

Supporting information for:

Automated Fragmentation Polarizable

Embedding DFT Calculations of NMR Shielding

Constants of Proteins with Application to

Chemical Shift Predictions

Casper Steinmann,^{*,†,‡} Lars Andersen Bratholm,[¶] Jógvan Magnus Haugaard
Olsen,[‡] and Jacob Kongsted[‡]

*[†]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol
BS8 1TS, United Kingdom*

*[‡]Department of Physics, Chemistry and Pharmacy, University of Southern Denmark,
DK-5230 Odense M, Denmark*

[¶]Department of Chemistry, University of Copenhagen, DK-2100 Copenhagen, Denmark

E-mail: css@sdu.dk

1 Pearson r -values

Tables of mean correlation values, \bar{r} , when comparing models **1**, **2** and **3** to model **4**. The mean correlation values are obtained by averaging the correlation coefficients between models over residues. In Tables S1 to S6 we list atom specific correlation values for carbon atoms.

Table S1: Mean correlation, \bar{r} , of computed C $^{\alpha}$ NMR shielding constants of the smaller models (**1**, **2** and **3**) versus using model **4** for residues in GB3. All NMR shielding constants were computed at the PE-KT3/pcSseg-1 level of theory. This table is the data behind Figure 5 in the manuscript.

	model 1	model 2	model 3
ALA	0.77	0.88	0.81
ASN	0.85	0.97	0.82
ASP	0.78	0.92	0.80
GLN	0.85	0.99	0.87
GLU	0.70	0.92	0.68
GLY	0.39	0.94	0.47
ILE	0.19	0.81	0.54
LEU	0.86	0.99	0.84
LYS	0.70	0.98	0.71
MET	0.60	0.90	0.75
PHE	0.60	0.97	0.53
THR	0.60	0.86	0.62
TRP	0.70	1.00	0.71
TYR	0.80	0.98	0.91
VAL	0.43	0.84	0.54
BB	0.66	0.93	0.71

Table S2: Mean correlation, \bar{r} , of computed C $^{\beta}$ NMR shielding constants of the smaller models (**1**, **2** and **3**) versus using model **4** for residues in GB3. All NMR shielding constants were computed at the PE-KT3/pcSseg-1 level of theory.

	model 1	model 2	model 3
ALA	0.66	0.94	0.79
ASN	0.69	0.95	0.83
ASP	0.76	0.77	0.77
GLN	0.91	0.99	0.91
GLU	0.52	0.90	0.50
ILE	0.29	0.71	0.95
LEU	0.65	0.95	0.69
LYS	0.43	0.95	0.65
MET	0.78	1.00	0.71
PHE	0.08	0.96	0.16
THR	0.17	0.87	0.34
TRP	0.56	0.92	0.56
TYR	0.61	0.93	0.41
VAL	0.54	0.95	0.52
BB	0.51	0.85	0.59

Table S3: Mean correlation, \bar{r} , of computed C^γ NMR shielding constants of the smaller models (**1**, **2** and **3**) versus using model **4** for residues in GB3. All NMR shielding constants were computed at the PE-KT3/pcSseg-1 level of theory.

	model 1	model 2	model 3
ASN	0.82	0.90	0.92
ASP	0.90	0.89	0.91
GLN	0.87	0.89	0.99
GLU	0.87	0.90	0.98
ILE	0.78	0.80	0.67
LEU	0.83	0.93	0.96
LYS	0.96	0.96	0.91
MET	0.99	0.99	0.97
PHE	0.65	0.60	0.72
THR	0.65	0.92	0.87
TRP	0.98	1.00	0.96
TYR	0.69	0.82	0.88
VAL	0.74	0.83	0.77
BB	0.72	0.76	0.77

Table S4: Mean correlation, \bar{r} , of computed C^δ NMR shielding constants of the smaller models (**1**, **2** and **3**) versus using model **4** for residues in GB3. All NMR shielding constants were computed at the PE-KT3/pcSseg-1 level of theory.

	model 1	model 2	model 3
GLN	0.90	0.93	0.98
GLU	0.51	0.57	0.87
ILE	0.99	1.00	1.00
LEU	0.67	0.99	0.63
LYS	0.89	0.93	0.96
PHE	0.97	0.97	0.97
THR	0.00	0.00	0.00
TRP	0.93	0.98	0.97
TYR	0.68	0.72	0.76
BB	0.44	0.47	0.48

Table S5: Mean correlation, \bar{r} , of computed C^ε NMR shielding constants of the smaller models (**1**, **2** and **3**) versus using model **4** for residues in GB3. All NMR shielding constants were computed at the PE-KT3/pcSseg-1 level of theory.

	model 1	model 2	model 3
LYS	0.90	0.94	0.99
MET	0.99	1.00	0.99
PHE	0.82	0.90	0.79
TRP	0.97	0.98	0.99
TYR	0.56	0.63	0.80
BB	0.28	0.30	0.30

Table S6: Mean correlation, \bar{r} , of computed C^ζ NMR shielding constants of the smaller models (**1**, **2** and **3**) versus using model **4** for residues in GB3. All NMR shielding constants were computed at the PE-KT3/pcSseg-1 level of theory.

	model 1	model 2	model 3
PHE	0.87	0.97	0.81
TRP	0.97	1.00	0.98
TYR	0.75	0.87	0.99
BB	0.17	0.19	0.19

2 Cumulative Distribution Function for C^α

In Figure S1A we show the cumulative distribution function (CDF) for C^α NMR shielding constants calculated at the PE-KT3/pcSseg-1 level of theory for all residues for model **2**. In Figure S1B we show a similar CDF for C^α but predicted from ProCS15. The CDF is made by sorting all C^α NMR shielding constants according to magnitude. When selecting snapshots, we follow the procedure outlined below for $K_j = 3$ snapshots: The three snapshots whose predicted chemical shielding constants of residue j correspond to the ones having the 16.7 %, 50 % and 83.3 % highest predicted shielding constants (the mid-points of the three intervals 0-33.3 %, 33.3-66.7 % and 66.7-100 %) as illustrated with gray lines in Figure S1. This

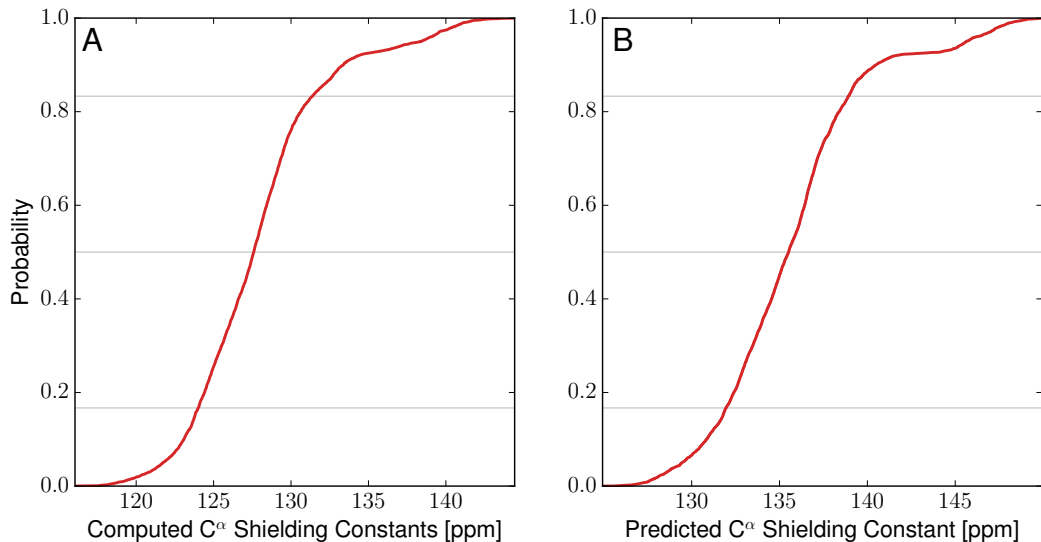


Figure S1: Cumulative distribution function for C^α NMR shielding constants calculated at A) the PE-KT3/pcSseg-1 level of theory for all residues using model **2** or B) predicted using ProCS15.

procedure is repeated for every new value of K_j required for a residue j .

3 ProCS15B Averaging for $\bar{K}_j = 5$

In Table S7 we show which snapshots from our molecular dynamics simulation are to be used for each residue to minimize the variance when computing NMR shielding constants. The predictions are made using ProCS15. Residues that have a large computed variance, such as illustrated in Figure 7 in the manuscript, have more snapshots included to lower the variance in the computed shieldings.

Table S7: List of snapshots which minimizes the variance for each residue. Residue 1 and 56 excluded. Residue counting starts from zero.

Residue	Snapshot
2	11, 13, 25, 27, 70
3	11, 20, 27
4	14, 50, 70, 97
5	14, 19, 83
6	63, 67, 96
7	32, 57, 89
8	22, 23, 26, 37, 83
9	3, 13, 51, 65, 69
10	3, 5, 15, 21, 26, 27, 29, 50, 52, 61, 62, 63, 66, 83, 88, 91, 96
11	8, 13, 22, 29, 32, 35, 48, 50, 51, 54, 71, 93
12	5, 36, 57, 79, 80, 83
13	1, 18, 22, 33, 61, 64, 68, 88
14	9, 26, 54, 68, 79
15	21, 32, 35, 88
16	19, 21, 27, 34, 60
17	6, 35, 38, 46
18	1, 14, 39, 50, 58, 73, 83, 94, 99
19	21, 54, 69, 72
20	61, 94, 98
21	15, 16, 35, 38, 43, 51, 54, 64, 65, 85, 98
22	11, 14, 33, 41, 43, 50, 89
23	54, 60, 71
24	8, 16, 96
25	9, 34, 50, 51, 56
26	46, 65, 87
27	47, 71, 80
28	31, 32, 46, 50, 98

Table S7: Cont'd.

Residue	Snapshot
29	0, 50, 67
30	15, 22, 62, 80, 84, 99
31	16, 34, 38, 39, 56, 68, 75, 77, 78, 84, 96
32	8, 30, 33, 38, 83, 84
33	12, 15, 19
34	24, 47, 78
35	32, 48, 76
36	24, 88, 95
37	25, 28, 51, 52, 99
38	5, 24, 88
39	27, 28, 84
40	17, 20, 23, 33, 38, 41, 50, 59, 61, 76, 83, 91, 98
41	22, 37, 91
42	20, 21, 64, 69, 89
43	7, 11, 80
44	11, 21, 26, 62, 80, 93, 95
45	5, 55, 65, 93, 98
46	20, 65, 69, 87
47	46, 62, 73, 86, 93
48	28, 52, 91
49	29, 48, 81
50	11, 24, 37
51	12, 15, 37
52	18, 73, 76, 78
53	7, 10, 17
54	4, 31, 64
55	50, 64, 97, 98

4 ProCS15B Averaging for $\bar{K}_j = 10$

In Table S8 we show which snapshots from our molecular dynamics simulation are to be used for each residue to minimize the variance when computing NMR shielding constants. The predictions are made using ProCS15. Residues that have a large computed variance, such as illustrated in Figure 7 in the manuscript, have more snapshots included to lower the variance in the computed shieldings.

Table S8: List of snapshots which minimizes the variance for each residue. Residue 1 and 56 excluded. Residue counting starts from zero.

Residue	Snapshot
2	0, 13, 19, 21, 26, 50, 54, 61, 82, 92, 94
3	34, 51, 54, 60, 67, 79
4	7, 19, 23, 26, 58, 84, 89
5	14, 19, 83
6	63, 67, 96
7	43, 80, 92, 95
8	4, 6, 12, 16, 37, 43, 59, 65, 79, 96, 99
9	21, 37, 39, 43, 57, 58, 68, 88, 96, 99
10	4, 5, 10, 12, 13, 14, 16, 17, 18, 22, 25, 26, 27, 36, 37, 38, 43, 47, 49, 50, 52, 55, 57, 58, 60, 67, 72, 75, 76, 79, 81, 82, 84, 91, 92, 96, 99
11	0, 5, 9, 10, 16, 18, 21, 33, 36, 39, 46, 48, 51, 54, 56, 57, 59, 60, 66, 70, 72, 75, 79, 91, 93, 96
12	3, 32, 38, 51, 60, 62, 67, 68, 69, 80, 83, 91, 93
13	4, 5, 17, 20, 21, 25, 28, 37, 51, 52, 58, 61, 77, 80, 85, 88, 96
14	1, 26, 30, 37, 50, 64, 66, 72, 75, 87, 92
15	3, 7, 22, 39, 48, 55, 86, 89, 93
16	13, 19, 20, 29, 38, 54, 77, 81, 83, 86, 98
17	20, 25, 29, 48, 50, 58, 64, 84, 85, 89
18	5, 10, 16, 19, 25, 26, 33, 39, 52, 54, 55, 56, 57, 58, 62, 71, 78, 87, 93, 95
19	0, 22, 23, 26, 34, 41, 43, 55, 58, 71, 76, 94, 99
20	61, 94, 98
21	7, 9, 10, 11, 17, 21, 22, 24, 33, 34, 36, 46, 48, 56, 60, 61, 62, 63, 69, 78, 80, 82, 94, 95
22	6, 7, 18, 35, 36, 55, 59, 62, 63, 67, 68, 85, 89, 91, 97
23	21, 28, 59, 65, 77, 85
24	0, 5, 17, 19, 54, 83
25	1, 8, 18, 23, 27, 50, 65, 72, 92
26	46, 65, 87
27	47, 71, 80
28	7, 9, 19, 22, 24, 52, 62, 63, 86, 88, 98

Table S8: Cont'd.

Residue	Snapshot
29	0, 50, 67
30	0, 1, 16, 19, 24, 61, 63, 68, 73, 75, 80, 89, 96, 99
31	8, 14, 16, 17, 20, 24, 28, 31, 34, 35, 47, 54, 56, 61, 63, 64, 66, 68, 79, 82, 83, 87, 89, 91, 95, 98
32	5, 7, 8, 16, 31, 34, 72, 79, 81, 83, 91, 93, 95, 98
33	9, 11, 12, 15, 19, 37, 48, 79, 84
34	24, 47, 78
35	6, 16, 61, 73
36	1, 16, 55, 58, 73, 89
37	4, 6, 17, 23, 32, 47, 51, 62, 84
38	1, 5, 10, 24, 27, 48, 75, 80, 88
39	4, 14, 65, 84, 85
40	1, 3, 8, 9, 11, 15, 16, 20, 24, 25, 35, 36, 52, 60, 61, 62, 63, 66, 70, 72, 79, 80, 82, 84, 88, 89, 91, 95, 99
41	18, 22, 29, 62, 69, 80, 99
42	23, 25, 41, 50, 63, 66, 68, 77, 81, 86, 94, 99
43	11, 20, 33, 66, 77
44	1, 3, 8, 27, 29, 32, 37, 51, 59, 68, 73, 77, 87, 91
45	13, 22, 24, 39, 58, 61, 63, 71, 75, 94
46	7, 35, 37, 62, 64, 82
47	0, 4, 5, 10, 21, 24, 25, 36, 51, 67, 72, 98
48	28, 52, 91
49	15, 37, 43, 86
50	11, 24, 37
51	15, 18, 26, 36, 85
52	6, 25, 26, 28, 38, 48, 87, 92
53	17, 25, 50, 81, 92
54	4, 31, 64
55	6, 37, 49, 62, 66, 67, 72