma(^{1}H^{\alpha}) = 0.45 \times \sigma_{20} + 0.14 \quad [s2]$$

$$\sigma(^{1}H^{N}) = 0.46 \times \sigma_{20} + 0.28 \quad [s3]$$

$$\sigma(^{13}C^{\alpha}) = 0.63 \times \sigma_{20} + 0.38 \quad [s4]$$

$$\sigma(^{13}C^{\beta}) = 0.61 \times \sigma_{20} + 0.45 \quad [s5]$$

$$\sigma(^{13}C') = 0.38 \times \sigma_{20} + 0.69 \quad [s6]$$

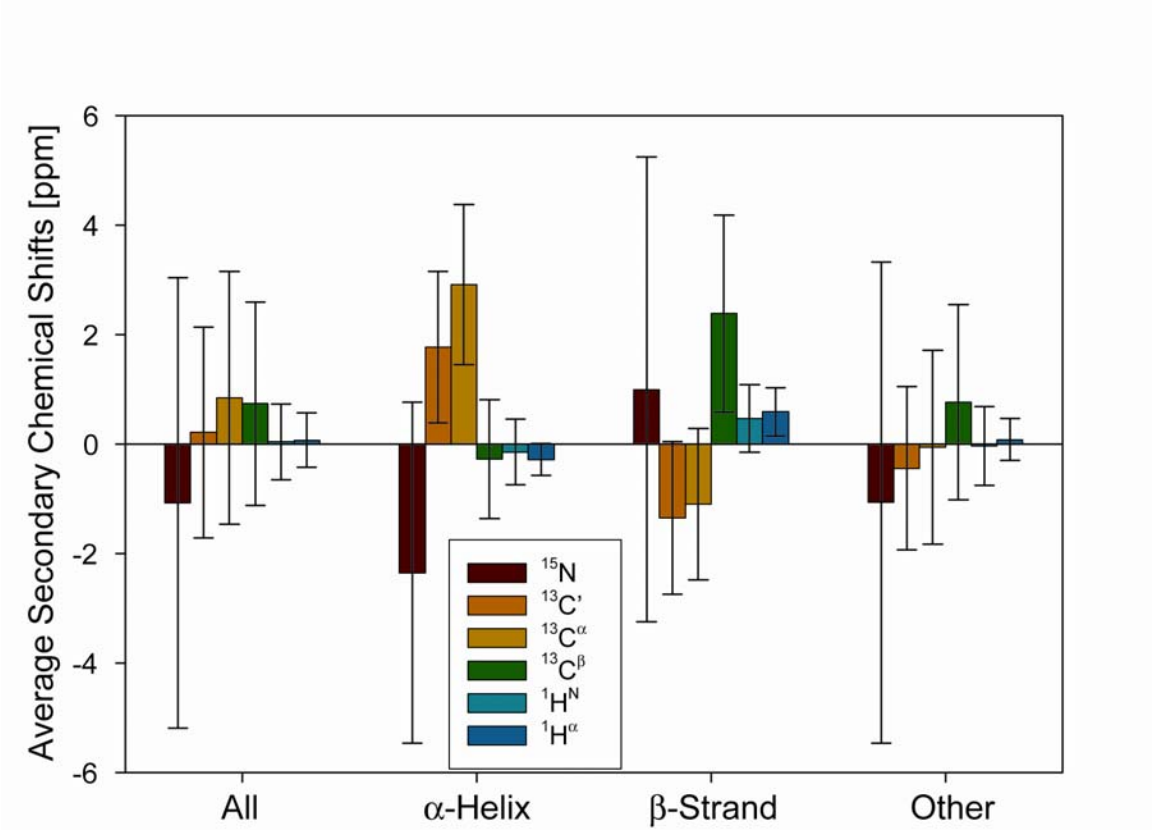

Figure S1. Distribution of secondary chemical shifts in the SPARTA database ("All": for all residues in the database; "α-Helix": for residues in α-helix; "β-Strand": for residues in β-strand, "Other": for residues not in α-helix or β-strand).

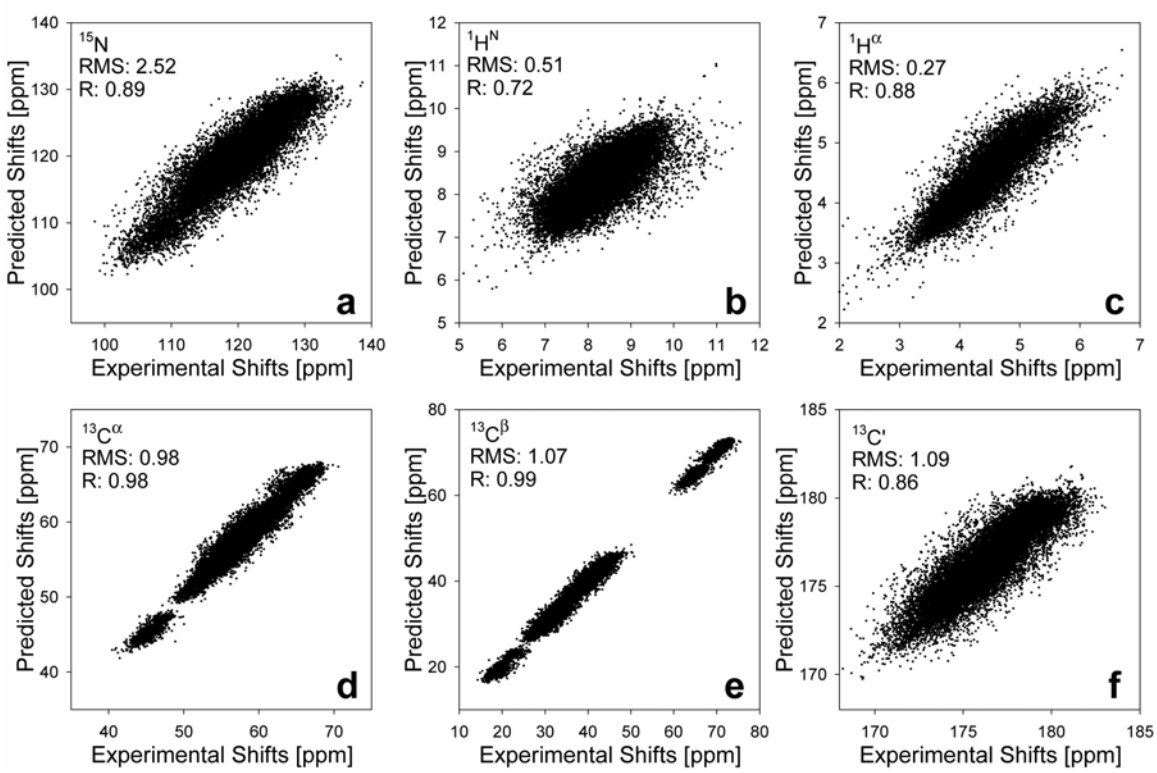

Figure S2. Scatter plots comparing experimental and SPARTA-predicted chemical shifts for backbone ^{15}N, $^{1}H^{N}$, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\alpha}$ and $^{13}C'$ nuclei. The RMS deviation (in ppm) and correlation coefficient (R) between the experimental and SPARTA-predicted shifts are indicated. For $^{1}H^{N}$ and $^{1}H^{\alpha}$ shifts, the SPARTA-predicted shifts include the hydrogen bond corrections for $^{1}H^{N}$ and $^{1}H^{\alpha}$ atoms involved in intramolecular H-bonds.

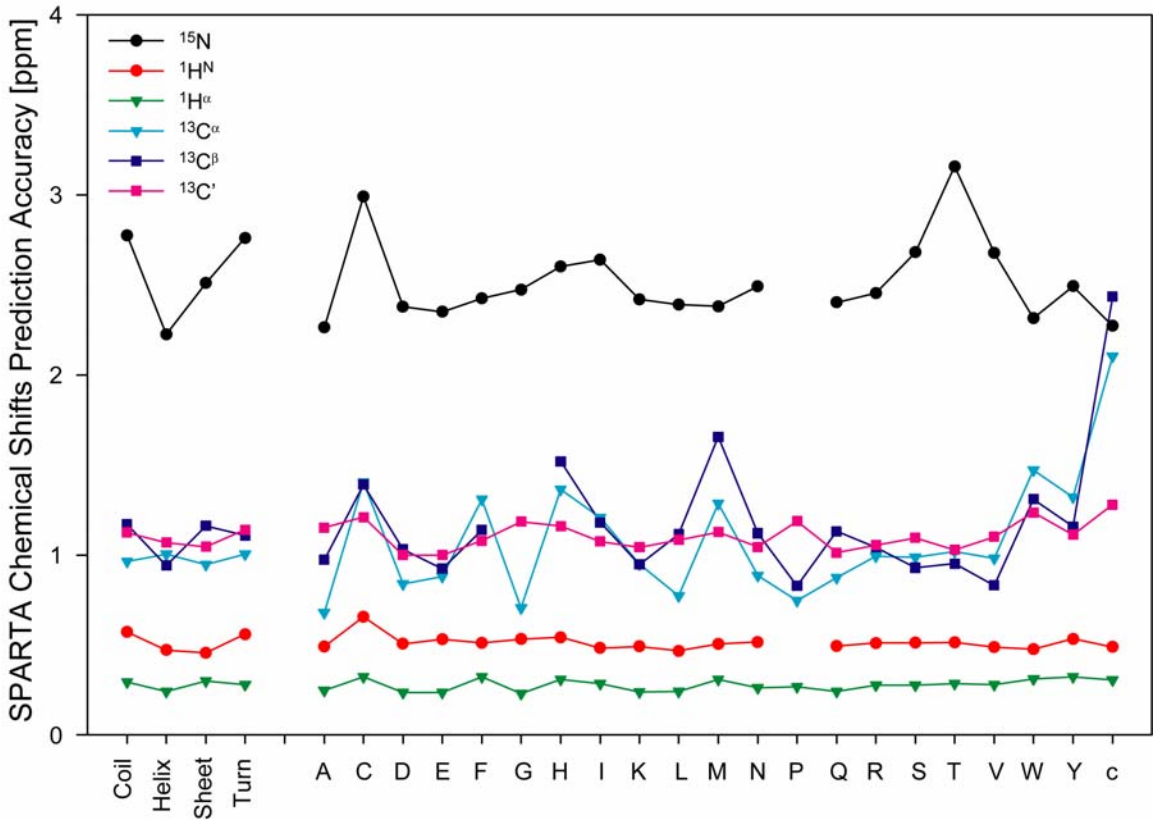

Figure S3. Accuracy of SPARTA chemical shift predictions for each secondary structure type and amino acid type (c: oxidized Cys).

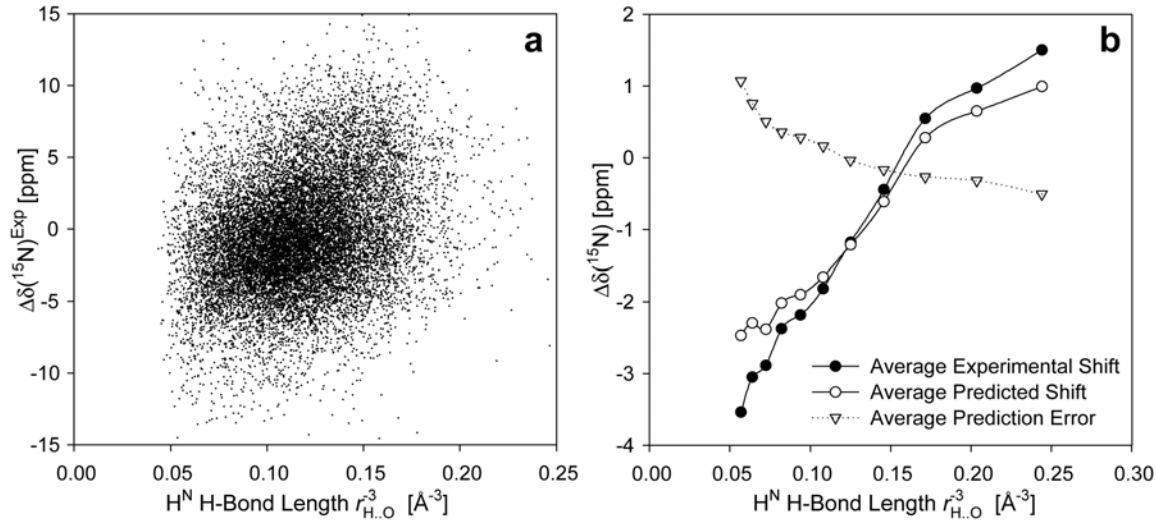

Figure S4. Plots of ^{15}N secondary chemical shift *versus* the inverse cube of the hydrogen bond length of the attached $^{1}H^{N}$ atoms. Only the secondary chemical shifts of the ^{15}N atoms with residual solvent exposure <0.3 are plotted. (**a**): Scatter plot between experimental secondary chemical shifts, $\Delta\delta(^{15}N)^{Exp}$, and $r_{H..O}^{-3}$ of all hydrogen-bonded ^{15}N atoms in the SPARTA database. (**b**): Plot of the average ^{15}N secondary chemical shifts $\Delta\delta(^{15}N)$ binned according to hydrogen bond length of the attached $^{1}H^{N}$ atom, $r_{H..O}^{-3}$. The bin size is 0.1 Å for, and only bins with >50 data are plotted.

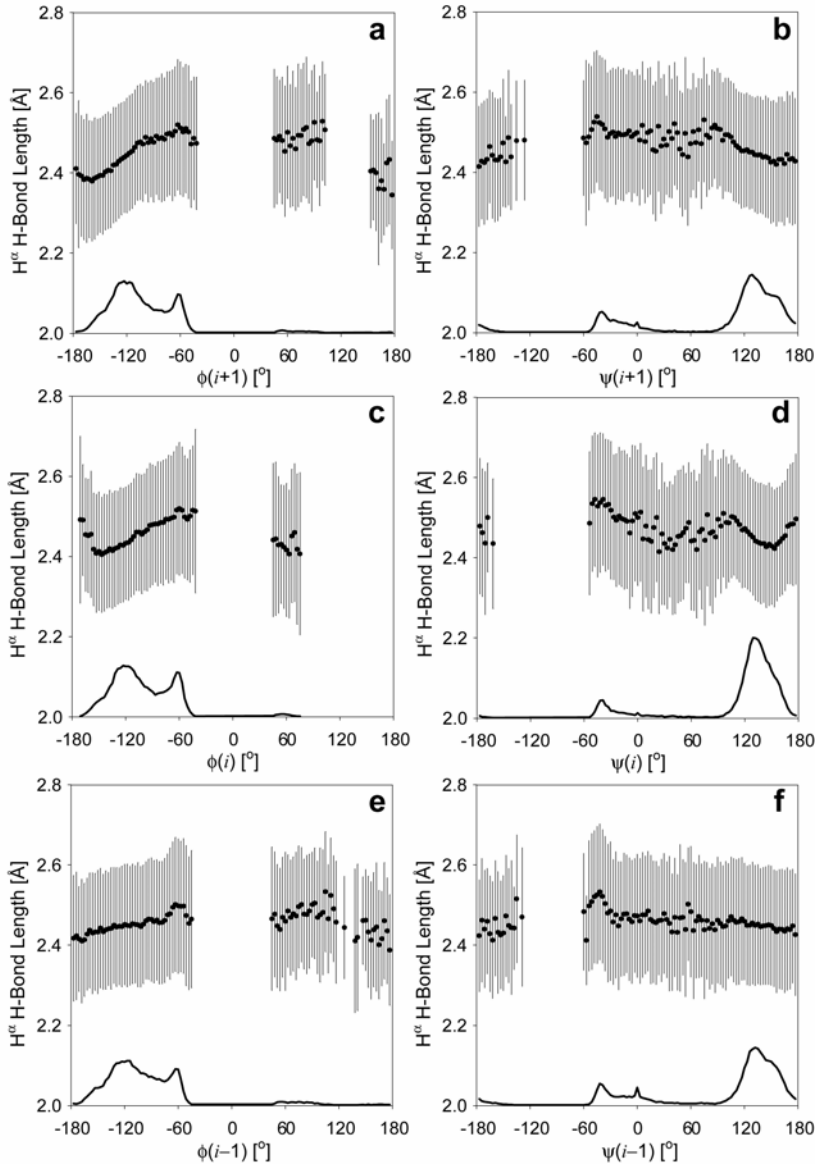

Figure S5. Plots of $^{1}H^{\alpha}$ hydrogen bond length *versus* ϕ/ψ torsion angles for 1197 proteins with high resolution X-ray coordinates (≤ 1.6Å). Horizontal axes represent the ϕ/ψ torsion angles for the following (*i*+1, **a** and **b**), center (*i*, **c** and **d**) or preceding (*i*–1, **e** and **f**) residue. Vertical axes represent the average (filled circle) and the standard deviation (vertical bar) of the distribution of $^{1}H^{\alpha}$ hydrogen bond lengths within each ϕ or ψ bin. The bin size is 3º for ϕ and ψ angles, and only bins with >20 data are plotted. The normalized density for each bin is shown at the bottom of each panel.

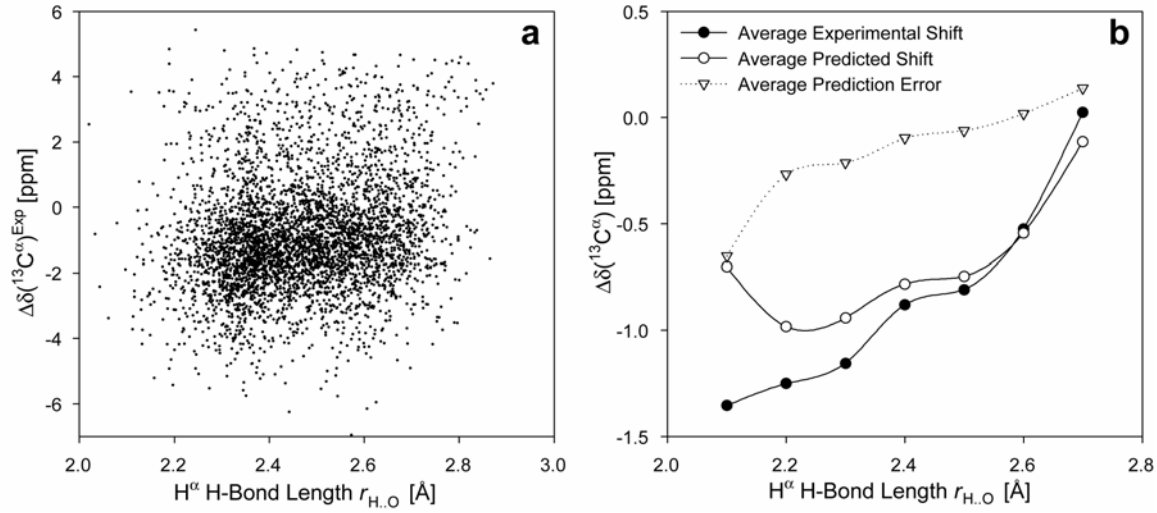

Figure S6. Plots of secondary chemical shifts *versus* hydrogen bond lengths for $^{13}C^{\alpha}$ atoms. (**a**): Scatter plot of experimental secondary chemical shift, $\Delta\delta(^{13}C^{\alpha})^{Exp}$ *versus* hydrogen bond length, $r_{H..O}$, for all $^{13}C^{\alpha}$ atoms with attached H-bonded H^{α} atoms in the SPARTA database. (**b**): Plot of the average $^{13}C^{\alpha}$ secondary chemical shifts $\Delta\delta(^{13}C^{\alpha})$, binned according to hydrogen bond length, $r_{H..O}$, using a bin size of 0.1 Å. Only bins with >50 data are plotted.

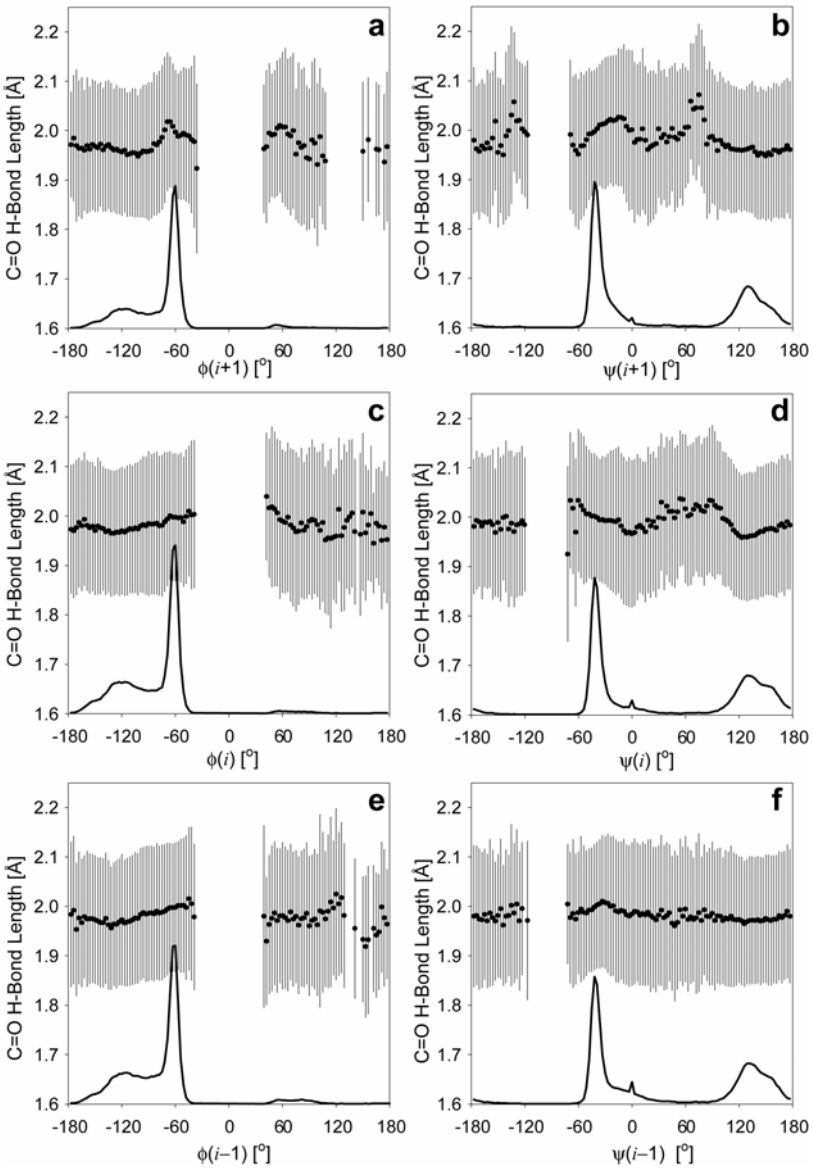

Figure S7. Plots of C=O hydrogen bond length *versus* ϕ/ψ torsion angles for 1197 proteins with high resolution X-ray coordinates (≤ 1.6Å). Horizontal axes represent the ϕ/ψ torsion angles for the following (*i*+1, **a** and **b**), center (*i*, **c** and **d**) or preceding (*i*–1, **e** and **f**) residue. Vertical axes represent the average (filled circle) and the standard deviation (vertical bar) of the distribution of hydrogen bond lengths of C=O with residual solvent exposure < 0.3 within each ϕ or ψ bin. The bin size is 3º for ϕ and ψ angles, and only bins with >30 data are plotted. The normalized density of the bins is shown as a solid line at the bottom of each panel.

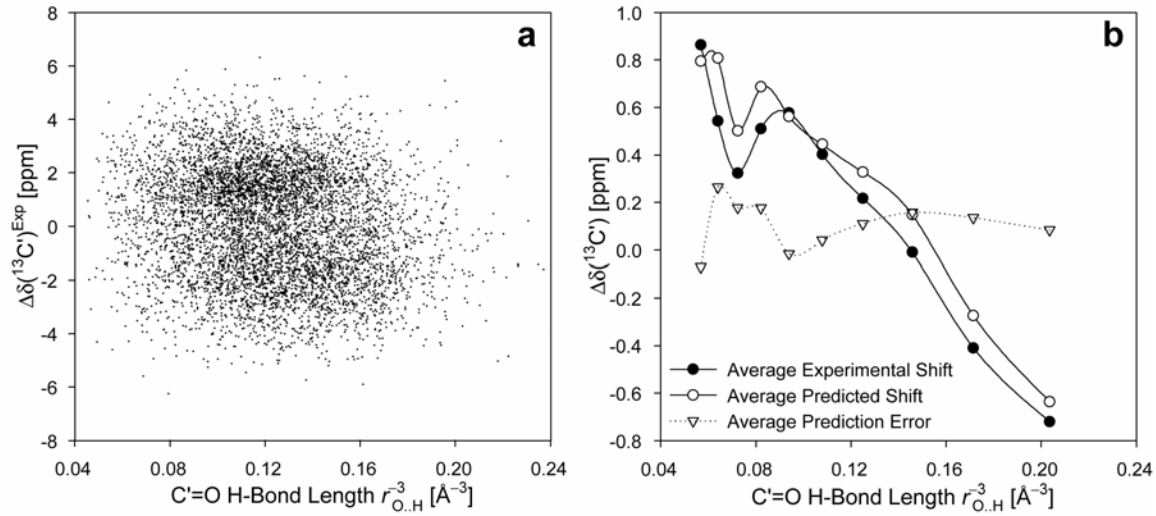

Figure S8. Plots of ^{13}C' secondary chemical shifts versus the inverse cube of the hydrogen bond length of the attached carbonyl oxygen atoms. In total, 12293 carbonyl oxygen atoms (out of 19803 from the database) are identified as intramolecularly hydrogen bonded, of which 6786 (49%) could be identified as not solvent exposed with a residual solvent exposure cutoff value of 0.2. Only those not solvent exposed are plotted. (**a**): Scatter plot between experimental secondary chemical shifts, $\Delta\delta(^{13}C')^{Exp}$, of all hydrogen-bonded ^{13}C' atoms in the SPARTA database and the hydrogen bond lengths of the attached carbonyl oxygen atoms, $r_{O..H}$. (**b**): Plot of the average ^{13}C' secondary chemical shifts $\Delta\delta(^{13}C')$ binned according to H-bond length of the attached carbonyl oxygen atoms (with residual solvent exposure < 0.2), $r_{O..H}$. The bin size of hydrogen bond length is 0.1 Å, and only bins with >50 data are plotted.

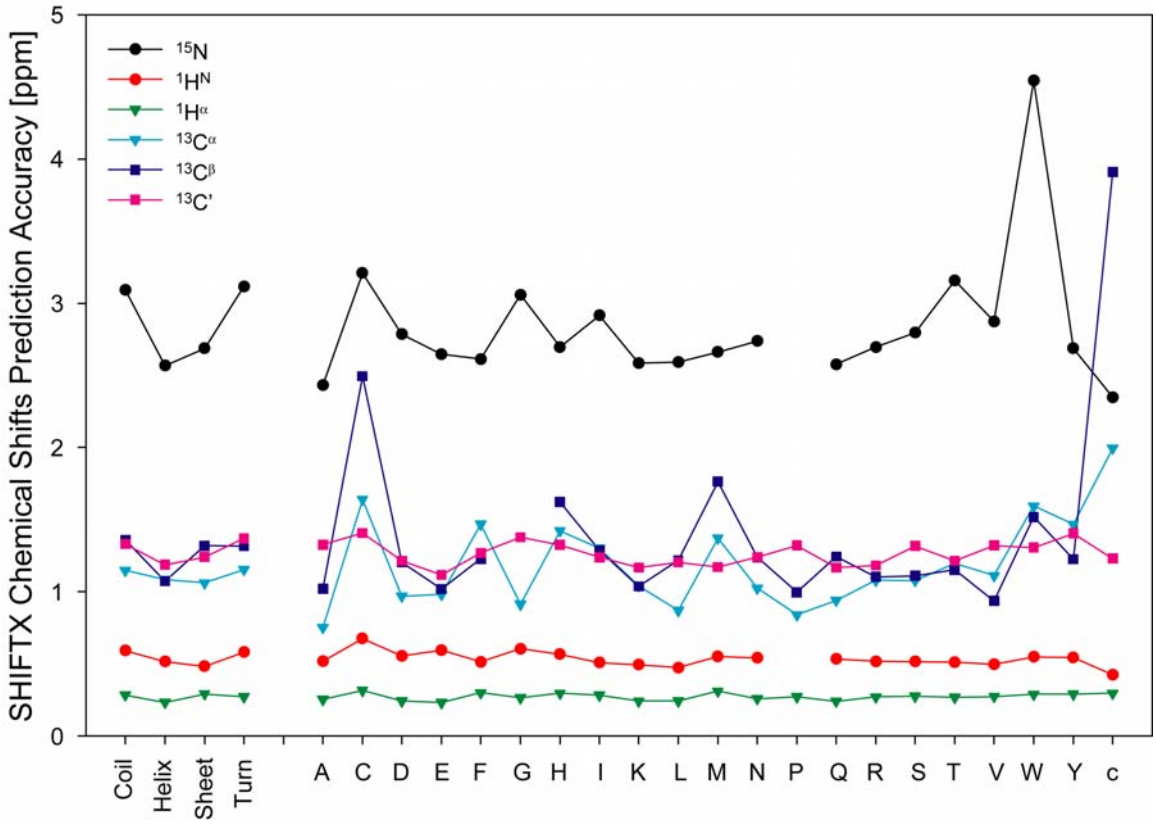

Figure S9. Accuracy of SHIFTX predicted secondary chemical shifts for each secondary structure type and residue type for the 200 proteins in the SPARTA database (c: oxidized Cystine).