Supporting Information

for

Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium.

Katarina Hart[1], Nicolas Foloppe[2], Christopher M. Baker,[3] Elizabeth J. Denning,[3] Lennart Nilsson[1*] and Alexander D. MacKerell Jr.[3*]

[1]Department of Biosciences and Nutrition, Center for Biosciences, Karolinska Institutet, SE-141 83 HUDDINGE, Sweden

[2]51 Natal Road, Cambridge CB1 3NY, UK

[3]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, Maryland 21201, USA

*Corresponding authors: Lennart.Nilsson@ki.se, alex@outerbanks.umaryland.edu

Running title: CHARMM DNA force field optimization

Keywords: nucleic acids, RNA, empirical force field, quantum mechanics, EcorRI dodecamer, oligonucleotide, A-DNA, B-DNA, Z-DNA, crystal survey

Table S1) Modified dihedral parameters .

ε:C2'-C3'-O3'-P[1]	K_φ	n	δ
C27			
CN8-CN7-ON2-P	2.5	1	180
C27_1			
CN8-CN7-ON2-P	1.7	1	180
C27_2			
CN8-CN7-ON2-P	2.1	1	180
C27_2 with C27_2b C2'-C3'-C4'-O4' parameters			
CN8-CN7-ON2-P2	1.9	1	180
C27_3			
CN8-CN7-ON2-P	2.2	1	180
CN8-CN7-ON2-P	0.7	2	180
ζ:C3'-O3'-P-O5' [1]			
C27			
ON2-P2-ON2-CN7	1.2	1	180
ON2-P2-ON2-CN7	0.1	2	180
ON2-P2-ON2-CN7	0.1	3	180
C27_1			
Same_as_C27			
C27_2			
ON2-P2-ON2-CN7	0.9	1	180
ON2-P2-ON2-CN7	0.4	2	180
ON2-P2-ON2-CN7	0.2	3	180
C27_3			
Same_as_C27			
Sugar pucker: C2'-C3'-C4'-O4' [1]			
C27			
CN8-CN7-CN7-ON6	0.0	3	0
C27_2b and C27_3b			
CN8-CN7-CN7-ON6	1.0	4	0
CN8-CN7-CN7-ON6	0.3	5	180
CN8-CN7-CN7-ON6	0.3	6	180

1) Atom naming in the header lines is the standard nomenclature for DNA[1] while the names associated with the parameters are the CHARMM atom type names. Note that in the final model (CHARMM36) the P atom type name was changed to P2 to avoid a conflict in the ζ parameters with those in RNA.

2) K_φ, n, and δ are the force constant (kcal/mol), multiplicity or periodicity and phase angle associated with the dihedral term in the potential energy function.

Table S2a) Average percent BII as a function of residue for JunFos from ^{31}P NMR experiments and the simulations.

Base Step	Exper	C27		C27_2b		Amber bsc0	
		Ave	Diff	Ave	Diff	Ave	Diff
5'-G_1pC_2	22	20±1.3	-2	50±6.2	28	30±6.6	8
C_2pA_3	69	6±0.2	-63	30±5.4	-39	10±2.3	-59
A_3pT_4	0	0±0.1	0	2±0.7	2	3±1.4	3
T_4pT_5	22	1±0.3	-21	2±0.3	-20	1±1.1	-21
T_5pC_6	13	2±0.2	-11	8±0.5	-5	2±0.6	-11
C_6pT_7	13	1±0.1	-12	6±0.3	-7	0±0.2	-13
T_7pG_8	37	8±1.3	-29	21±2	-16	2±0.7	-35
G_8pA_9	49	12±1.7	-37	37±1.8	-12	46±3.1	-3
A_9pG_{10}	26	1±0.2	-25	7±0.8	-19	7±4.1	-19
$G_{10}pT_{11}$	0	1±0.3	1	9v1.7	9	2±1.0	2
$T_{11}pC_{12}$	9	6±2.6	-3	21±2.8	12	1±0.1	-8
$C_{12}pA_{13}$	80	7±2.5	-73	31±9.5	-49	2±0.7	-78
$A_{13}pG_{14}$-3'	42	20±8.1	-22	56±12.7	14	5±4.1	-37
5'-$C_{15}pT_{16}$	23	29±4.9	6	62±3.2	39	1±0.5	-22
$T_{16}pG_{17}$	52	23±5.7	-29	44±6.7	-8	7±2.5	-45
$G_{17}pA_{18}$	50	28±9.3	-22	61±5.9	11	63±4.4	13
$A_{18}pC_{19}$	29	1±0.3	-28	13±3.2	-16	1±0.2	-28
$C_{19}pT_{20}$	0	4±0.8	4	11±1.0	11	1±0.5	1
$T_{20}pC_{21}$	2	2±0.2	0	10±1.0	8	1±0.2	-1
$C_{21}pA_{22}$	85	4±0.5	-81	20±3.3	-65	2±0.3	-83
$A_{22}pG_{23}$	28	8±0.5	-20	21±1.7	-7	33±6.0	5
$G_{23}pA_{24}$	32	16±1.0	-16	37±1.2	5	17±3.0	-15
$A_{24}pA_{25}$	13	1±0.1	-12	6±0.5	-7	5±1.1	-8
$A_{25}pT_{26}$	0	1±0.1	1	6±0.5	6	0±1.1	0
$T_{26}pG_{27}$	26	4±0.4	-22	9±2.0	-17	2±0.7	-24
$G_{27}pC_{28}$-3'	66	4±0.5	-62	35±13.7	-31	10±2.0	-56
Average	30	8±1.7	-22	24±3.4	-7	10±1.9	-20
Correlation		0.30		0.49		0.26	

Results obtained over 100 ns simulations. Statisical analysis for the individual base steps accounting for the symmetry of the sequence based on 5 20 ns blocks from which the averages and standard deviations were calculated with the errors for the average differences are the standard error over all the base seteps. Correlations are between the experimental and average simulation values over the base steps. Experimental data from Heddi et al.[2]

Table 2b) Average percent BII as a function of residue for NF-κb from ^{31}P NMR experiments and the simulation.

Base Step	Exper	C27		C27_2b	
		Ave	Diff	Ave	Diff
5'-G1pA2	32	47±2	15	88±2	56
A2pC3	58	2±0	-56	15±2	-43
C3pT4	0	4±2	4	20±3	20
T4pT5	12	0±0	-12	1±0	-11
T5pT6	0	1±0	1	2±0	2
T6pC7	16	4±0	-12	10±1	-6
C7pC8	32	2±0	-30	17±1	-15
C8pA9	75	5±1	-70	28±3	-47
A9pG10	28	4±1	-24	32±2	4
G10pG11-3'	59	18±7	-41	64±11	5
5'-C14pC15	34	20±1	-14	67±3	33
C15pT16	11	2±7	-9	29±7	18
T16pG17	42	3±12	-39	35±11	-7
G17pG18	42	6±1	-36	37±6	-5
G18pA19	36	8±1	-28	32±7	-4
A19pA20	2	3±0	1	10±4	8
A20pA21	12	3±1	-9	9±2	-3
A21pG22	38	2±0	-36	8±1	-30
G22pT23	0	2±1	2	8±1	8
T23pC24-3'	59	5±0	-54	38±1	-21
Average	29	7±2	-22	28±3	-2
Correl		0.20		0.45	

Results obtained over 100 ns simulations. Statistical analysis for the individual base steps accounting for the symmetry of the sequence based on 5 20 ns blocks from which the averages and standard deviations were calculated with the errors for the average differences are the standard error over all the base steps. Correlations are between the experimental and average simulation values over the base steps. Experimental data obtained using the method of Heddi et al.[2] (Brigette Hartmann, Personal Communication).

Table S3) Order parameters, S^2, for EcoRI from NMR experiments and excess MD simulation using C27_2b. Results from the simulations are presented individually for strand 1 (s1) and strand 2 (s2).

C1' atom	Exper		C27_2b/high salt	
Base	S^2	SD	s1	s2
1	0.52	0.02	0.45	0.57
2	0.78	0.03	0.75	0.78
3	0.74	0.03	0.56	0.59
4	0.88	0.03	0.81	0.80
5	0.84	0.03	0.80	0.88
6	N.A.	N.A.	0.56	0.60
7	0.92	0.02	0.67	0.66
8	0.86	0.02	0.81	0.76
9	0.68	0.03	0.60	0.55
10	0.85	0.02	0.82	0.81
11	0.71	0.02	0.63	0.65
12	N.A.	N.A.	0.70	0.68
Difference Analysis				
Average_all			-0.09	-0.07
Correlation			0.84	0.70
Average_non_terminal			-0.09	-0.08
Correlation			0.70	0.64
C3' atom	Exper		C27_2b/high salt	
Base	S^2	SD	s1	s2
1	0.39	0.02	0.51	0.58
2	N.A.	N.A.	0.75	0.80
3	N.A.	N.A.	0.50	0.52
4	N.A.	N.A.	0.81	0.82
5	0.90	0.02	0.76	0.89
6	0.79	0.03	0.35	0.41
7	N.A.	N.A.	0.46	0.47
8	0.79	0.03	0.78	0.65
9	0.67	0.04	0.50	0.40
10	N.A.	N.A.	0.82	0.78
11	N.A.	N.A.	0.58	0.62
12	0.43	0.05	0.64	0.63
Difference Analysis				
Average_all			-0.11	-0.11
Correlation			0.24	0.25
Average_non_terminal			-0.19	-0.20
Correlation			0.47	0.84

Table S3 continued

C6/C8 atoms	Exper		C27_2b/high salt	
Base	S^2	SD	s1	s2
1	0.77	0.04	0.57	0.72
2	0.81	0.07	0.86	0.86
3	0.92	0.04	0.84	0.85
4	N.A.	N.A.	0.88	0.89
5	N.A.	N.A.	0.90	0.90
6	N.A.	N.A.	0.90	0.90
7	0.83	0.02	0.86	0.86
8	0.87	0.04	0.85	0.85
9	0.79	0.06	0.85	0.85
10	0.88	0.04	0.88	0.88
11	0.88	0.04	0.82	0.81
12	0.91	0.08	0.80	0.78
Difference Analysis				
Average_all			-0.04	-0.02
Correlation			0.47	0.27
Average_non_terminal			-0.01	-0.02
Correlation			-0.15	-0.08

Analysis over the 5 to 100 ns portions of the trajectories. Difference and correlation coefficient calculated over all nucleotides for which experimental data is available and excluding the data for the terminal nucleotides.

Figure S1) RMS difference versus time for 1ZF1 for C27 (black) and C27_2b (red) from MD simulations in a 75 % ethanol solution (A, B) or aqueous solution (C, D) versus the canonical A (A, C) and B (B, D) forms of DNA.

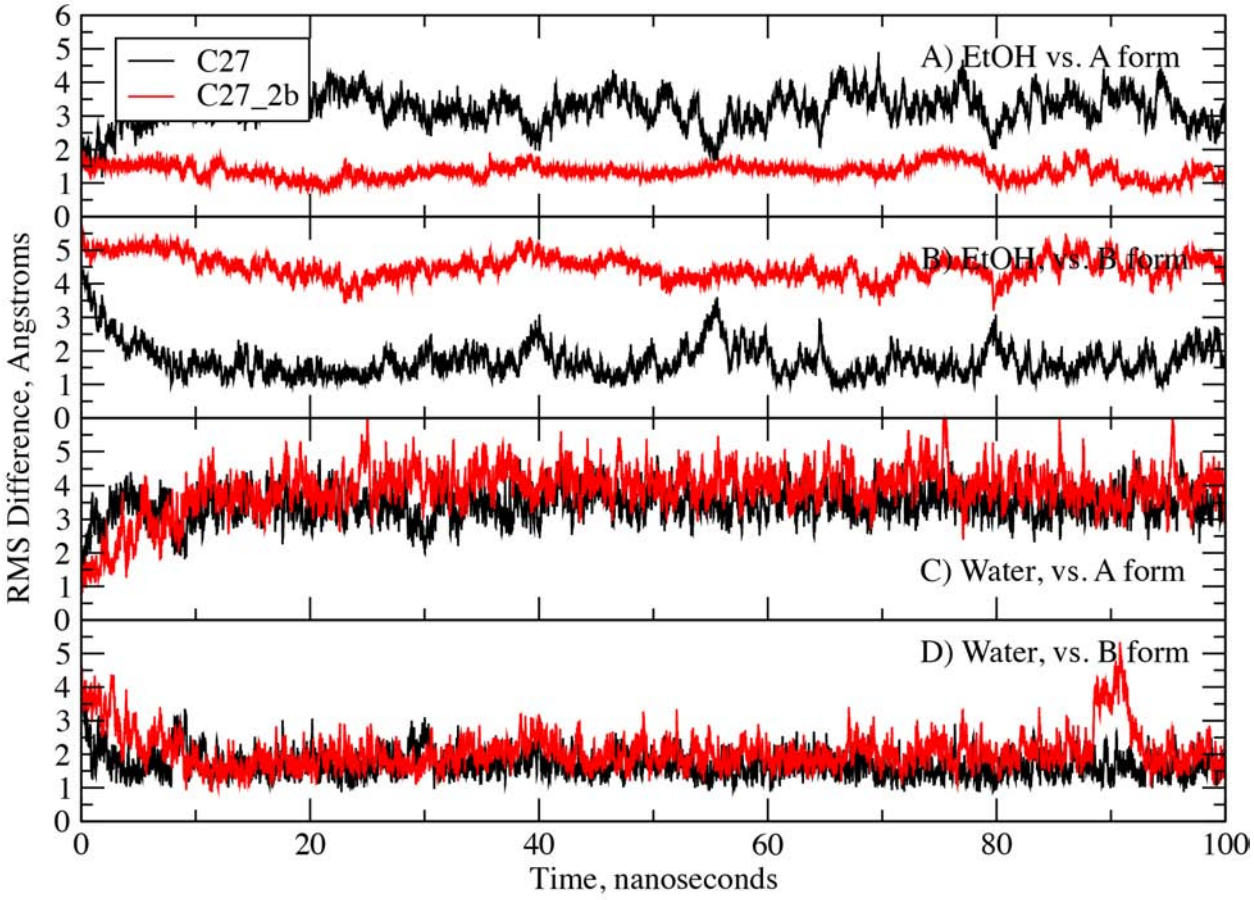
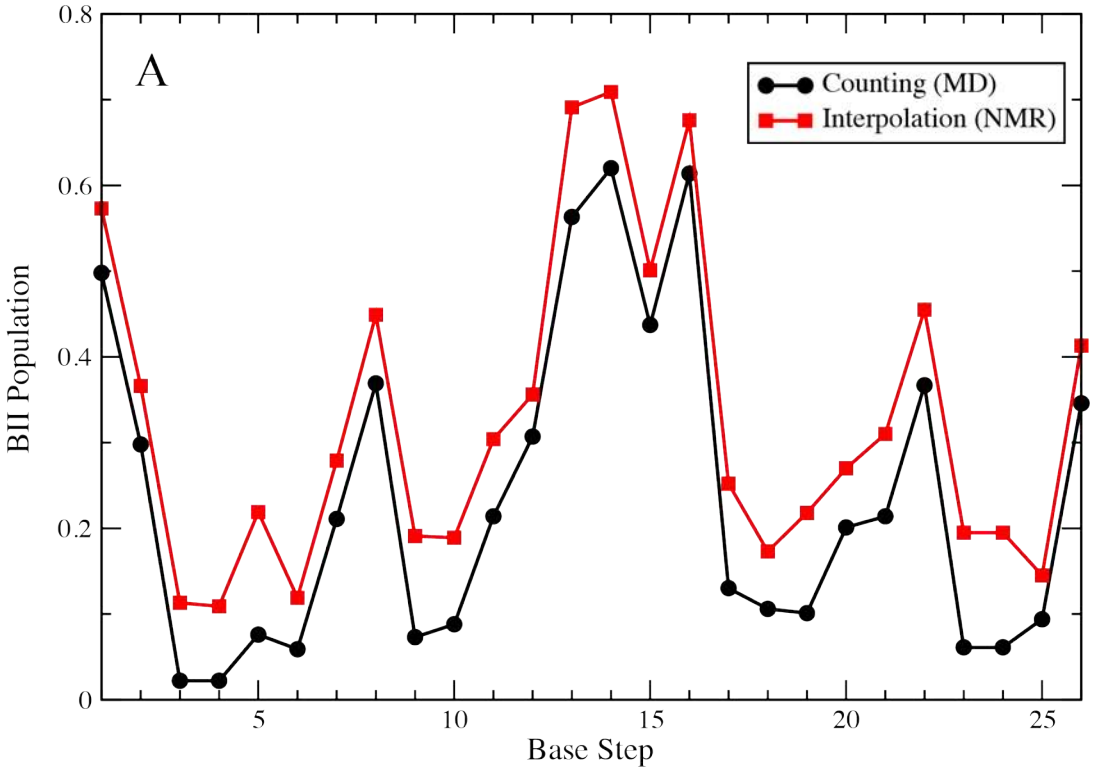


Figure S2) A) BII population from the JunFos C27_2b simulation calculated using simple counting of BI and BII states as performed in the MD analysis versus the interpolation method based on the average difference between ε and ζ, analogous to that performed in the ^{31}P NMR experimental estimates of BII. B) Correlation between the BII populations from the counting and interpolation methods from the C27_2b simulation.



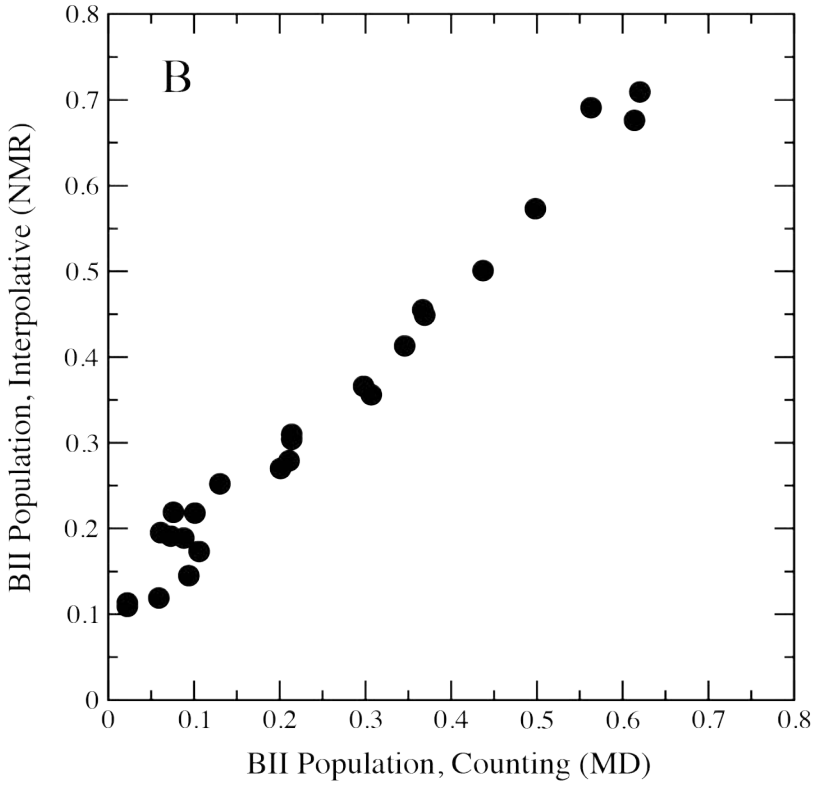


Figure S3) RMS fluctuations as a function of nucleotide from simulations of the 1ZF7 (A, B, C) and 3BSE (D, E, F) duplexes for all non-hydrogen atoms (A, C), the backbone (phosphate and sugar) non-hydrogen atoms (B, E) and the base non-hydrogen atoms (C, F).

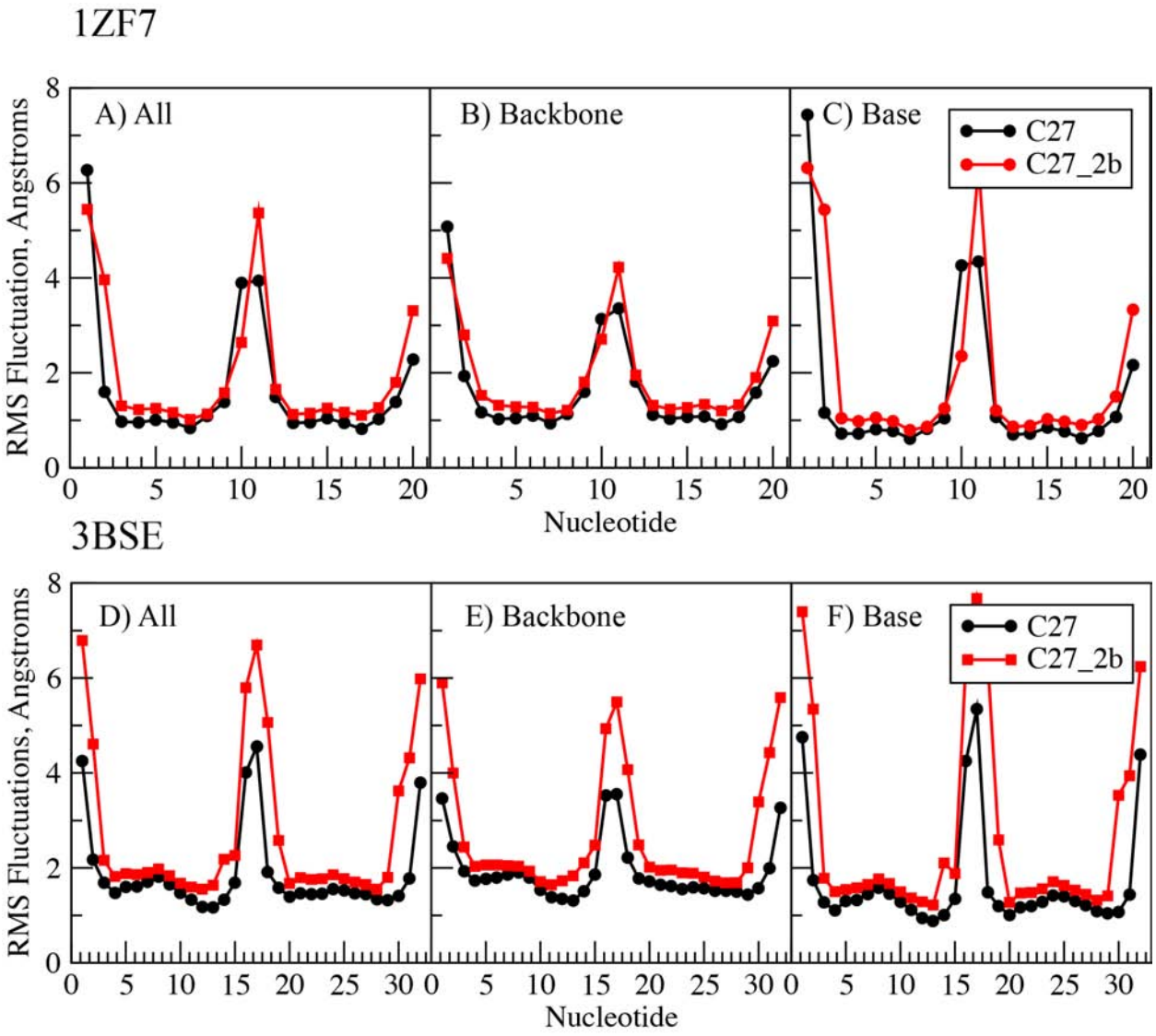
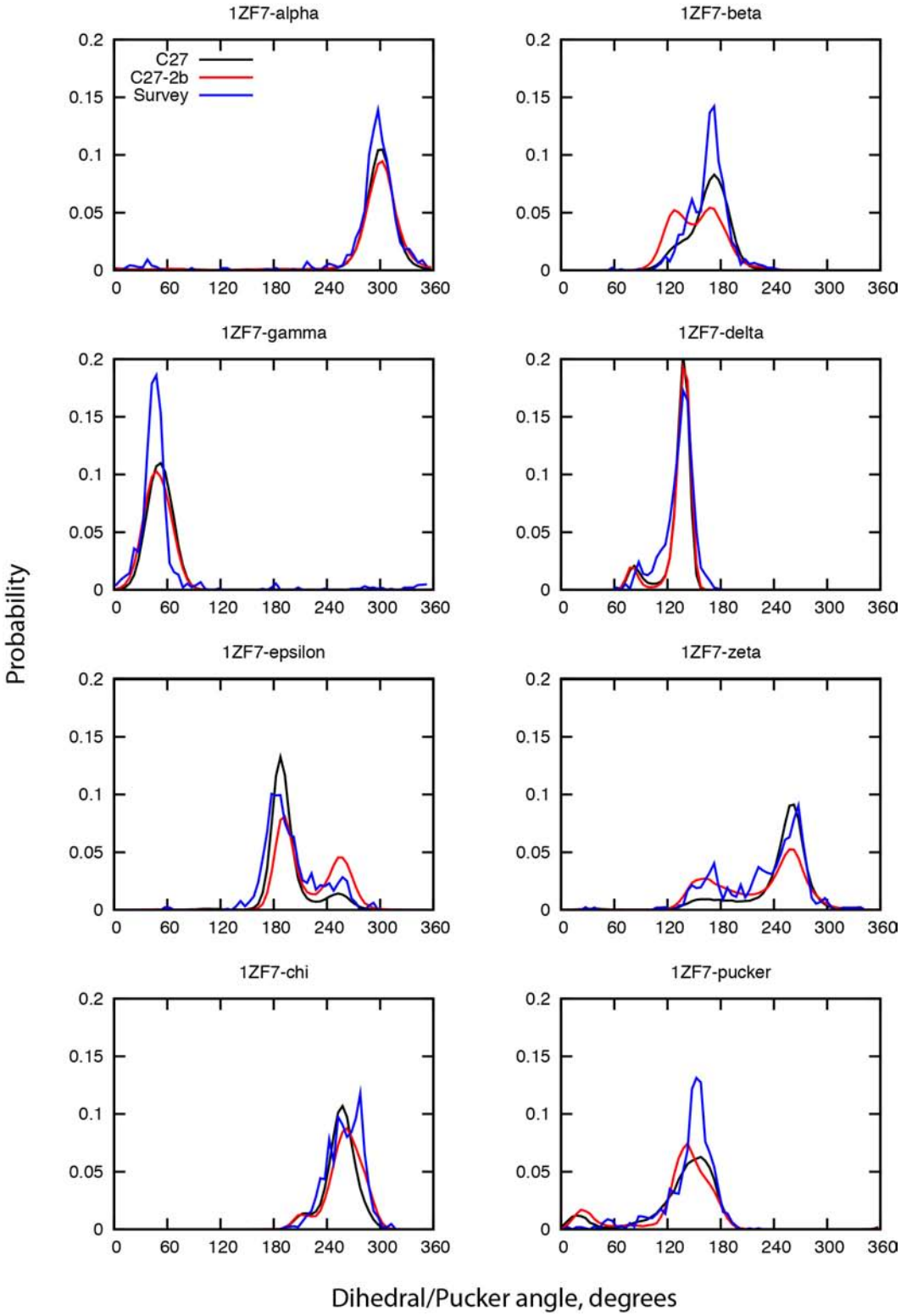
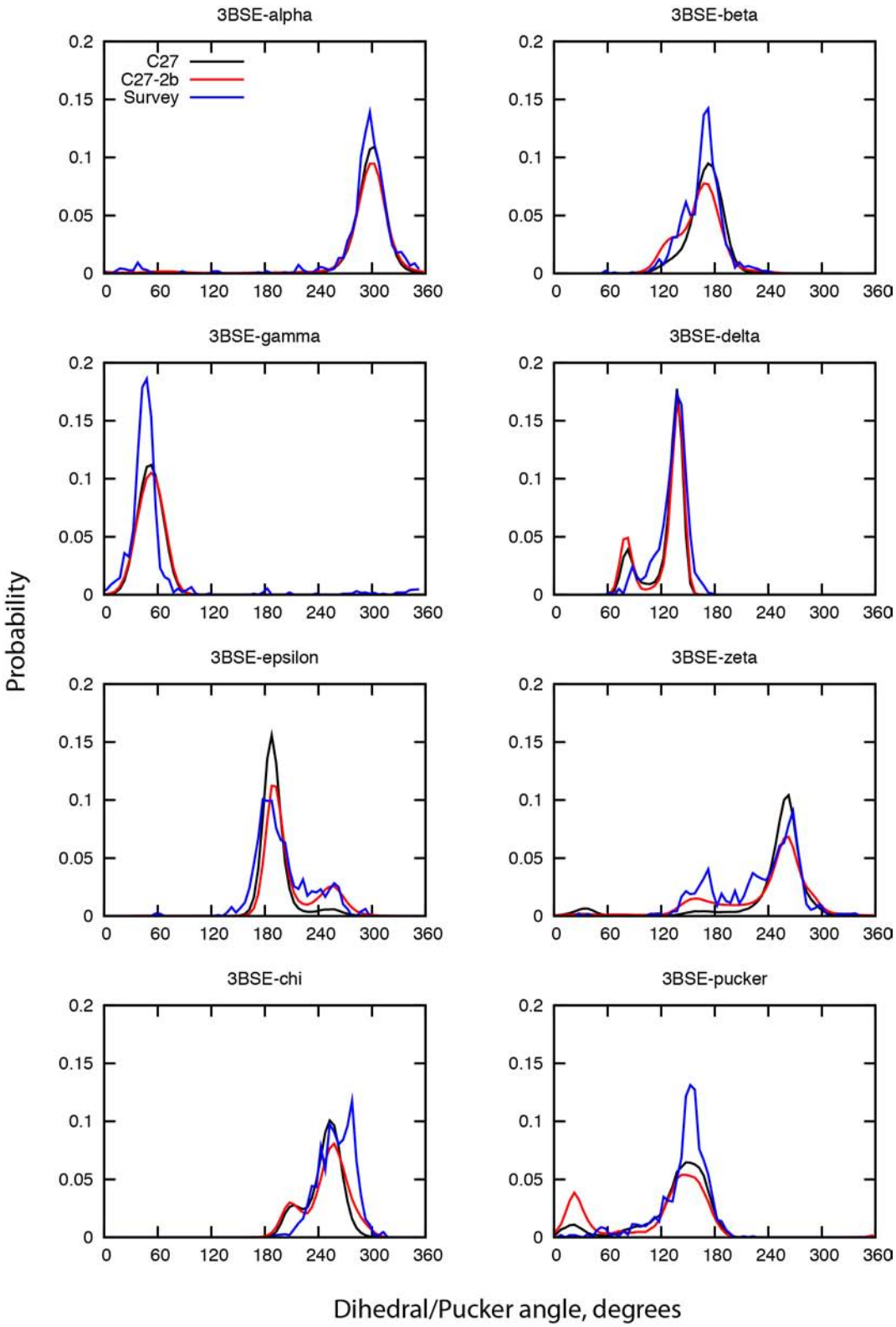


Figure S4) Dihedral and sugar pucker probability distributions from MD simulations using C27 (black) and C27_2b (red) of A) 1ZF7, B) 3BSE, C) 1ZF1 in 75% ethanol and D) 1LJX crystal. Data from a survey of the crystallographic structures (blue) is included for B form structures in A and B and A form structures in C

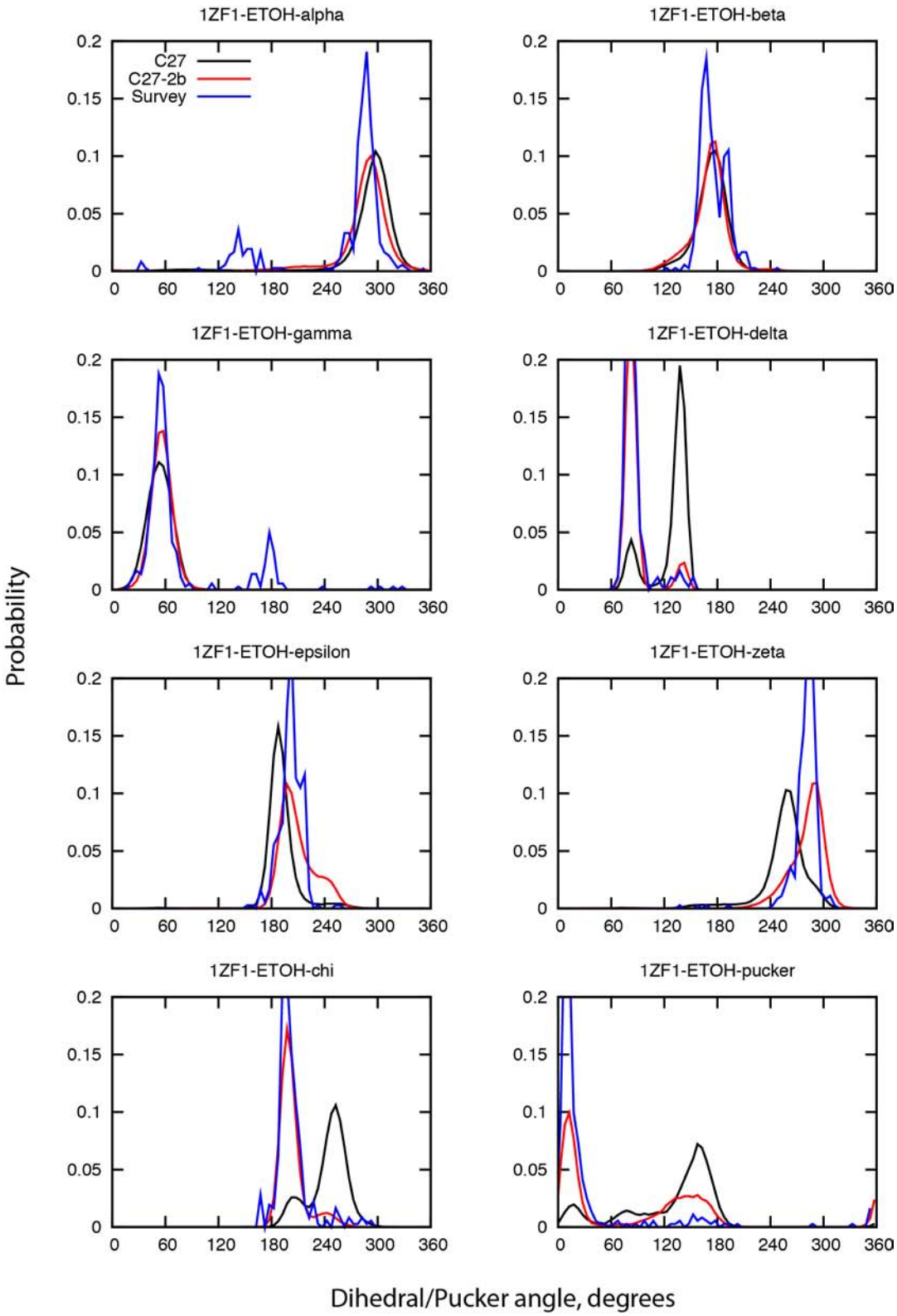
S4a)



S4b)


3BSE-alpha
3BSE-beta
3BSE-gamma
3BSE-delta
3BSE-epsilon
3BSE-zeta
3BSE-chi
3BSE-pucker
C27
C27-2b
Survey
Probability
Dihedral/Pucker angle, degrees
0
0.05
0.1
0.15
0.2
0
60
120
180
240
300
360

S4c)


1ZF1-ETOH-alpha
1ZF1-ETOH-beta
1ZF1-ETOH-gamma
1ZF1-ETOH-delta
1ZF1-ETOH-epsilon
1ZF1-ETOH-zeta
1ZF1-ETOH-chi
1ZF1-ETOH-pucker
C27
C27-2b
Survey
Probability
Dihedral/Pucker angle, degrees

S4d) 1LJX

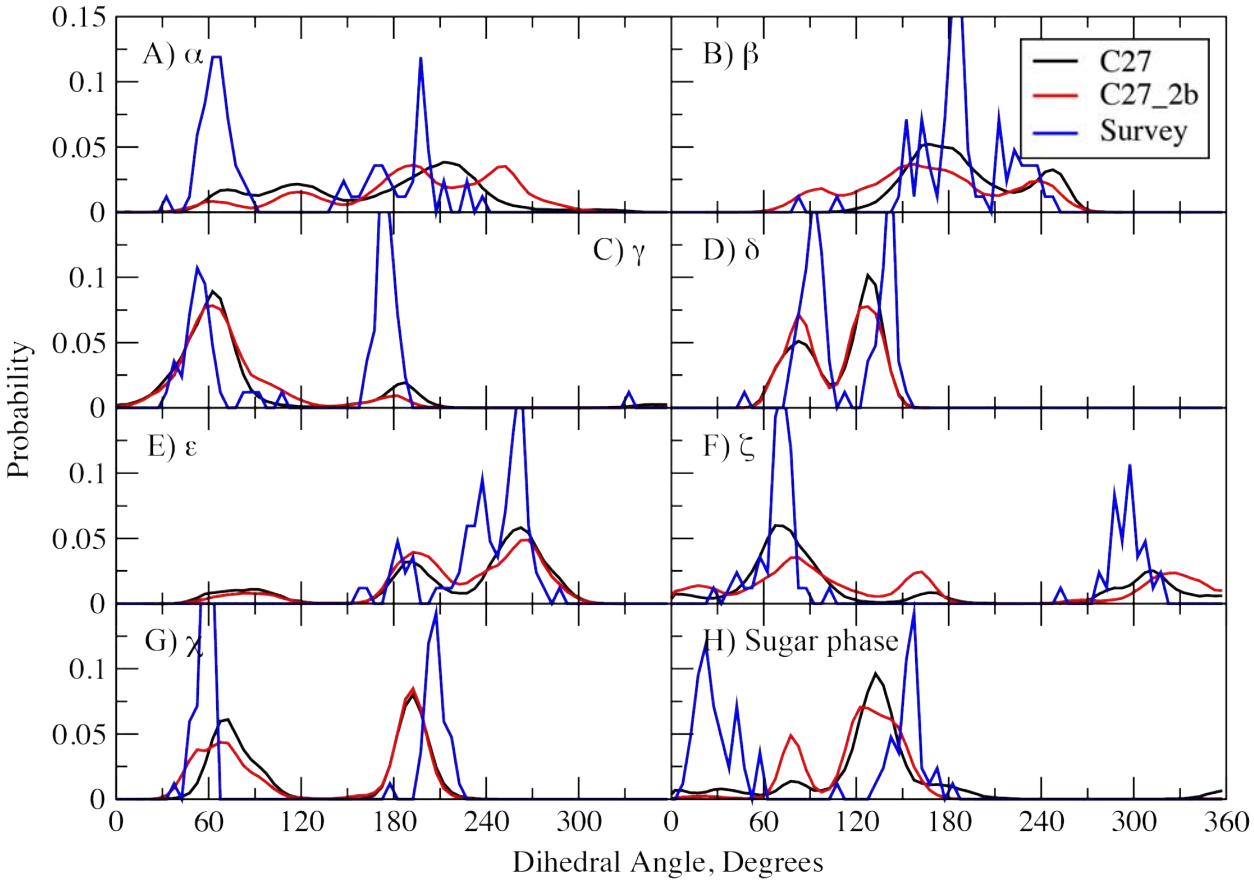
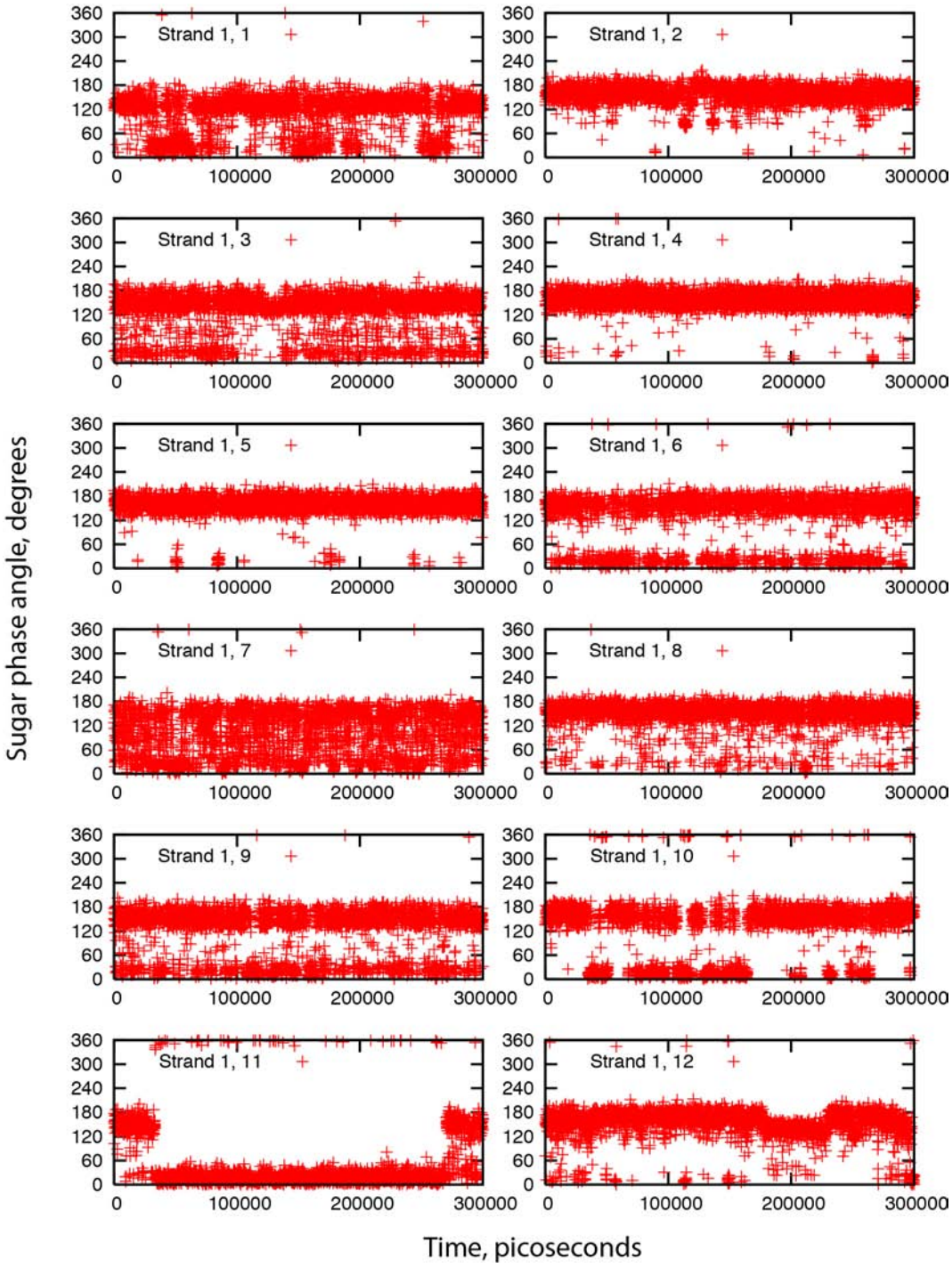

A) α
B) β
C) γ
D) δ
E) ε
F) ζ
G) χ
H) Sugar phase
C27
C27_2b
Survey
Probability
Dihedral Angle, Degrees
0.15
0.1
0.05
0
0
60
120
180
240
300
360

Figure S5) Time series of the sugar pucker phase angles from the C27_2b MD simulation of EcoRI. Values presented at 100 ps intervals for the individual nucleotides. Upper and lower panels represent strands 1 and 2, respectively.



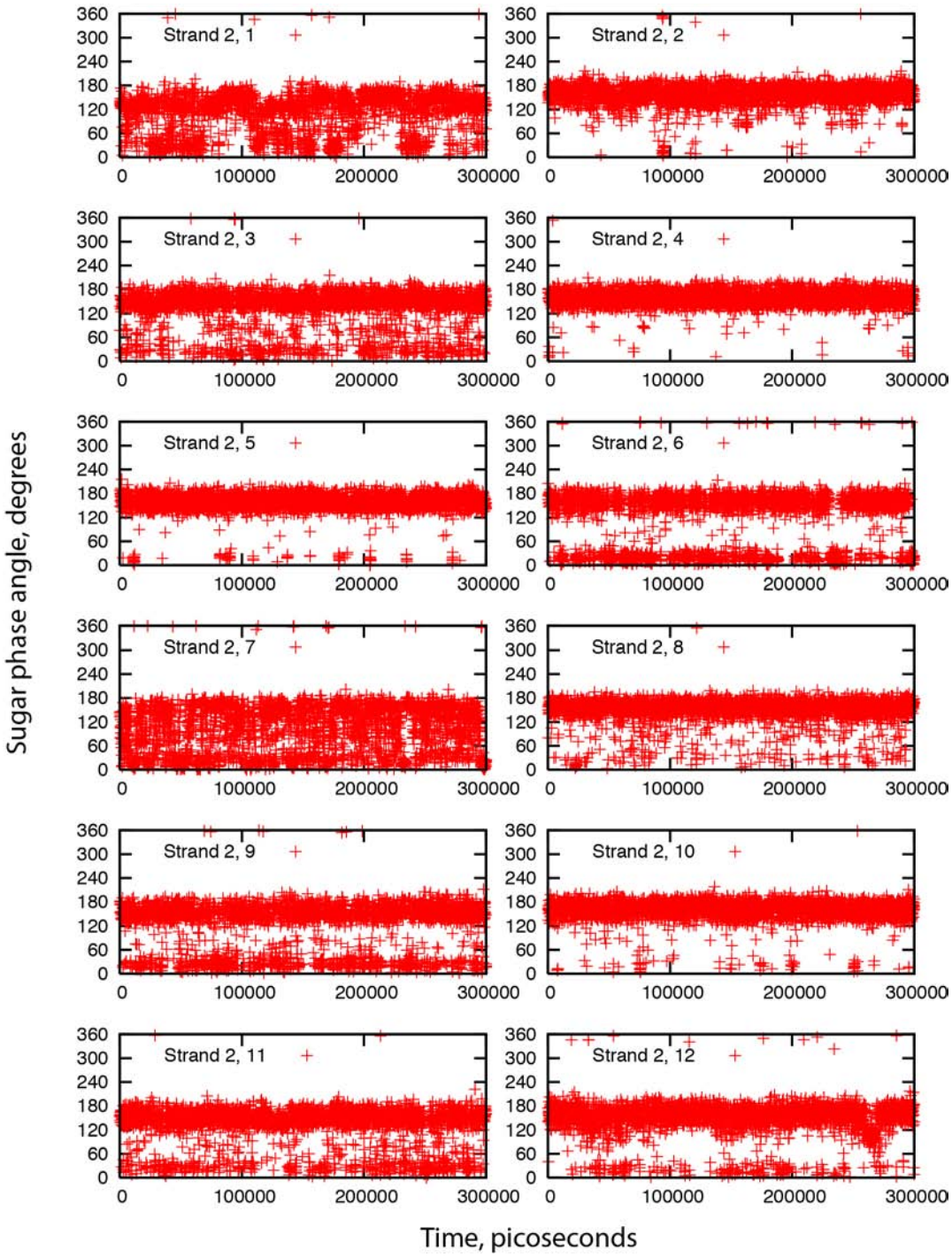
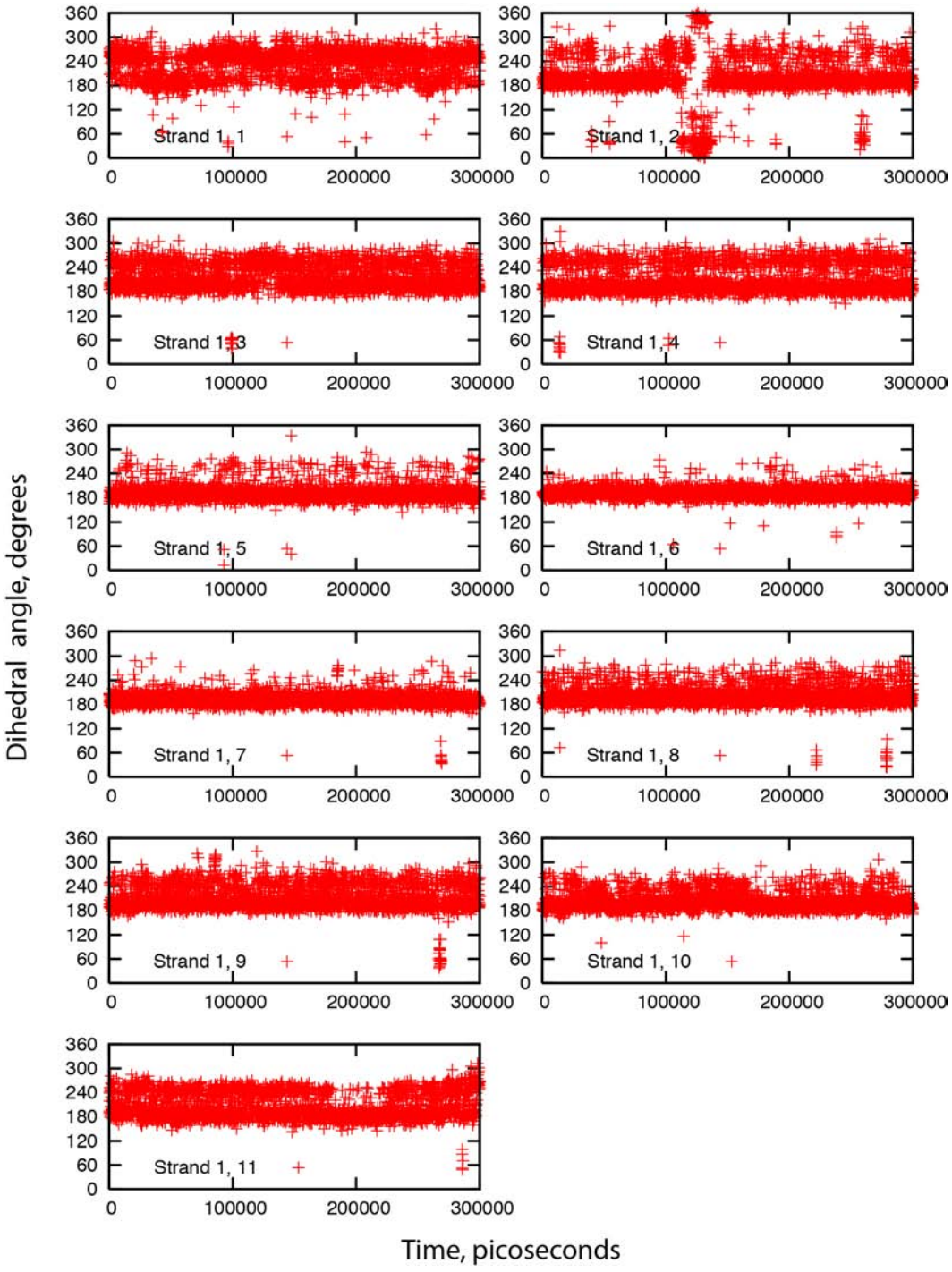

Strand 2, 1
Strand 2, 2
Strand 2, 3
Strand 2, 4
Strand 2, 5
Strand 2, 6
Strand 2, 7
Strand 2, 8
Strand 2, 9
Strand 2, 10
Strand 2, 11
Strand 2, 12
Sugar phase angle, degrees
Time, picoseconds
0
60
120
180
240
300
360
100000
200000
300000

Figure S6) Time series of the epsilon dihedral angle from the C27_2b MD simulation of EcoRI. Values presented at 100 ps intervals. Upper and lower panels represent strands 1 and 2, respectively. Note that in the BI state ε is approximately 190˚ in the BII state it is approximately 270˚.



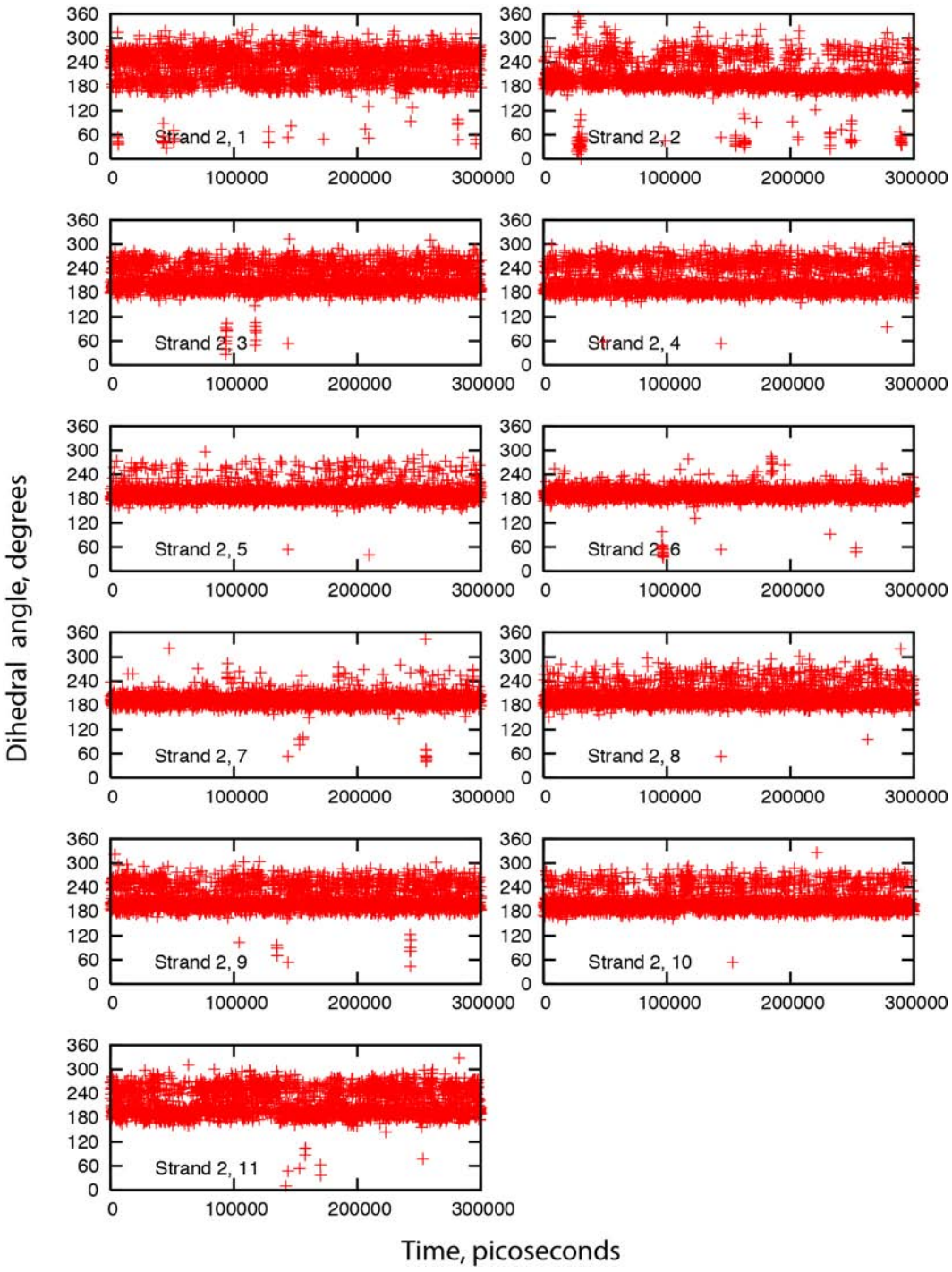
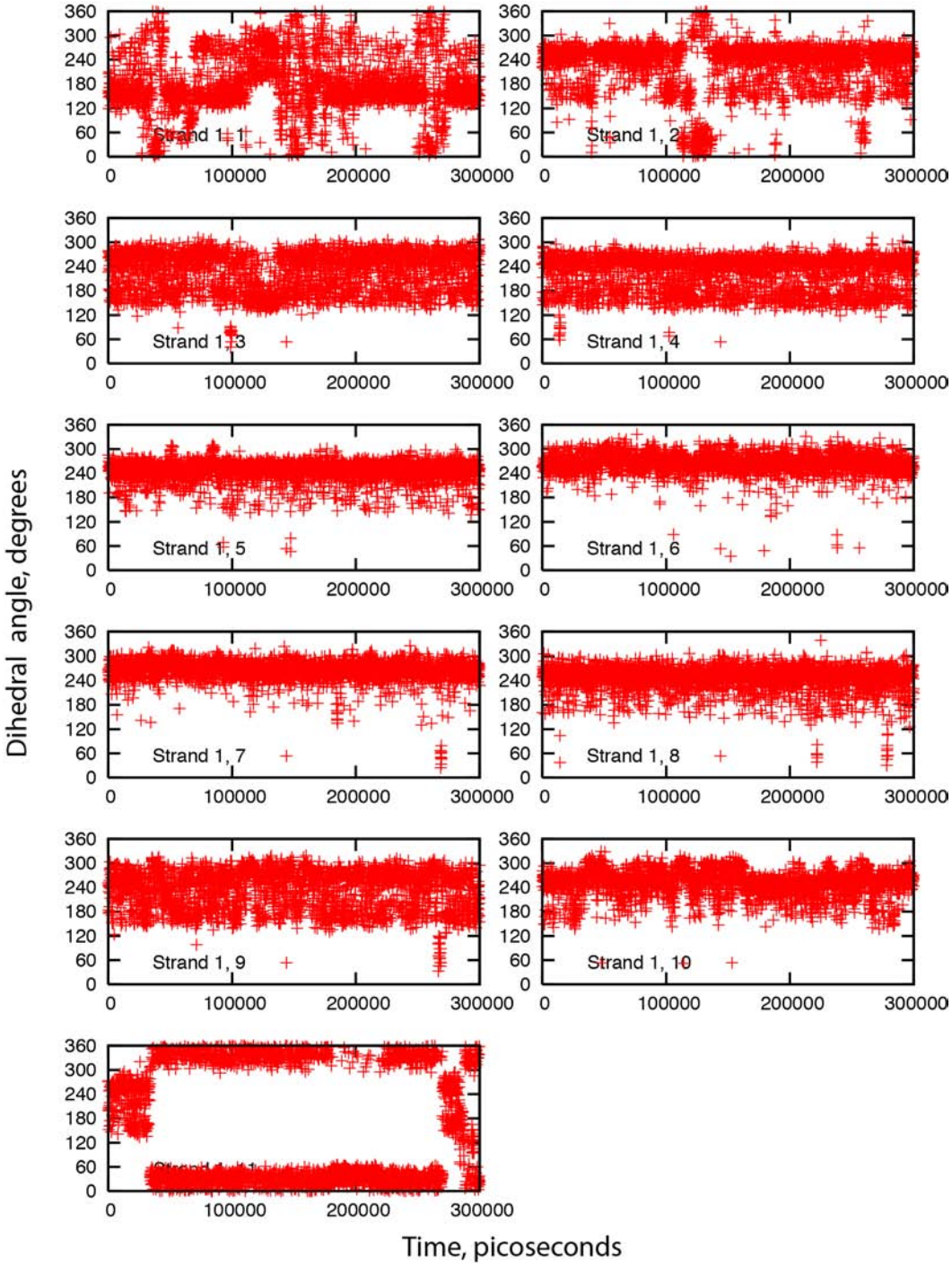

Dihedral angle, degrees
Time, picoseconds
Strand 2, 1
Strand 2, 2
Strand 2, 3
Strand 2, 4
Strand 2, 5
Strand 2, 6
Strand 2, 7
Strand 2, 8
Strand 2, 9
Strand 2, 10
Strand 2, 11
0
60
120
180
240
300
360
0
100000
200000
300000

Figure S7) Time series of the zeta dihedral angle from the C27_2b MD simulation of EcoRI. Values presented at 100 ps intervals. Upper and lower panels represent strands 1 and 2, respectively. Note that in the BI state ζ is approximately 270˚and in the BII state it is approximately 180˚.



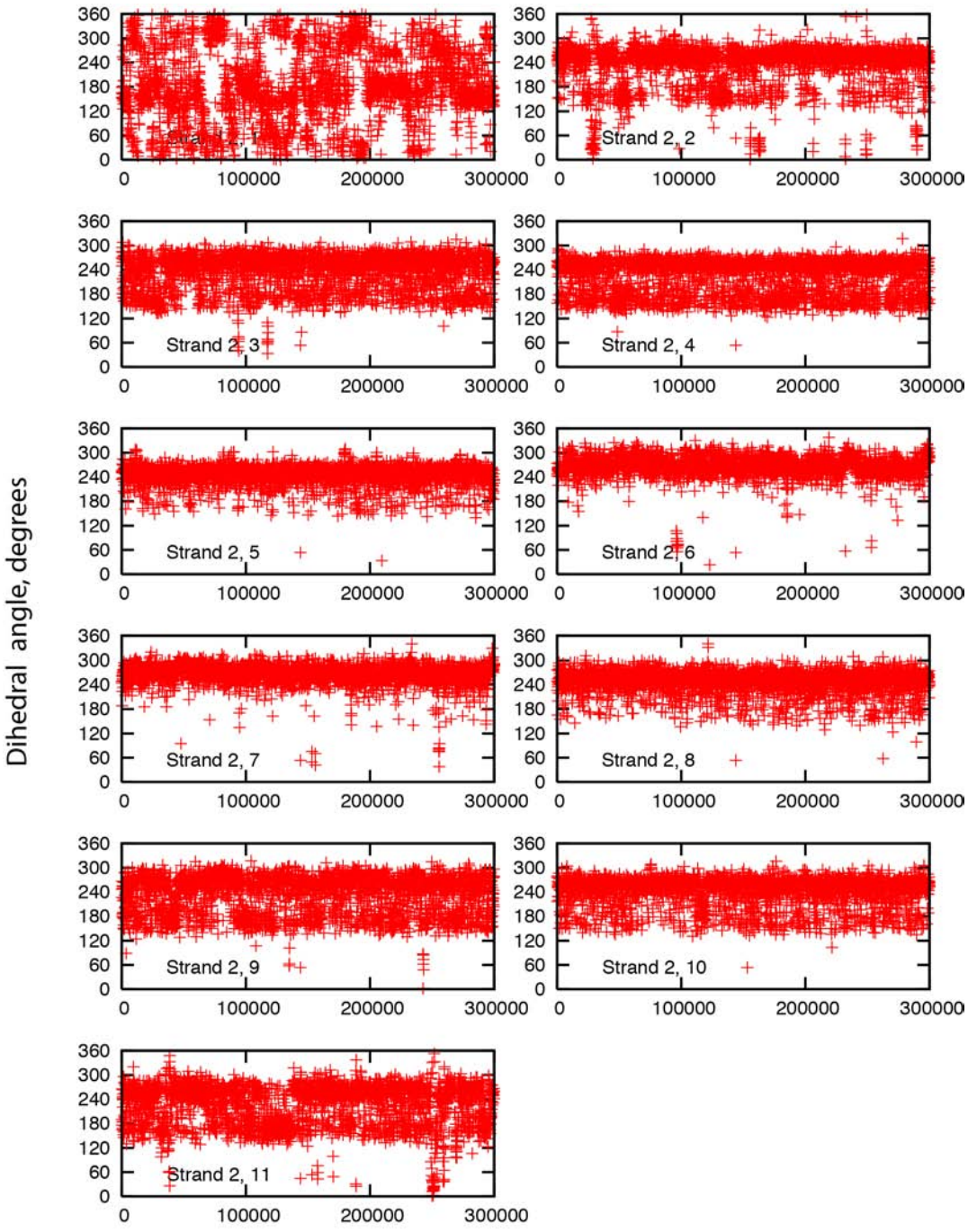
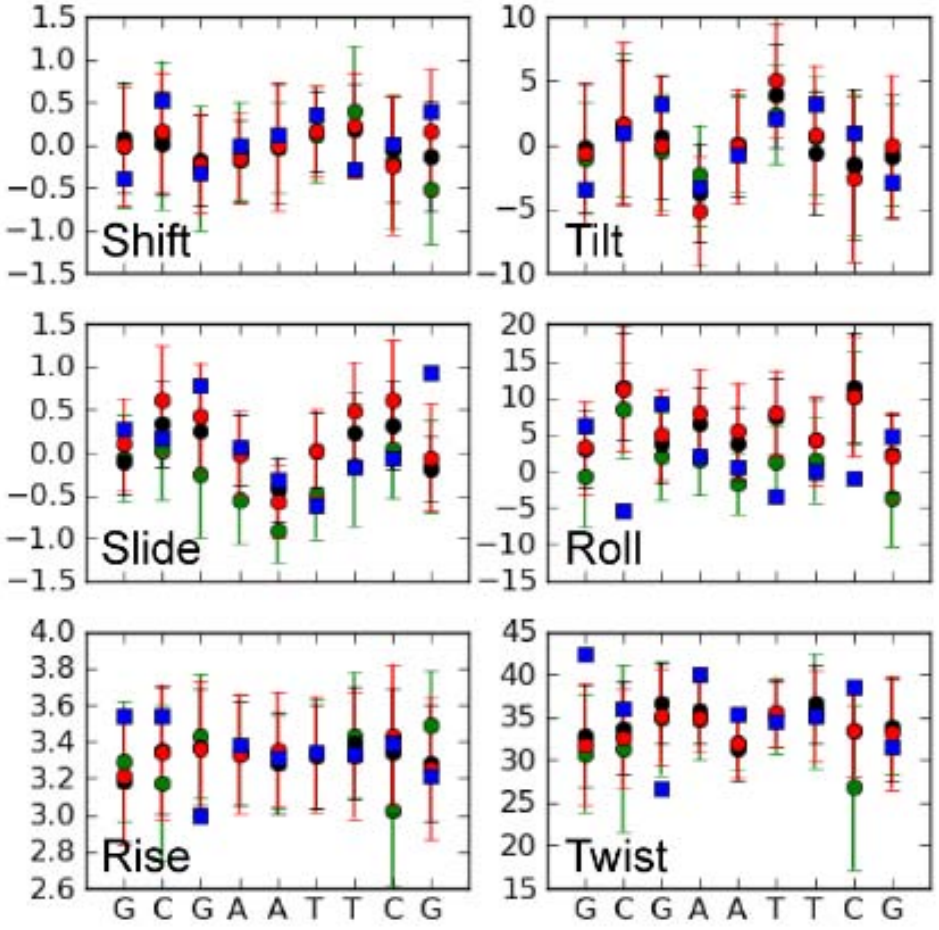

Dihedral angle, degrees
Time, picoseconds
Strand 2, 2
Strand 2, 3
Strand 2, 4
Strand 2, 5
Strand 2, 6
Strand 2, 7
Strand 2, 8
Strand 2, 9
Strand 2, 10
Strand 2, 11
0
60
120
180
240
300
360
0
100000
200000
300000

Figure S8) Selected helicoidal parameters as a function of nucleotide from the 100 ns EcoRI simulations with the C27 (black), C27_2b (red) and AMBER Parm99bsc0 (green) force fields along with values from the EcoRI crystal structure 1BNA (blue). Error bars represent the RMS fluctuations. The X labels indicate the top base pair (bp) for the bp to bp+1 step. The terminal nucleotides are omitted from the analysis.



References

1. Saenger, W., *Principles of Nucleic Acid Structure*. 1984, New York: Springer-Verlag.
2. Heddi, B., N. Foloppe, N. Bouchemal, E. Hantz, and B. Hartmann, *Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition.* J Am Chem Soc, 2006. **128**: p. 9170-7.