Supplementary Information: Alchemical free energy calculations for nucleotide mutations in protein-DNA complexes

Vytautas Gapsys* and Bert L. de Groot*

Computational Biomolecular Dynamics Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

E-mail: vgapsys@gwdg.de; bgroot@gwdg.de

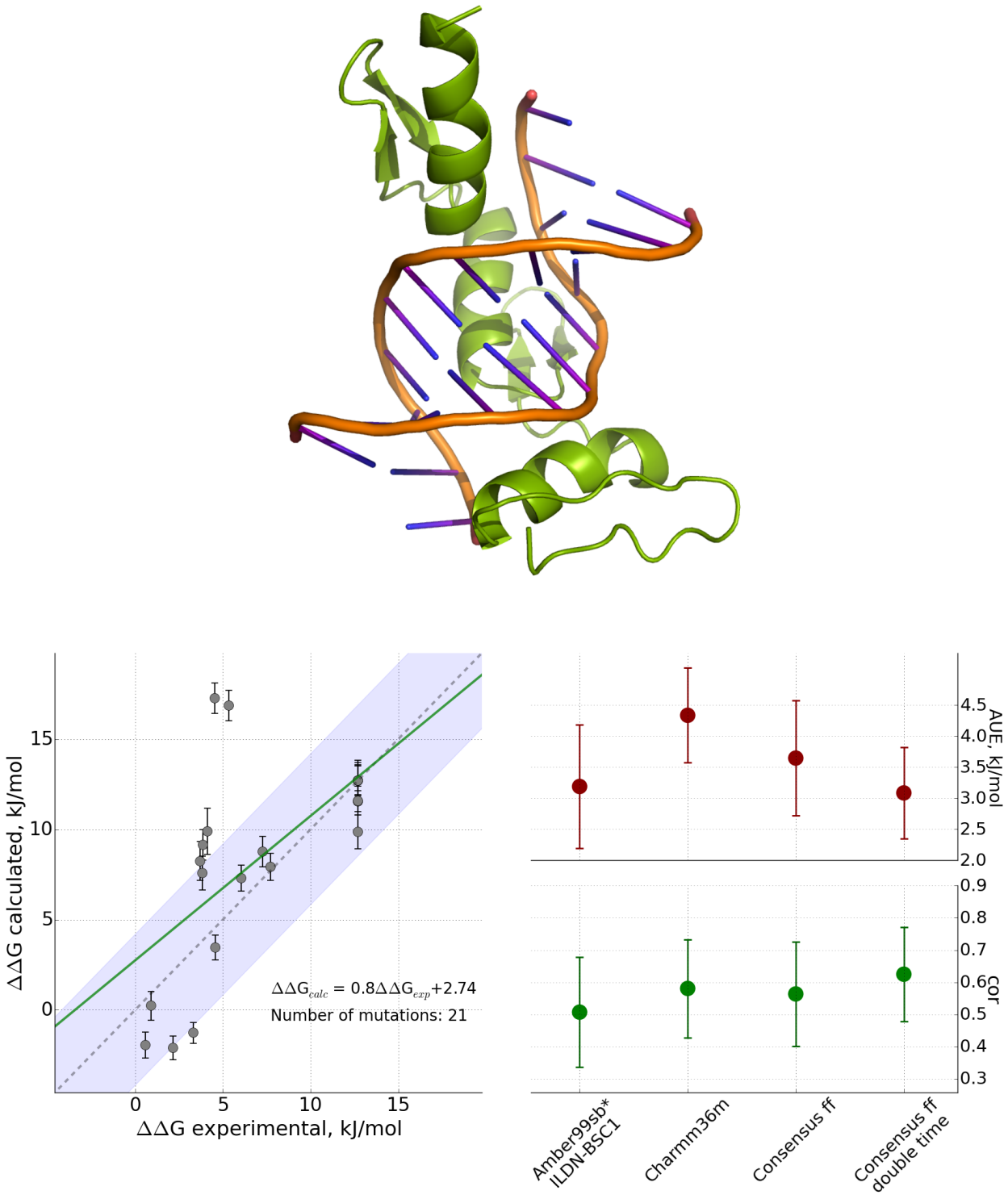

Figure S1: The structure and the calculation results for the DNA bound to the zinc finger protein Zif268. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

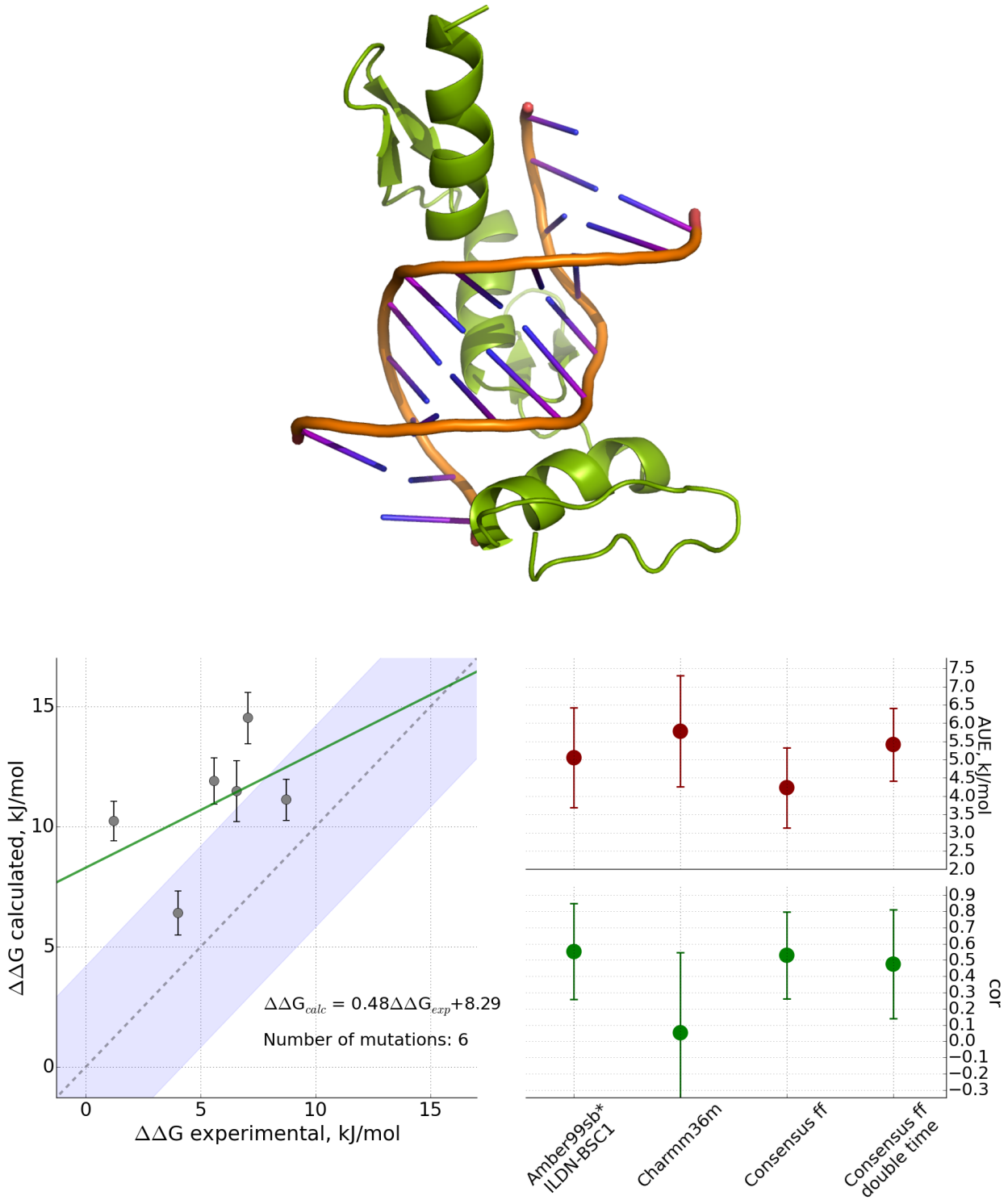

Figure S2: The structure and the calculation results for the DNA bound to the zinc finger protein Zif268 D20A mutant. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

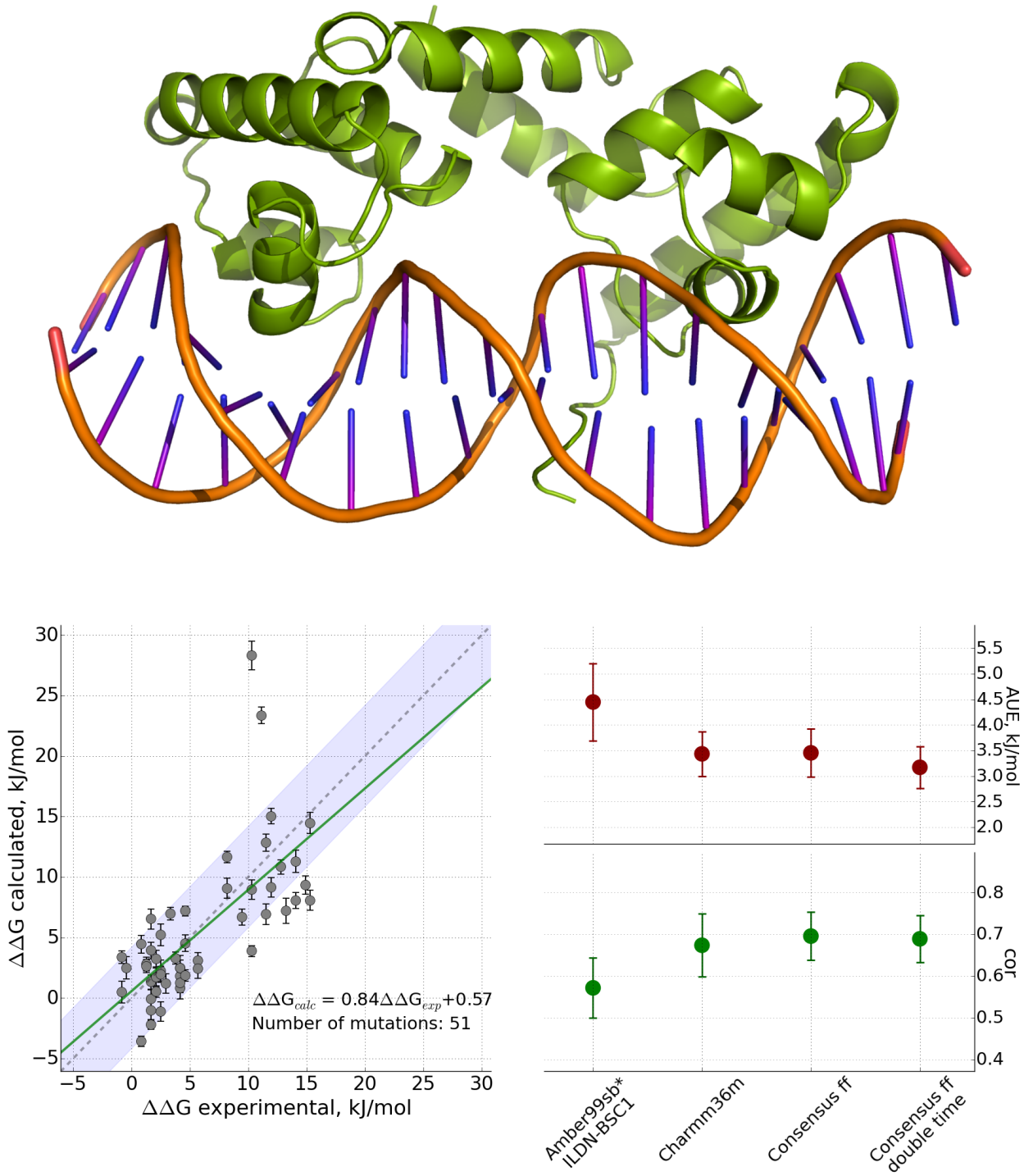

Figure S3: The structure and the calculation results for the DNA bound to the λ repressor λR. Panel on the left shows the experimental $\Delta\Delta$G values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

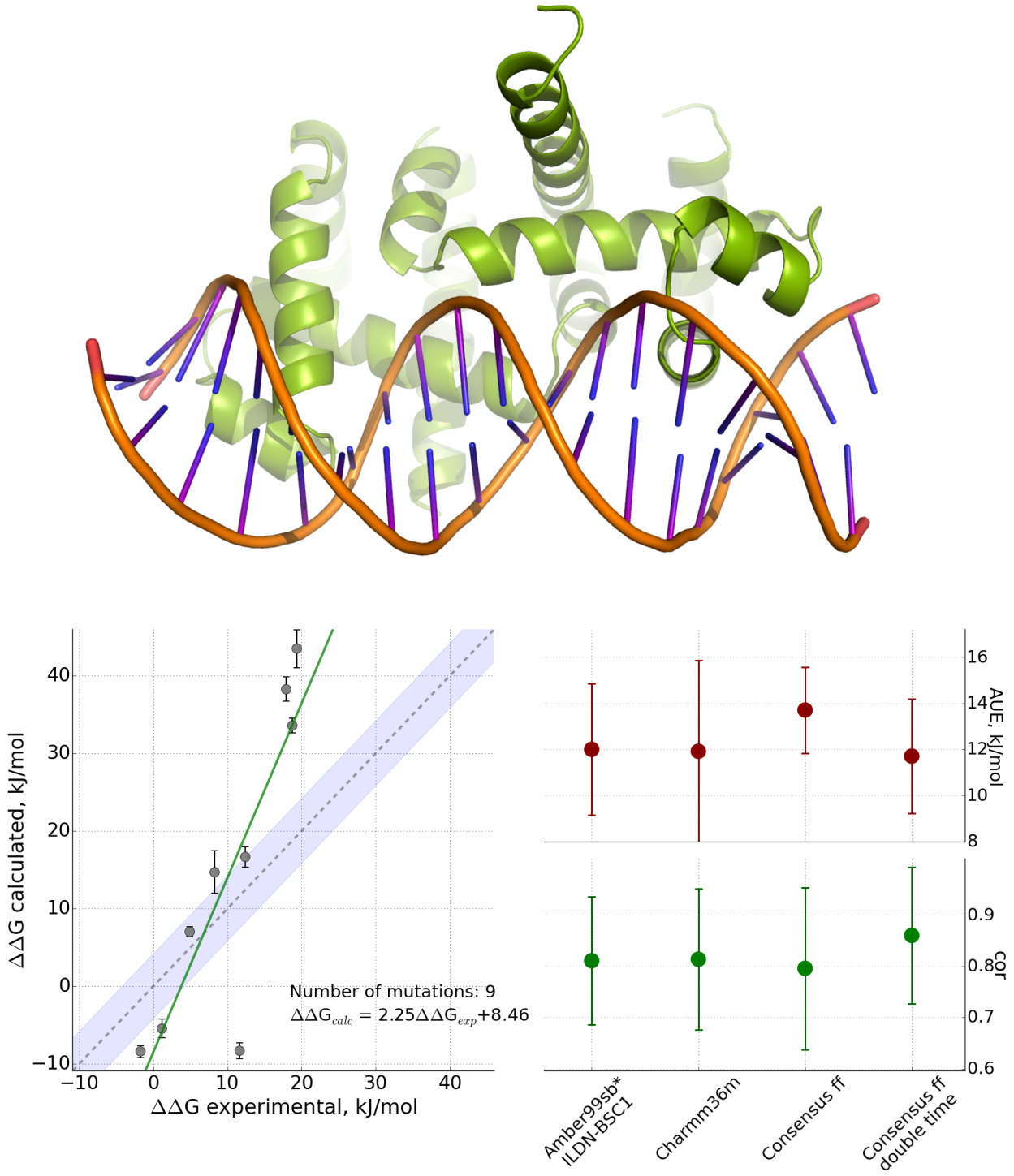

Figure S4: The structure and the calculation results for the DNA bound to the Trp repressor TrpR. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

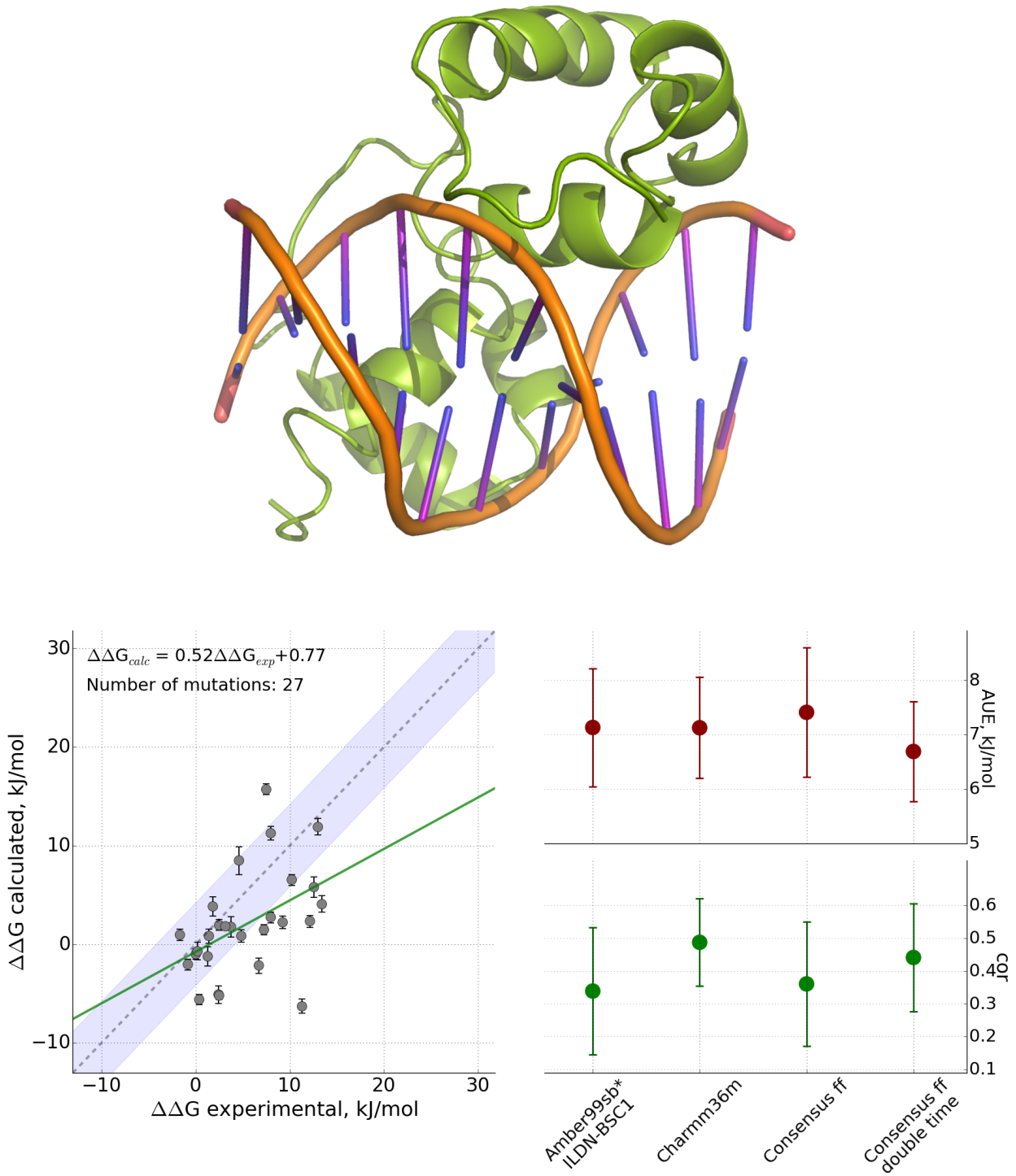

Figure S5: The structure and the calculation results for the DNA bound to the protooncogene c-Myb. Panel on the left shows the experimental $\Delta\Delta G$ values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

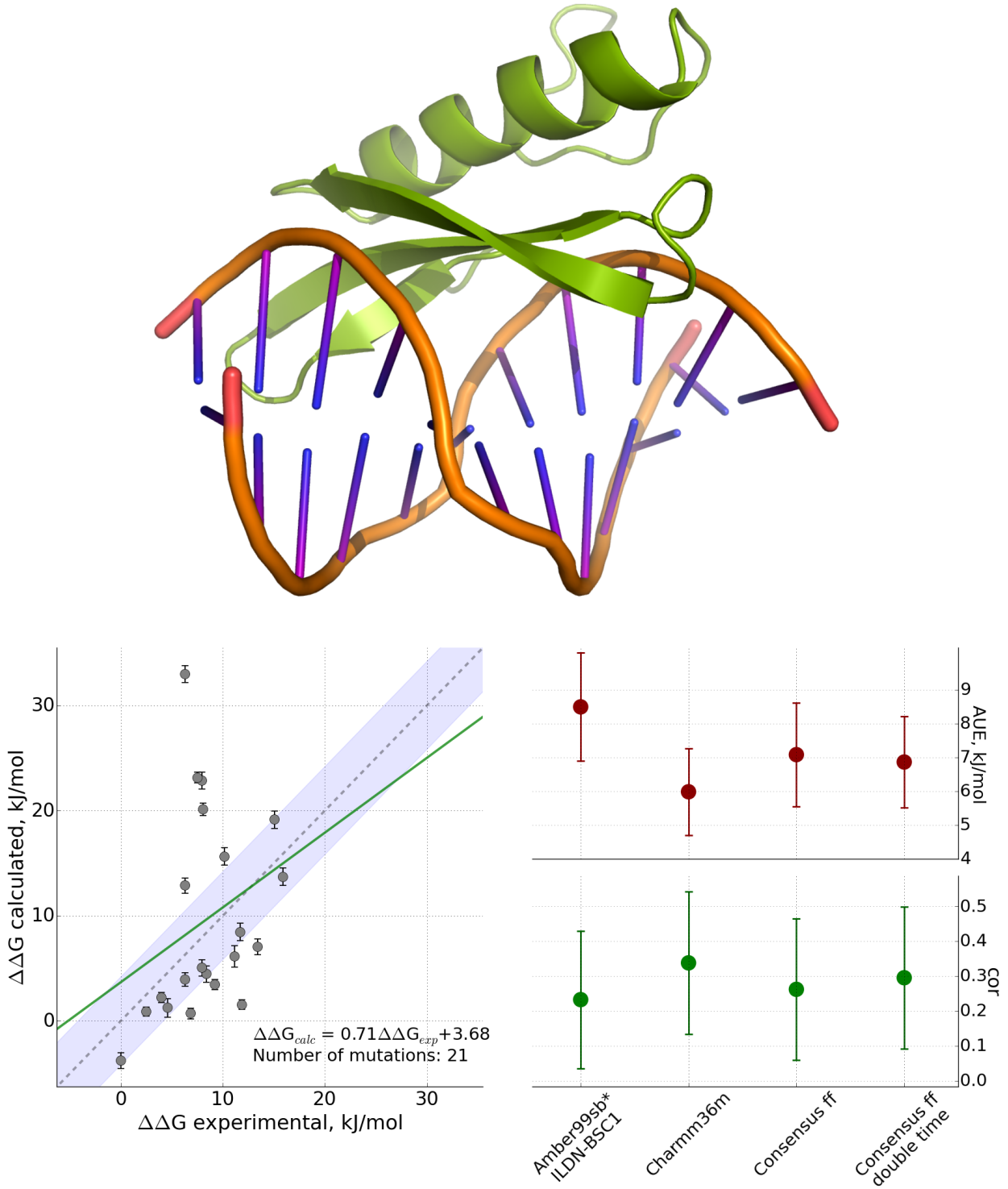

Figure S6: AtERF1. The structure and the calculation results for the DNA bound to the ethylene responsive transcription factor AtERF1. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

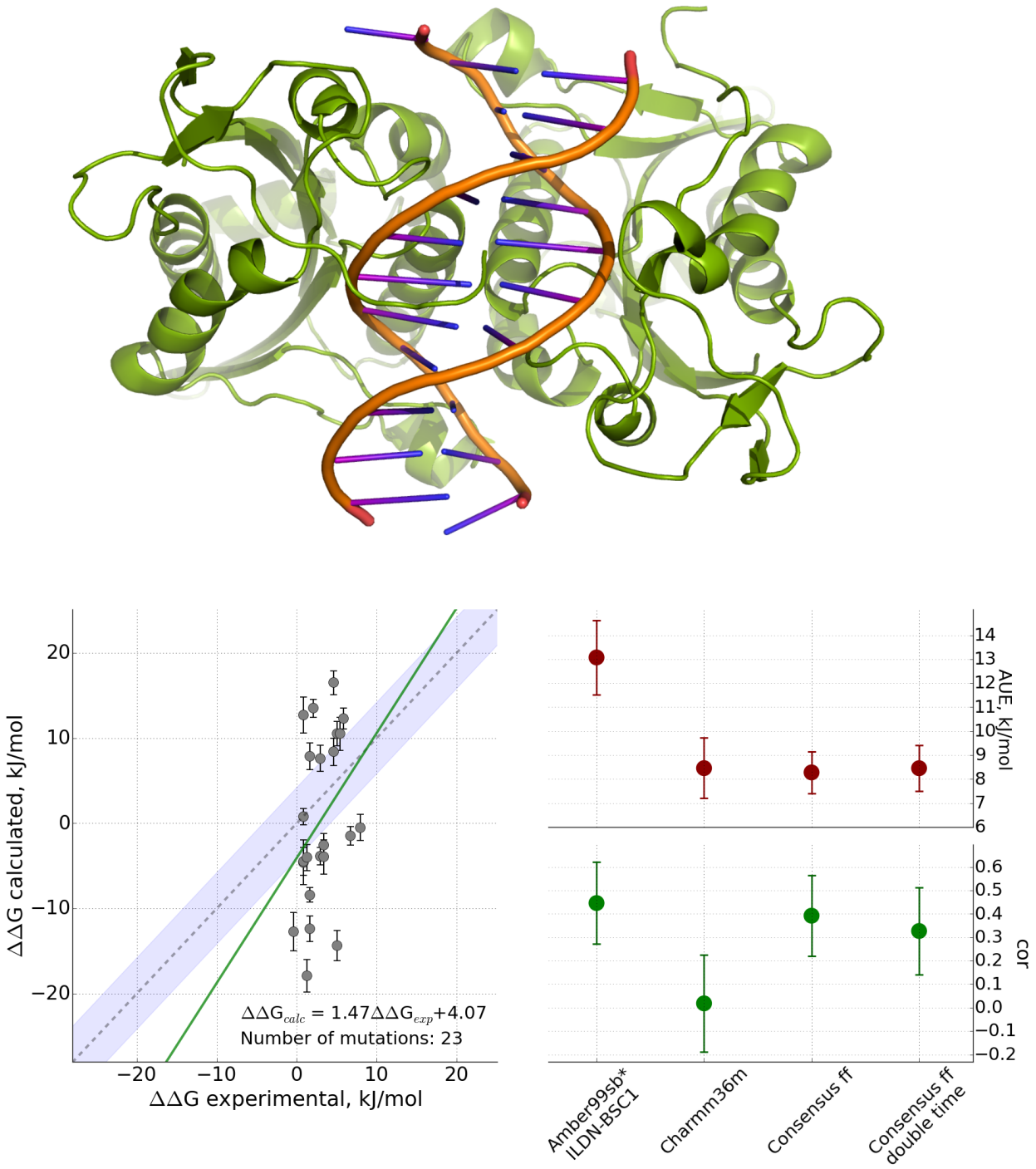

Figure S7: The structure and the calculation results for the DNA bound to the endonuclease BamHI. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

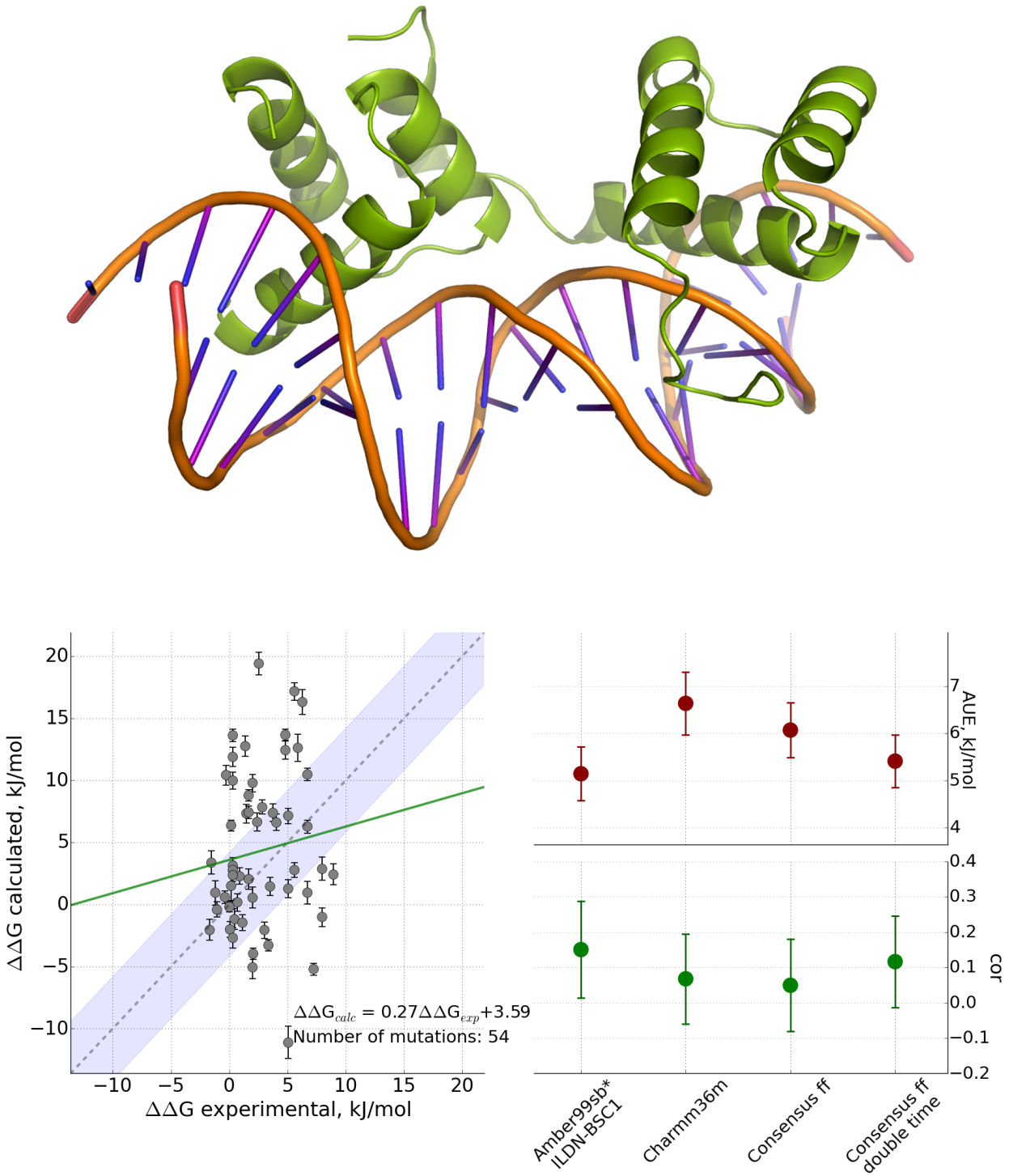

Figure S8: The structure and the calculation results for the DNA bound to the MAT **a**1/α2 homeodomain. Panel on the left shows the experimental $\Delta\Delta$G values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

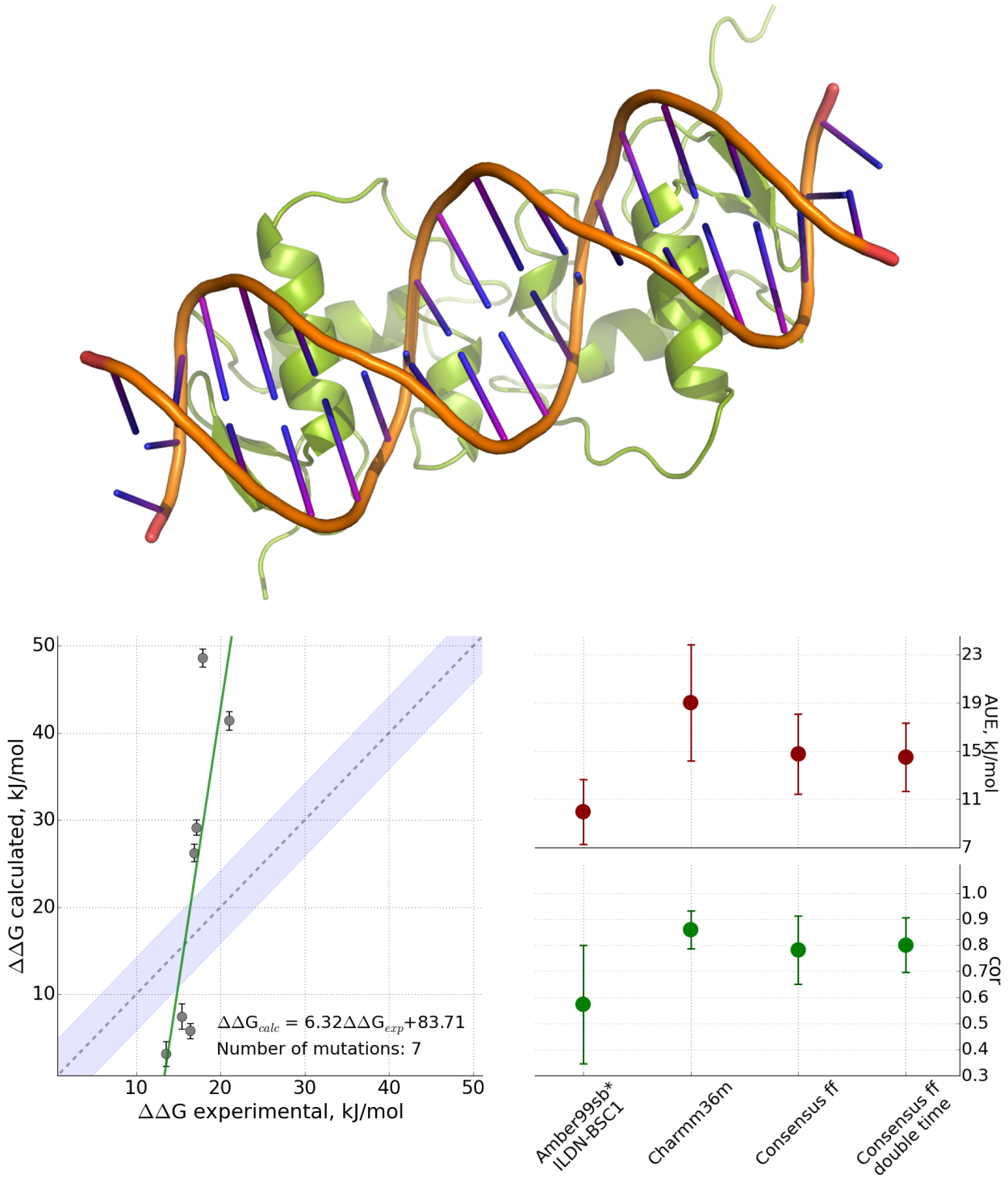

Figure S9: The structure and the calculation results for the DNA bound to the estrogen receptor α (ER). Panel on the left shows the experimental $\Delta\Delta$G values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

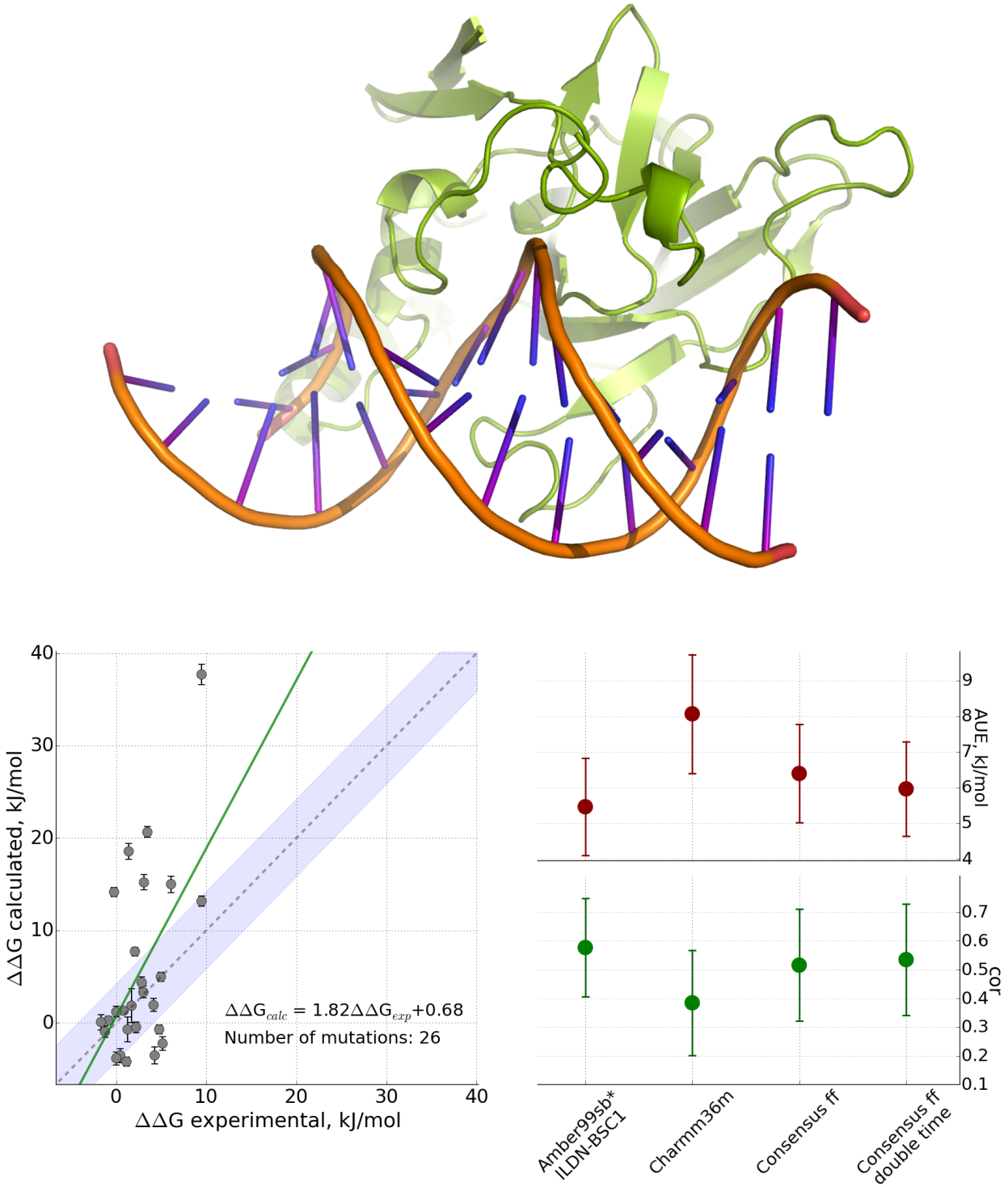

Figure S10: The structure and the calculation results for the DNA bound to the transcriptional activator Ndt80. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

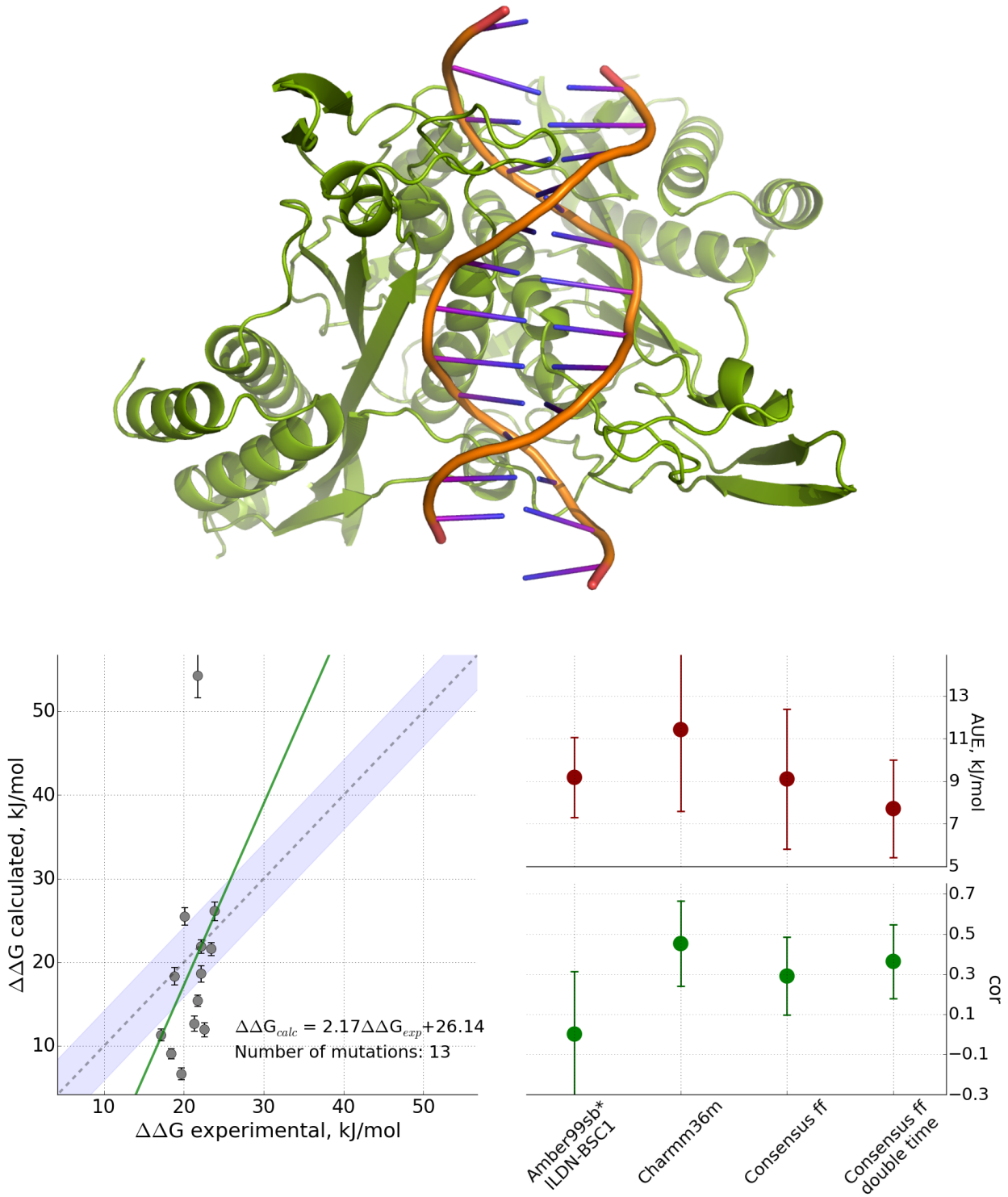

Figure S11: The structure and the calculation results for the DNA bound to the endonuclease EcoRI. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

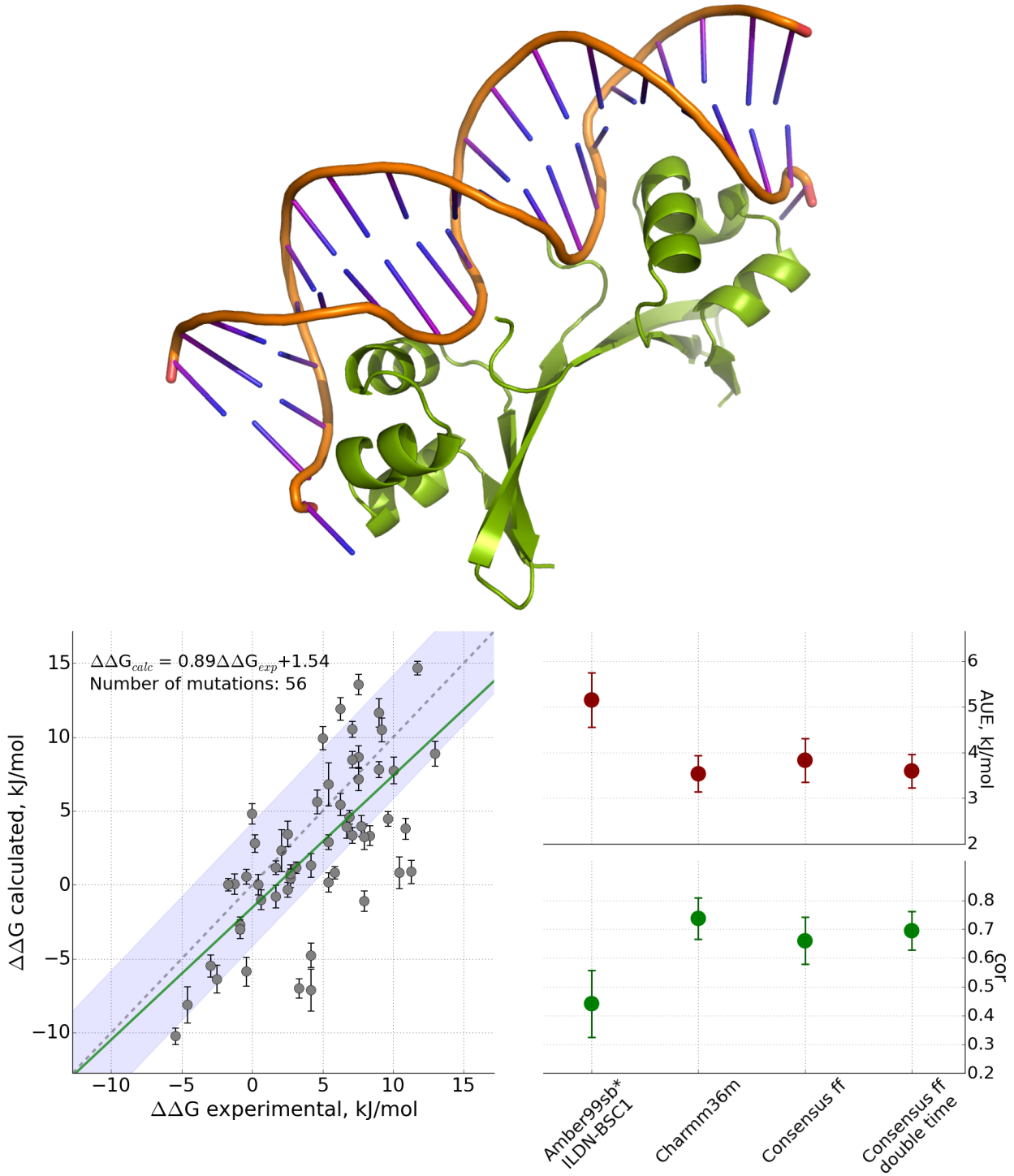

Figure S12: The structure and the calculation results for the DNA bound to the CRO repressor CroR. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

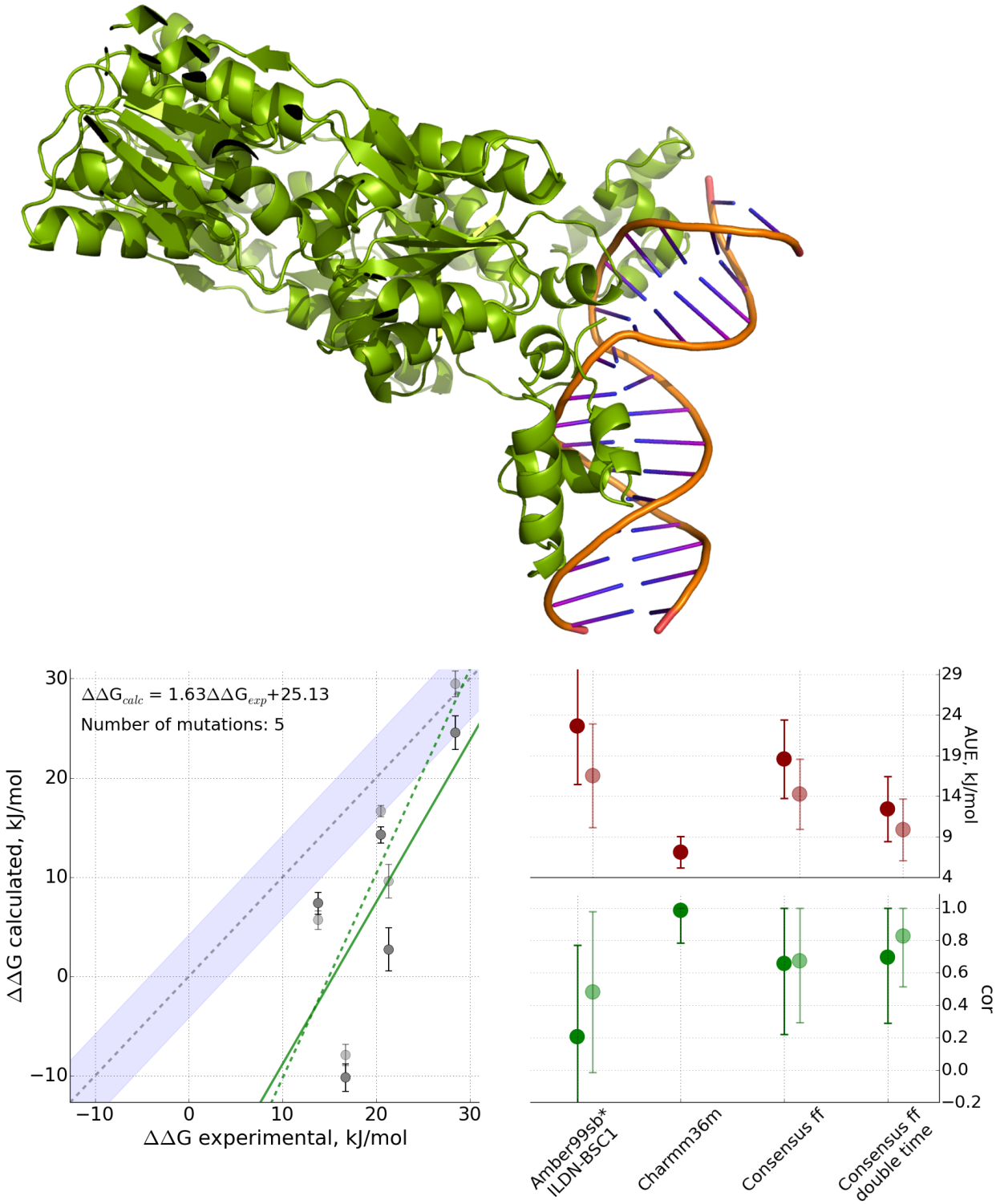

Figure S13: The structure and the calculation results for the DNA bound to the *Lac* repressor LacR. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

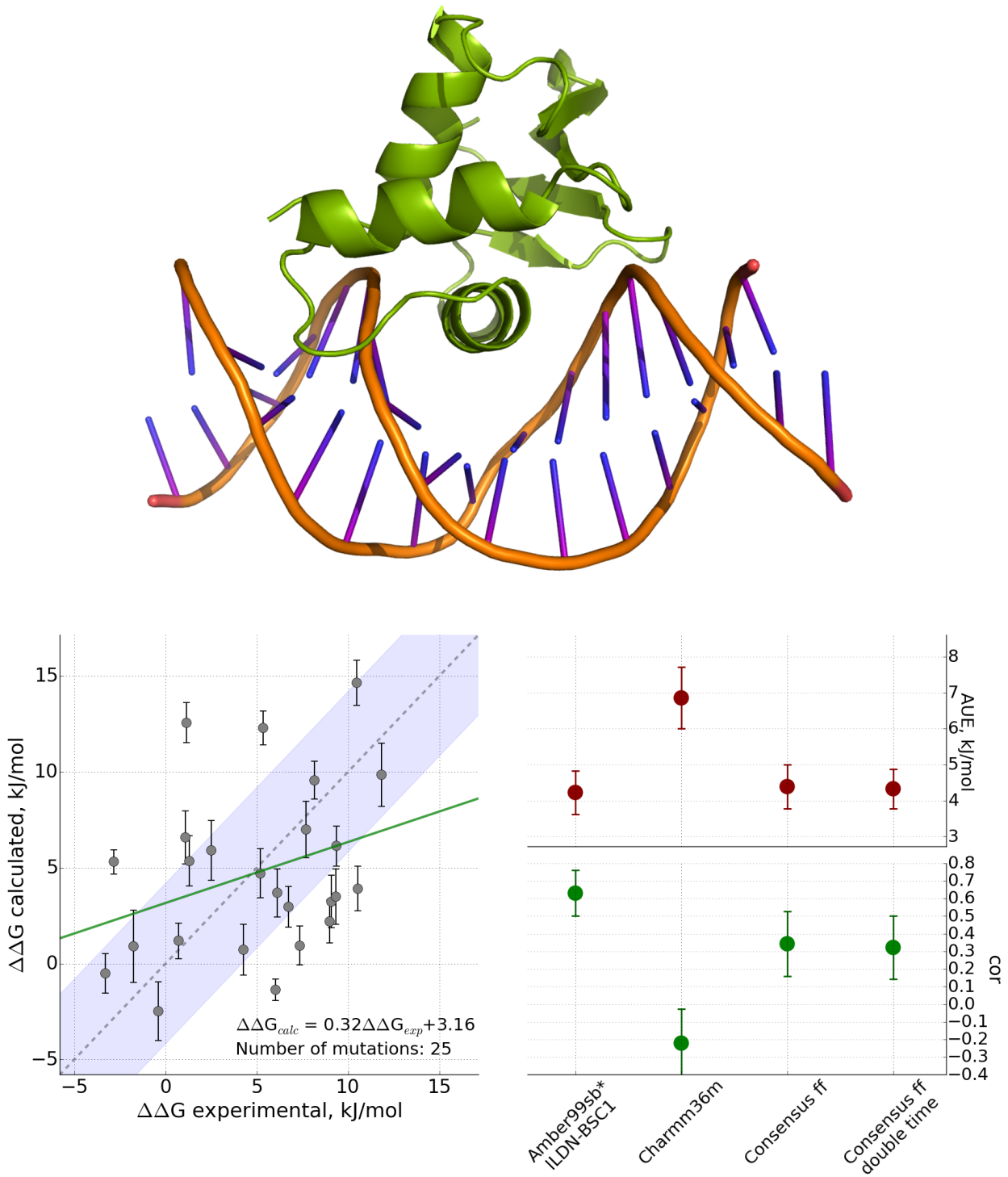

Figure S14: The structure and the calculation results for the DNA bound to the ETS domain of the transcription factor PU.1. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

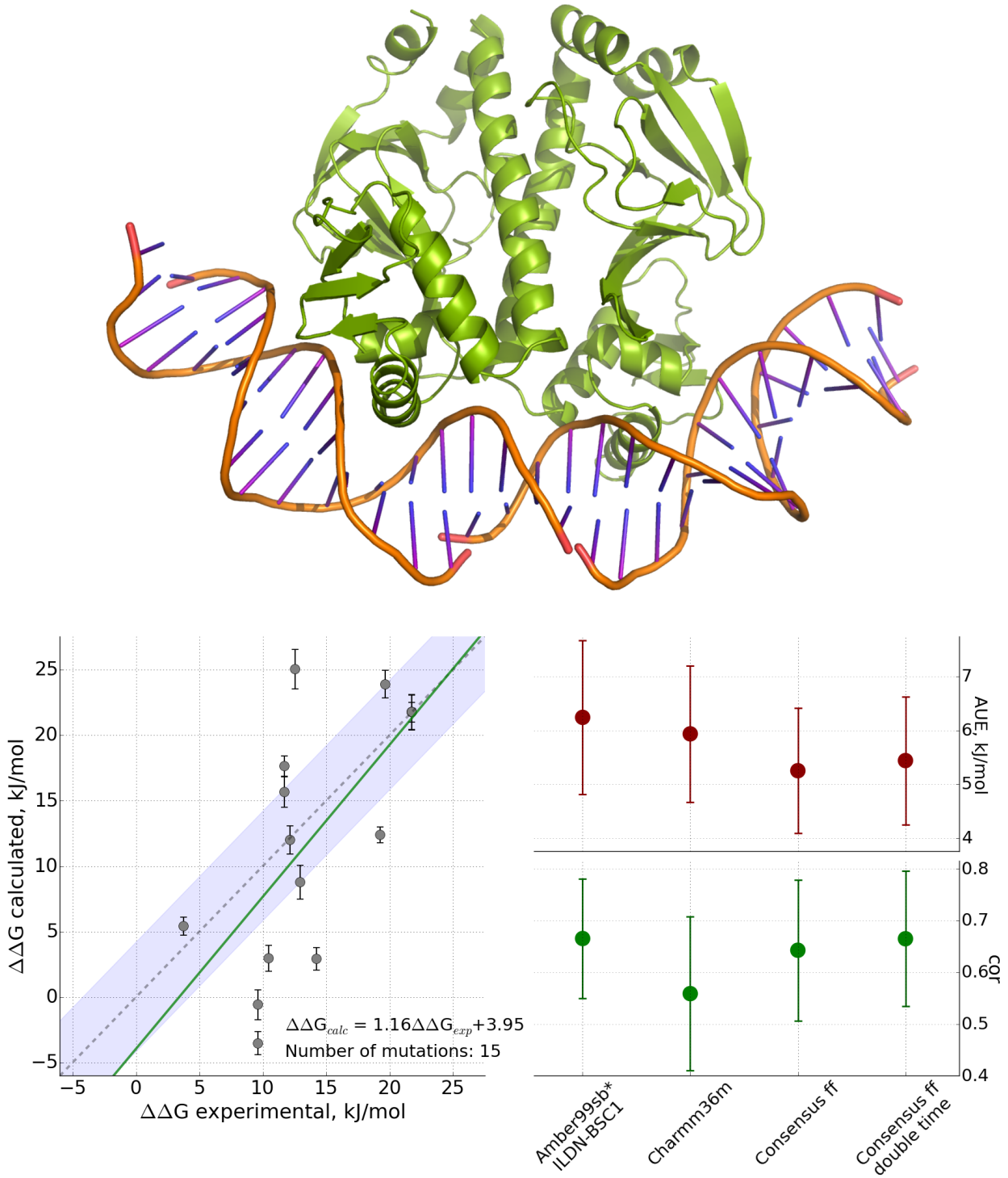

Figure S15: The structure and the calculation results for the DNA bound to the catabolite gene activator protein CAP. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

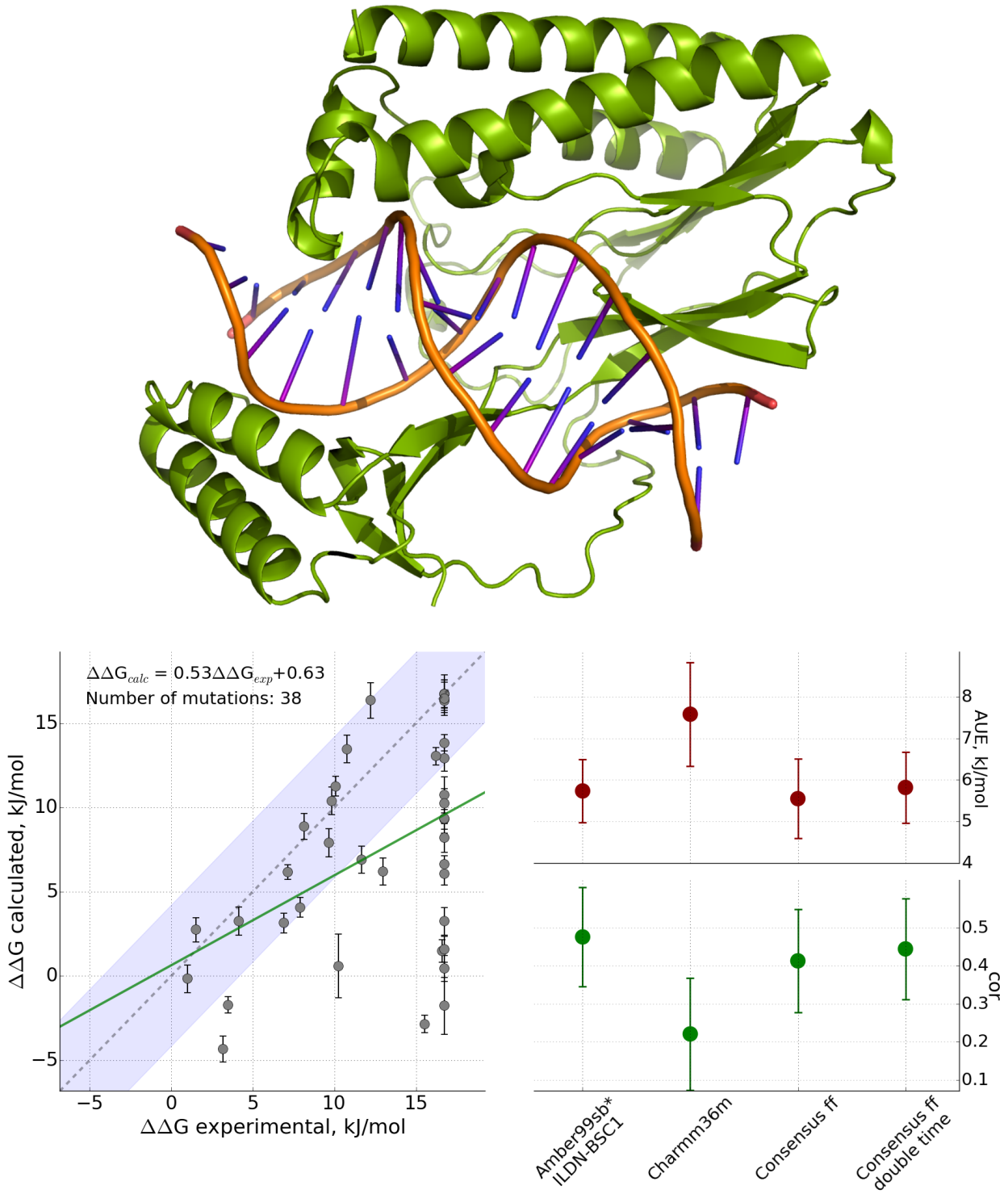

Figure S16: The structure and the calculation results for the DNA bound to the Tus protein. Panel on the left shows the experimental ΔΔG values against the consensus force field (with the doubled simulation time) calculation. Panel on the top-right depicts the average unsigned error (AUE) of the calculations from the experimental results, while panel on the bottom-right shows correlations between the calculations and experiments.

Force field comparison

For a better understanding of the error cancellation effects in the concensus force field approach, we performed a detailed investigation where the discrepancies between Amber99sb*ILDN-BSC1 and Charmm36m occur. For the concensus approach to yield results more accurate than those of the individual force fields taken separately, the errors made by the individual force fields must be pointing in opposite directions. However, this effect is not apparent when all the analyzed data values are pooled together (Fig S17).

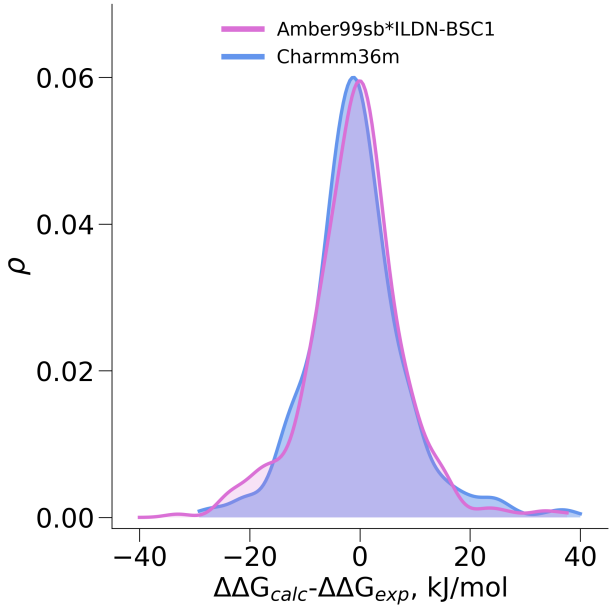

Figure S17: Distributions of the differences between the calculated and experimental ΔΔG values for the Amber99sb*ILDN-BSC1 and Charmm36m force fields.

Clearer trends were revealed by dividing the calculated values into groups, where both force fields were either making an error in the same direction with respect to the experimental $\Delta\Delta$G, or the errors were of opposite signs (Fig S18). From this analysis it is evident that for more than 30% of the investigated mutations the force field results disagree in estimating the $\Delta\Delta$G values to be smaller or larger than the experimental measurement.

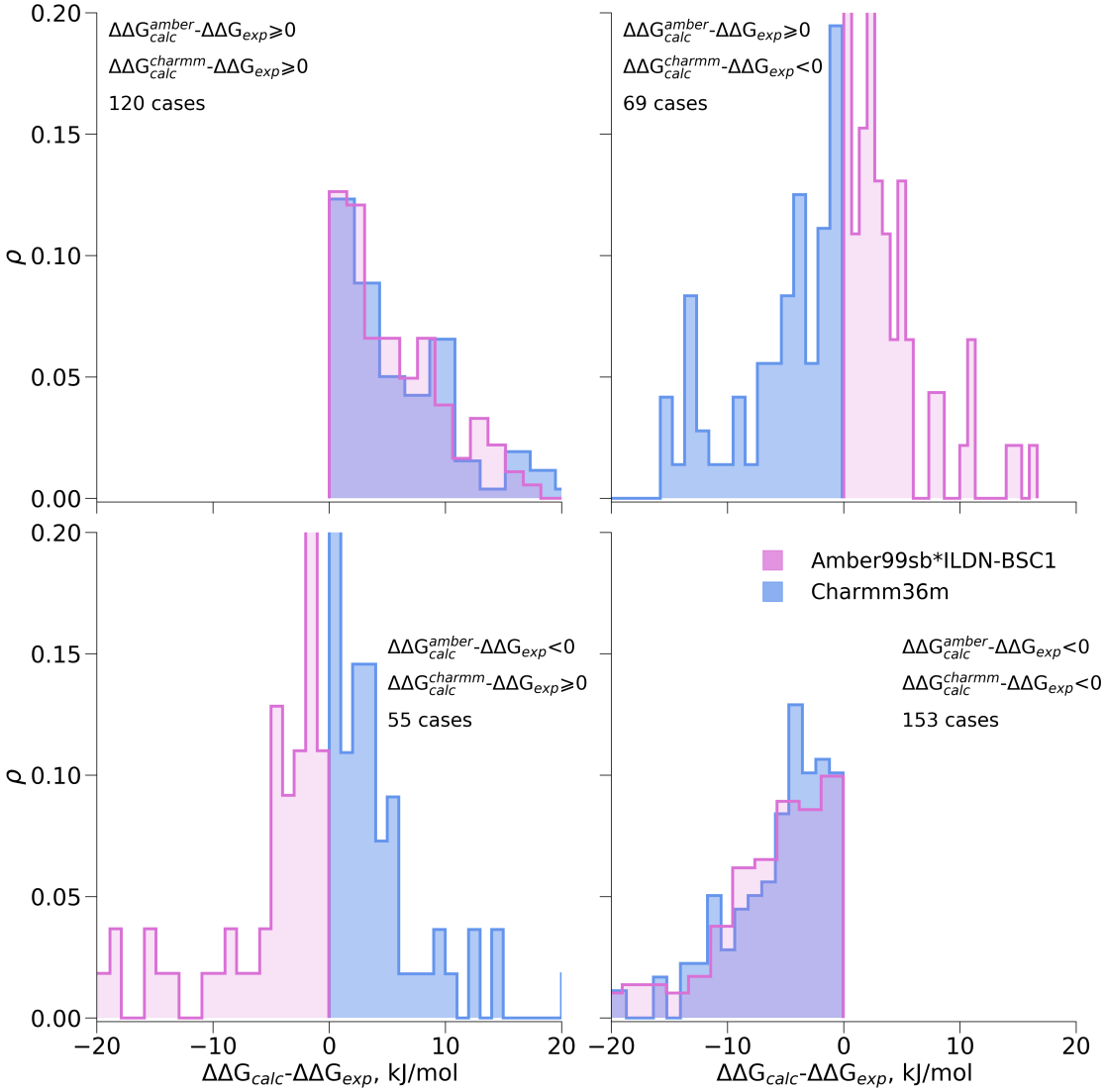

Figure S18: Distributions of the differences between the calculated and experimental $\Delta\Delta$G values divided into groups by the direction of an error made by individual force fields with respect to the experimental value.

The compensatory effects observed in Fig S18, however, cannot be unambiguously ascribed to the individual protein-DNA complexes (Figs S19). It appears that for most of the systems analysed there is no clear trend for the force fields to yield ΔΔG values that would disagree in terms of a direction from the experimental measurement.

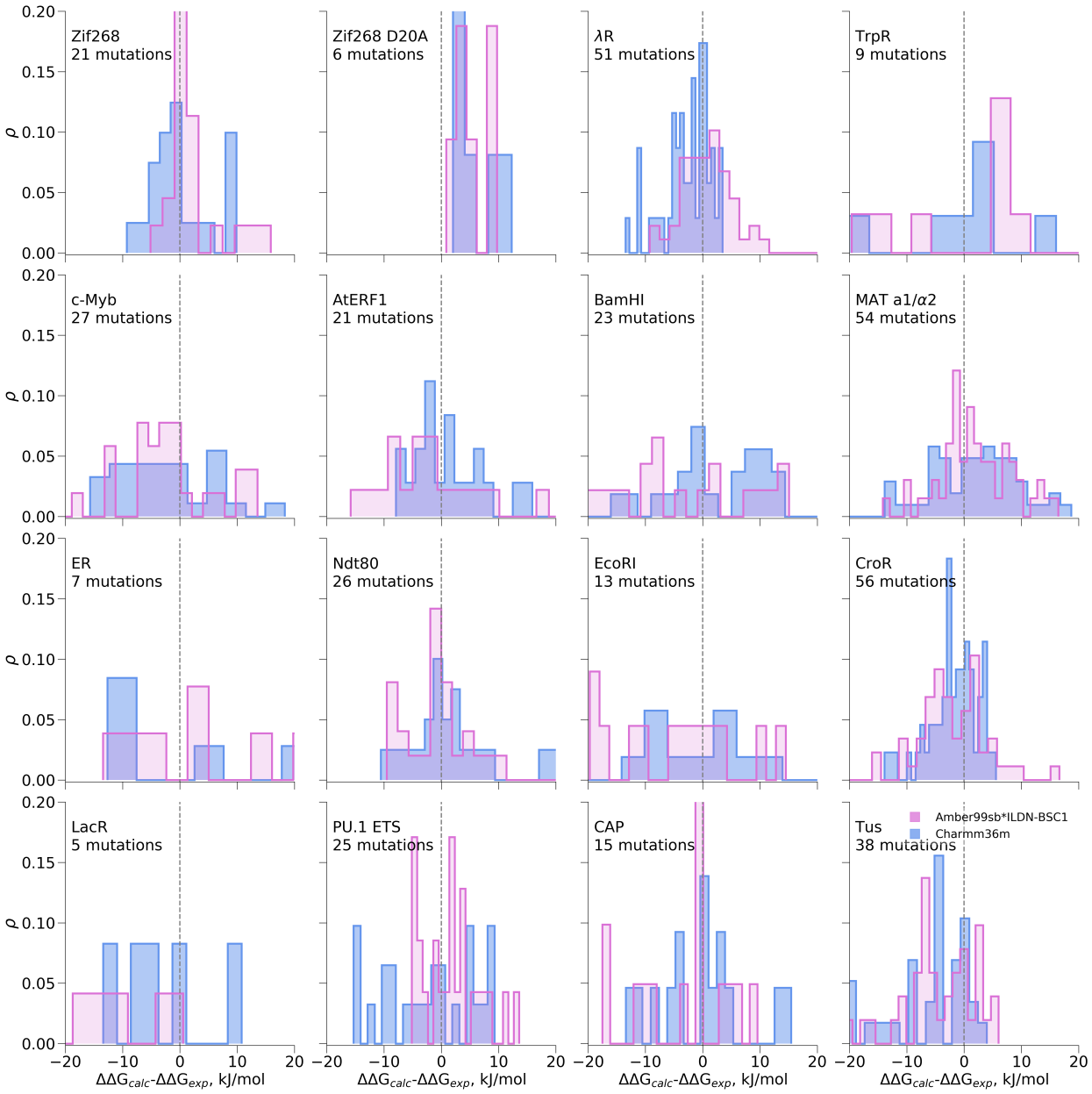

Figure S19: Distributions of the differences between the calculated and experimental ΔΔG values depicted for the individual protein-DNA complexes separately.

While from the distributions of the $\Delta\Delta G_{calc}$-$\Delta\Delta G_{exp}$ (Fig S19) no clear signal could be discerned, averaging of these differences for every protein-DNA system separately, reveals several cases where the force fields on average point in the opposite directions from the experimentally measured values (Fig S20). Namely, for the λR and PU.1 ETS complexes, Amber99sb*ILDN-BSC1 on average predicts $\Delta\Delta G$ to be too high, while Charmm36m estimates too low $\Delta\Delta G$ values; the situation is reversed for the BamHI and EcoRI complexes. These systems could potentially be of interest for the further investigation into the differences between the force fields.

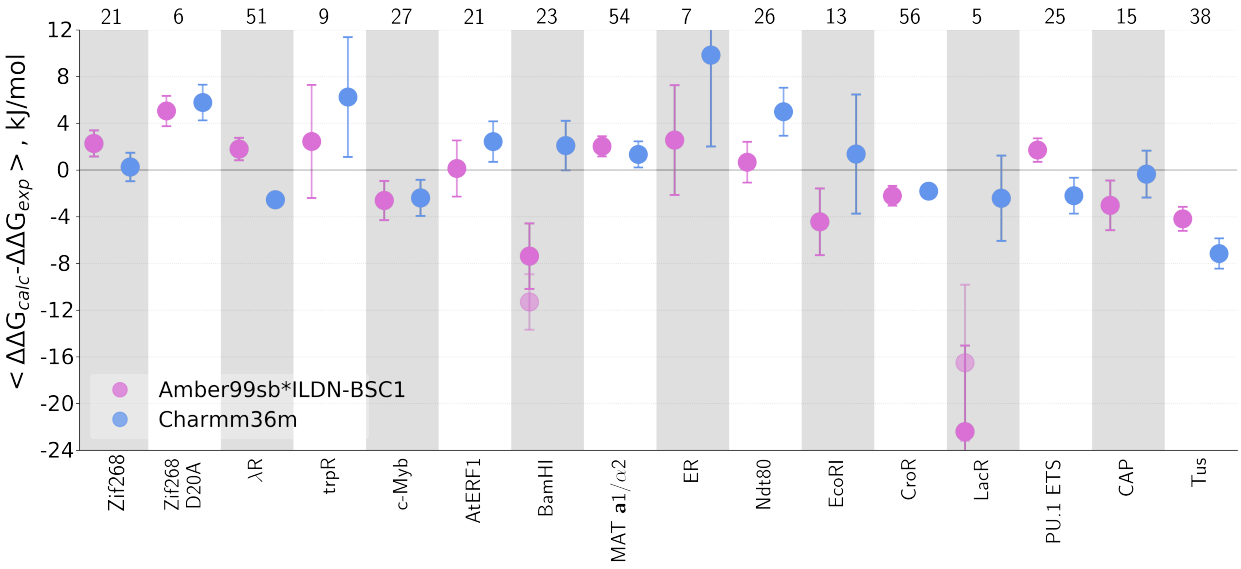

Figure S20: Averages of the absolute differences between the calculated and experimental $\Delta\Delta G$ values for the individual protein-DNA complexes.

Table S1: Experimental and calculated $\Delta\Delta G$ values for the protein-DNA complexes upon nucleotide mutation

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
Zif268				
GCGTGGGCGT				
CCGTGGGCGT	6.64 ± 1.95	13.16 ± 3.50	4.11	[1]
G**A**GTGGGCGT	5.76 ± 1.60	1.14 ± 1.11	4.55	[1]
G**G**GTGGGCGT	4.80 ± 2.59	13.48 ± 1.42	3.86	[1]
G**T**GTGGGCGT	-0.10 ± 0.79	-3.80 ± 0.79	0.57	[1]
GCG**G**GGCGT	6.76 ± 1.74	-6.33 ± 1.58	0.89	[1]
GCGTGGG**A**GT	4.59 ± 2.02	9.99 ± 2.69	6.06	[1]
GCGTGGG**G**GT	9.43 ± 1.97	6.43 ± 2.49	7.72	[1]
GCGTGGG**T**GT	-1.82 ± 1.11	-0.74 ± 0.64	3.33	[1]
GC**C**TGGGCGT	10.32 ± 2.68	9.39 ± 2.42	12.71*	[1]
GCG**A**GGGCGT	12.71 ± 1.89	12.71 ± 1.56	12.71*	[1]
GCG**C**GGGCGT	12.71 ± 0.95	9.84 ± 0.74	12.71*	[1]
GCGT**C**GGCGT	12.71 ± 1.80	12.71 ± 2.01	12.71*	[1]
GCGTG**C**GCGT	12.71 ± 3.98	12.71 ± 4.59	12.71*	[1]
GCGTGG**C**CGT	12.71 ± 3.17	12.71 ± 1.68	12.71*	[1]
GCGTGGGC**C**T	12.71 ± 2.46	7.49 ± 1.59	12.71*	[1]
GCGTGGGCGG				
GCGTGGG**A**GG	5.63 ± 1.95	9.50 ± 1.62	3.82	[2]
GCGTGGG**G**GG	9.79 ± 1.61	7.74 ± 2.39	7.28	[2]
GCGTGGG**T**GG	2.92 ± 1.16	-7.12 ± 0.74	2.15	[2]
GCGTGGGC**A**G	14.60 ± 1.44	1.87 ± 1.50	3.71	[2]
GCGTGGGC**C**G	18.56 ± 2.52	15.18 ± 4.11	5.33	[2]
GCGTGGGC**T**G	20.39 ± 2.00	14.15 ± 1.15	4.55	[2]
Zif268 D20A				
GCGTGGGCGG				
GCGTGGG**A**GG	14.15 ± 1.81	9.66 ± 1.39	5.59	[2]
GCGTGGG**G**GG	11.49 ± 1.76	10.74 ± 1.40	8.72	[2]
GCGTGGG**T**GG	7.40 ± 1.11	15.55 ± 1.48	6.56	[2]
GCGTGGGC**A**G	6.70 ± 1.16	6.12 ± 1.53	4.03	[2]
GCGTGGGC**C**G	16.78 ± 2.07	12.26 ± 1.32	7.06	[2]
GCGTGGGC**T**G	6.97 ± 1.17	13.50 ± 1.45	1.22	[2]
λR				

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
ATACCTCTGGCGGTGATAT				
ATACC**A**CTGGCGGTGATAT	1.98 ± 1.34	-0.98 ± 1.47	-0.84	[3]
ATACCTCTGGCGGTGAT**G**T	6.26 ± 0.74	-3.59 ± 0.64	1.67	[3]
ATACCTCTGGCGGTGAT**T**T	10.46 ± 1.29	2.61 ± 1.18	1.67	[3]
ATACCTCTGGCGGTGAT**C**T	5.74 ± 1.25	-2.34 ± 0.88	2.09	[3]
ATACCTCTGGCGGTGA**G**AT	6.15 ± 2.25	2.91 ± 1.15	4.60	[3]
ATACCTCTGGCGGTGA**A**AT	12.14 ± 1.62	13.57 ± 1.52	11.55	[3]
ATACCTCTGGCGGTGA**C**AT	14.01 ± 1.22	9.30 ± 0.65	8.20	[3]
ATACCTCTGGCGGTG**G**TAT	4.15 ± 1.96	2.62 ± 1.07	-0.84	[3]
ATACCTCTGGCGGTG**T**TAT	3.52 ± 2.32	1.49 ± 1.42	-0.42	[3]
ATACCTCTGGCGGTG**C**TAT	4.62 ± 5.13	4.28 ± 1.54	0.84	[3]
ATACCTCTGGCGGT**A**ATAT	3.85 ± 0.95	1.77 ± 0.92	1.26	[3]
ATACCTCTGGCGGT**T**ATAT	6.58 ± 1.92	-4.97 ± 1.86	4.18	[3]
ATACCTCTGGCGGT**C**ATAT	0.07 ± 2.57	-2.31 ± 2.12	2.51	[3]
ATACCTCTGGCGG**G**GATAT	17.26 ± 1.52	1.07 ± 2.03	11.97	[3]
ATACCTCTGGCGG**A**GATAT	10.51 ± 1.34	7.63 ± 2.12	8.20	[3]
ATACCTCTGGCGG**C**GATAT	8.43 ± 0.85	-0.67 ± 0.92	10.29	[3]
ATACCTCTGGCG**A**TGATAT	15.59 ± 1.16	6.09 ± 0.69	12.80	[3]
ATACCTCTGGCG**T**TGATAT	15.05 ± 1.76	13.88 ± 1.77	15.31	[3]
ATACCTCTGGCG**C**TGATAT	12.03 ± 1.85	4.14 ± 4.99	15.31	[3]
ATACCTCTGGC**A**GTGATAT	20.48 ± 1.08	9.58 ± 1.71	11.97	[3]
ATACCTCTGGC**T**GTGATAT	3.95 ± 1.88	10.50 ± 2.11	13.22	[3]
ATACCTCTGGC**C**GTGATAT	11.90 ± 1.32	10.69 ± 1.05	14.06	[3]
ATACCTCTGG**G**GGTGATAT	44.52 ± 1.77	12.11 ± 2.30	10.29	[3]
ATACCTCTGG**A**GGTGATAT	21.08 ± 2.59	-3.18 ± 1.82	10.29	[3]
ATACCTCTGG**T**GGTGATAT	32.71 ± 1.55	14.02 ± 2.28	11.13	[3]
ATACCTCTG**A**CGGTGATAT	9.40 ± 0.68	4.97 ± 0.69	4.60	[3]
ATACCTCTG**T**CGGTGATAT	9.65 ± 1.64	3.73 ± 1.19	9.46	[3]
ATACCTCTG**C**CGGTGATAT	0.20 ± 1.28	5.96 ± 1.08	5.69	[3]
ATACCTCT**C**GCGGTGATAT	-1.83 ± 1.38	1.71 ± 1.17	1.67	[3]
ATACCTCT**T**GCGGTGATAT	4.86 ± 1.32	1.63 ± 1.09	2.09	[3]
ATACCTCT**A**GCGGTGATAT	0.52 ± 0.94	-7.65 ± 0.71	0.84	[3]
ATACCTC**C**GGCGGTGATAT	3.76 ± 0.72	1.51 ± 0.67	1.26	[3]
ATACCTC**A**GGCGGTGATAT	3.77 ± 1.97	10.06 ± 0.95	11.55	[3]
ATACCTC**G**GGCGGTGATAT	1.86 ± 1.21	1.75 ± 1.03	4.18	[3]
ATACCT**T**TGGCGGTGATAT	0.50 ± 0.92	0.59 ± 0.63	2.09	[3]
ATACCT**A**TGGCGGTGATAT	8.62 ± 2.27	10.06 ± 2.12	14.90	[3]
ATACCT**G**TGGCGGTGATAT	9.51 ± 1.34	6.63 ± 0.99	14.06	[3]
ATACC**C**CTGGCGGTGATAT	3.33 ± 0.70	3.28 ± 0.76	3.77	[3]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
ATACC**G**CTGGCGGTGATAT	0.41 ± 1.71	-2.49 ± 1.34	1.67	[3]
ATAC**T**TCTGGCGGTGATAT	4.06 ± 1.42	0.92 ± 1.05	4.18	[3]
ATAC**A**TCTGGCGGTGATAT	2.03 ± 1.54	2.95 ± 1.66	4.18	[3]
ATAC**G**TCTGGCGGTGATAT	5.38 ± 1.46	5.07 ± 1.63	2.51	[3]
ATA**T**CTCTGGCGGTGATAT	10.85 ± 1.13	3.09 ± 1.17	3.35	[3]
ATA**A**CTCTGGCGGTGATAT	5.83 ± 1.89	-1.31 ± 1.20	2.51	[3]
ATA**G**CTCTGGCGGTGATAT	3.04 ± 1.25	-0.63 ± 1.63	2.93	[3]
AT**C**CCTCTGGCGGTGATAT	1.71 ± 1.18	0.83 ± 0.91	4.18	[3]
AT**T**CCTCTGGCGGTGATAT	4.28 ± 1.74	0.60 ± 3.22	5.69	[3]
AT**G**CCTCTGGCGGTGATAT	0.62 ± 0.93	3.15 ± 0.68	4.60	[3]
A**C**ACCTCTGGCGGTGATAT	-1.62 ± 0.99	-2.80 ± 0.68	1.67	[3]
A**A**ACCTCTGGCGGTGATAT	5.72 ± 1.12	2.14 ± 1.06	1.67	[3]
A**G**ACCTCTGGCGGTGATAT	0.93 ± 1.44	2.99 ± 1.05	2.51	[3]
trpR				
GAACTAGTTAACTAGTAC				
GAACTAGTTAA**GA**AGTAC	-13.82 ± 3.24	2.89 ± 1.72	1.13	[4]
GAACT**TC**TTAACTAGTAC	-8.04 ± 2.59	-8.61 ± 1.67	11.59	[4]
G**T**ACTAGTTAACTAGTAC	-10.25 ± 1.34	-6.62 ± 1.45	-1.80	[4]
GAACTAGTTAACTAGT**T**C	9.66 ± 2.06	4.39 ± 1.14	4.90	[4]
CAACTAGTTAACTAGTA**G**	17.52 ± 4.29	11.86 ± 7.30	8.24	[4]
GAACTAG**AATT**CTAGTAC	18.52 ± 9.62	14.75 ± 3.40	12.43	[4]
GAAC**G**AGTTAACT**C**GTAC	24.46 ± 1.97	42.74 ± 2.42	18.70	[4]
GAA**AG**AGTTAACT**CT**TAC	24.56 ± 5.24	52.02 ± 3.17	17.87	[4]
G**CC**C**CCC**TTA**T**C**ATCATT**	51.77 ± 7.27	35.20 ± 6.81	19.37	[4]
AtERF1				
GAGCCGCCACT				
G**C**GCCGCCACT	-8.61 ± 1.33	1.03 ± 0.82	0.00	[5]
G**G**GCCGCCACT	1.37 ± 0.67	0.40 ± 0.53	2.51	[5]
G**T**GCCGCCACT	1.76 ± 1.72	0.73 ± 1.74	4.55	[5,6]
GA**A**CCGCCACT	-0.42 ± 0.82	7.30 ± 1.22	9.20	[5]
GA**C**CCGCCACT	6.76 ± 2.08	10.15 ± 1.43	11.72	[5]
GA**T**CCGCCACT	2.93 ± 1.46	9.31 ± 1.03	11.15	[5,6]
GAG**A**CGCCACT	12.57 ± 1.26	13.18 ± 2.43	6.28	[5]
GAG**G**CGCCACT	0.91 ± 1.51	7.99 ± 2.90	8.37	[5]
GAG**T**CGCCACT	-8.99 ± 1.15	10.38 ± 0.70	6.83	[5,6]
GAGC**A**GCCACT	25.19 ± 1.97	20.56 ± 1.21	7.95	[5]
GAGC**G**GCCACT	33.68 ± 2.59	32.35 ± 1.51	6.28	[5]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
GAGC**T**GCCACT	16.36 ± 1.64	23.87 ± 1.48	8.09	[5,6]
GAGCC**A**CCACT	32.01 ± 1.61	14.29 ± 0.96	7.53	[5]
GAGCC**C**CCACT	18.70 ± 2.03	19.60 ± 2.70	15.06	[5]
GAGCC**T**CCACT	12.40 ± 1.51	18.86 ± 2.91	10.13	[5,6]
GAGCCG**A**CACT	6.36 ± 1.30	3.71 ± 2.03	7.95	[5]
GAGCCG**G**CACT	7.23 ± 2.04	0.63 ± 1.15	6.28	[5]
GAGCCG**T**CACT	-0.35 ± 0.71	4.74 ± 0.73	3.96	[5,6]
GAGCCGC**A**ACT	9.94 ± 2.41	17.48 ± 1.23	15.90	[5]
GAGCCGC**G**ACT	8.60 ± 1.52	5.46 ± 0.99	13.39	[5]
GAGCCGC**T**ACT	-0.84 ± 0.78	3.92 ± 0.68	11.86	[5,6]
MAT a1-α2				
CATGTAATTAATTACATCA				
AATGTAATTAATTACATCA	0.43 ± 1.27	-10.54 ± 2.96	2.00	[7]
GATGTAATTAATTACATCA	1.94 ± 1.24	-24.16 ± 2.00	5.04	[7]
TATGTAATTAATTACATCA	2.31 ± 0.73	-8.89 ± 1.14	3.36	[7]
C**C**TGTAATTAATTACATCA	0.10 ± 2.42	3.99 ± 0.93	1.68	[7]
C**G**TGTAATTAATTACATCA	-1.49 ± 1.37	-2.57 ± 0.81	0.06	[7]
C**T**TGTAATTAATTACATCA	-3.38 ± 1.70	1.04 ± 1.06	0.47	[7]
CA**A**GTAATTAATTACATCA	0.67 ± 1.73	0.44 ± 1.87	2.00	[7]
CA**C**GTAATTAATTACATCA	-1.60 ± 1.24	-2.56 ± 0.69	3.02	[7]
CA**G**GTAATTAATTACATCA	2.21 ± 1.97	0.33 ± 0.98	5.04	[7]
CAT**A**TAATTAATTACATCA	-6.94 ± 0.94	-3.48 ± 0.92	7.26	[7]
CAT**C**TAATTAATTACATCA	-1.35 ± 1.43	-0.70 ± 1.17	7.96	[7]
CAT**T**TAATTAATTACATCA	-3.34 ± 1.78	5.21 ± 1.42	6.72	[7]
CATG**A**AATTAATTACATCA	2.69 ± 1.96	3.06 ± 1.33	7.96	[7]
CATG**C**AATTAATTACATCA	6.10 ± 1.17	6.43 ± 0.85	6.72	[7]
CATG**G**AATTAATTACATCA	2.52 ± 1.57	2.34 ± 1.19	8.94	[7]
CATGT**C**ATTAATTACATCA	5.18 ± 1.59	9.60 ± 1.37	3.75	[7]
CATGT**G**ATTAATTACATCA	21.24 ± 1.02	6.10 ± 1.34	4.81	[7]
CATGT**T**ATTAATTACATCA	-1.52 ± 1.44	1.84 ± 1.47	0.70	[7]
CATGTA**C**TTAATTACATCA	6.55 ± 1.47	7.72 ± 0.93	5.04	[7]
CATGTA**G**TTAATTACATCA	-5.88 ± 1.14	-2.06 ± 0.73	2.07	[7]
CATGTA**T**TTAATTACATCA	1.71 ± 1.53	4.55 ± 1.38	0.32	[7]
CATGTAA**A**TAATTACATCA	11.93 ± 1.70	8.88 ± 0.85	-0.29	[7]
CATGTAA**C**TAATTACATCA	4.10 ± 0.95	-4.51 ± 0.56	0.00	[7]
CATGTAA**G**TAATTACATCA	3.71 ± 2.01	-0.83 ± 1.13	3.48	[7]
CATGTAAT**A**AATTACATCA	10.26 ± 1.71	4.38 ± 1.21	1.45	[7]
CATGTAAT**C**AATTACATCA	3.76 ± 0.77	8.96 ± 0.52	0.16	[7]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
CATGTAAT**G**AATTACATCA	10.16 ± 1.23	9.39 ± 1.15	2.00	[7]
CATGTAATT**C**ATTACATCA	4.82 ± 1.09	10.89 ± 0.97	2.82	[7]
CATGTAATT**G**ATTACATCA	7.26 ± 0.75	10.31 ± 0.89	1.68	[7]
CATGTAATT**T**ATTACATCA	1.84 ± 1.31	3.69 ± 1.02	0.32	[7]
CATGTAATTA**C**TTACATCA	-1.49 ± 0.92	-1.42 ± 1.85	1.14	[7]
CATGTAATTA**G**TTACATCA	1.94 ± 0.59	2.64 ± 0.82	0.91	[7]
CATGTAATTAA**A**TACATCA	0.57 ± 1.20	-1.36 ± 0.84	-1.04	[7]
CATGTAATTAA**G**TACATCA	2.97 ± 1.36	-7.04 ± 1.14	-1.68	[7]
CATGTAATTAAT**A**ACATCA	-1.45 ± 3.31	-3.94 ± 1.30	0.32	[7]
CATGTAATTAAT**C**ACATCA	4.87 ± 0.79	-3.69 ± 0.69	-0.36	[7]
CATGTAATTAAT**G**ACATCA	13.29 ± 1.34	-7.74 ± 1.26	5.58	[7]
CATGTAATTAATT**C**CATCA	-1.01 ± 1.83	4.03 ± 1.16	0.16	[7]
CATGTAATTAATT**G**CATCA	-0.74 ± 0.70	5.51 ± 0.53	0.32	[7]
CATGTAATTAATT**T**CATCA	13.22 ± 1.94	21.10 ± 1.81	5.58	[7]
CATGTAATTAATTA**A**ATCA	9.57 ± 1.25	10.39 ± 1.22	0.32	[7]
CATGTAATTAATTA**G**ATCA	14.07 ± 1.72	10.80 ± 1.11	4.81	[7]
CATGTAATTAATTA**T**ATCA	4.02 ± 1.04	10.84 ± 0.94	1.68	[7]
CATGTAATTAATTAC**C**TCA	7.52 ± 1.89	16.26 ± 1.01	0.32	[7]
CATGTAATTAATTAC**G**TCA	8.72 ± 1.47	12.26 ± 0.79	6.72	[7]
CATGTAATTAATTAC**T**TCA	15.06 ± 1.53	10.44 ± 1.56	1.39	[7]
CATGTAATTAATTACA**A**CA	12.71 ± 1.44	12.48 ± 1.75	5.91	[7]
CATGTAATTAATTACA**C**CA	6.58 ± 0.88	6.66 ± 0.66	4.06	[7]
CATGTAATTAATTACA**G**CA	14.31 ± 1.41	18.32 ± 1.03	6.28	[7]
CATGTAATTAATTACAT**A**A	7.35 ± 1.78	5.97 ± 1.54	2.38	[7]
CATGTAATTAATTACAT**G**A	17.58 ± 1.62	21.23 ± 1.13	2.54	[7]
CATGTAATTAATTACAT**T**A	13.70 ± 1.41	13.51 ± 1.26	0.32	[7]
CATGTAATTAATTACATC**C**	-0.90 ± 1.38	2.77 ± 1.17	-1.18	[7]
CATGTAATTAATTACATC**T**	2.40 ± 1.51	4.35 ± 1.77	-1.52	[7]
BamHI ATGGATCCATA				
TGGGATCC**CAC**	-23.28 ± 4.07	-2.14 ± 6.09	-0.42	[8]
CGGGATCC**CGC**	-7.03 ± 5.09	-2.17 ± 5.81	0.84	[8]
GCGGATCC**GCC**	-17.31 ± 3.19	8.40 ± 4.09	0.84	[8]
TAGGATCC**TAT**	0.85 ± 6.75	24.58 ± 5.82	0.84	[8]
A**C**GGATCC**G**T**G**	-8.64 ± 1.61	10.23 ± 1.89	0.84	[8]
CTGGATCCA**GC**	-23.27 ± 5.48	-12.53 ± 3.97	1.26	[8]
CCGGATCC**GGG**	-23.56 ± 4.96	15.50 ± 5.64	1.26	[8]
GTGGATCCA**CC**	-21.56 ± 5.93	-3.24 ± 2.72	1.67	[8]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
GTGGATCCA**C**A	-7.03 ± 2.29	-9.76 ± 0.98	1.67	[8]
GCGGATCC**GCG**	3.97 ± 8.09	11.74 ± 3.23	1.67	[8]
A**A**GGATCC**T**TA	13.67 ± 5.33	13.32 ± 3.97	2.09	[8]
TAGGATCC**TA**A	-1.95 ± 4.25	17.13 ± 4.10	2.93	[8]
GTGGATCCA**CG**	-11.43 ± 2.76	3.74 ± 1.52	2.93	[8]
TTGGATCCAA**A**	-7.23 ± 4.02	2.03 ± 3.32	3.35	[8]
CAGGATCC**TGG**	-16.68 ± 5.77	8.78 ± 6.66	3.35	[8]
A**G**GGATCC**C**T**T**	13.01 ± 7.90	3.77 ± 5.16	4.60	[8]
A**A**GGATCC**T**T**G**	19.55 ± 4.69	13.45 ± 2.19	4.60	[8]
GGGGATCC**CCC**	-10.66 ± 5.58	-18.05 ± 7.10	5.02	[8]
A**A**GGATCC**T**T**T**	20.07 ± 5.40	0.96 ± 3.30	5.02	[8]
GAGGATCC**TCT**	15.85 ± 4.39	5.12 ± 4.69	5.44	[8]
GAGGATCC**TCG**	8.67 ± 7.97	15.93 ± 2.91	5.86	[8]
GGGGATCC**CC**A	-0.85 ± 4.13	-2.09 ± 5.80	6.69	[8]
GGGGATCC**CCG**	-14.57 ± 4.81	13.59 ± 5.97	7.95	[8]
Tus-Ter AGTATGTTGTAACT				
A**A**TATGTTGTAACT	-2.91 ± 0.97	-0.54 ± 0.57	3.47	[9]
A**C**TATGTTGTAACT	4.43 ± 1.40	1.04 ± 1.02	1.51	[9]
A**T**TATGTTGTAACT	-5.31 ± 1.40	-3.37 ± 0.94	3.18	[9]
AG**A**ATGTTGTAACT	6.42 ± 1.39	0.09 ± 1.19	4.14	[9]
AG**G**ATGTTGTAACT	-0.24 ± 1.43	-0.10 ± 0.94	1.00	[9]
AGT**C**TGTTGTAACT	14.38 ± 1.80	12.54 ± 0.93	10.75	[9]
AGT**G**TGTTGTAACT	3.54 ± 0.89	2.74 ± 0.55	6.90	[9]
AGT**T**TGTTGTAACT	6.00 ± 1.95	7.80 ± 1.12	11.67	[9]
AGTA**A**GTTGTAACT	10.09 ± 1.65	7.66 ± 1.87	8.16	[9]
AGTA**C**GTTGTAACT	1.41 ± 0.92	6.72 ± 1.09	7.91	[9]
AGTA**G**GTTGTAACT	12.33 ± 1.76	8.44 ± 0.88	9.83	[9]
AGTAT**A**TTGTAACT	12.92 ± 1.02	12.93 ± 1.27	16.74*	[9]
AGTAT**C**TTGTAACT	10.06 ± 2.49	16.74 ± 1.48	16.74*	[9]
AGTAT**T**TTGTAACT	14.65 ± 2.61	16.74 ± 1.33	16.74*	[9]
AGTATG**A**TGTAACT	12.03 ± 1.33	3.77 ± 1.55	9.67	[9]
AGTATG**C**TGTAACT	8.49 ± 0.96	14.03 ± 1.36	10.08	[9]
AGTATG**G**TGTAACT	18.23 ± 1.54	14.49 ± 1.79	12.22	[9]
AGTATGT**A**GTAACT	16.74 ± 3.08	16.74 ± 2.82	16.74*	[9]
AGTATGT**C**GTAACT	10.50 ± 1.60	8.08 ± 0.99	16.74*	[9]
AGTATGT**G**GTAACT	11.50 ± 3.10	7.18 ± 2.30	16.74*	[9]
AGTATGTT**A**TAACT	16.74 ± 2.20	7.47 ± 1.84	16.74*	[9]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
AGTATGTT**C**TAACT	0.22 ± 2.54	16.16 ± 1.27	16.74*	[9]
AGTATGTT**T**TAACT	9.73 ± 3.35	-3.26 ± 1.50	16.74*	[9]
AGTATGTTG**A**AACT	16.74 ± 2.97	3.31 ± 2.21	16.74*	[9]
AGTATGTTG**C**AACT	16.74 ± 1.15	-4.64 ± 4.31	16.74*	[9]
AGTATGTTG**G**AACT	10.62 ± 1.79	-9.74 ± 3.08	16.74*	[9]
AGTATGTTGT**C**ACT	0.87 ± 1.63	11.53 ± 2.23	12.97	[9]
AGTATGTTGT**G**ACT	15.79 ± 0.90	11.86 ± 1.68	16.74*	[9]
AGTATGTTGT**T**ACT	11.17 ± 1.41	0.97 ± 2.59	16.74*	[9]
AGTATGTTGTA**C**CT	8.41 ± 1.39	-5.44 ± 1.97	16.61	[9]
AGTATGTTGTA**G**CT	21.71 ± 1.01	4.37 ± 1.03	16.23	[9]
AGTATGTTGTA**T**CT	5.88 ± 2.18	-2.70 ± 3.59	16.74*	[9]
AGTATGTTGTAA**A**T	9.90 ± 1.87	11.60 ± 1.23	16.74*	[9]
AGTATGTTGTAA**G**T	8.81 ± 1.69	11.69 ± 2.27	16.74*	[9]
AGTATGTTGTAA**T**T	9.41 ± 1.36	2.92 ± 0.74	7.15	[9]
AGTATGTTGTAAC**A**	-4.10 ± 1.77	0.54 ± 1.74	16.74*	[9]
AGTATGTTGTAAC**C**	-2.24 ± 0.69	-3.47 ± 0.65	15.52	[9]
AGTATGTTGTAAC**G**	-0.32 ± 2.20	1.50 ± 2.27	10.25	[9]
LacR				
TGTGAGCGCTCACAATT				
TGT**A**AGCGCT**T**ACAATT	-30.88 ± 4.39	10.54 ± 4.83	16.74	[10]
TG**A**GAGCGCTC**T**CAATT	14.35 ± 3.22	0.40 ± 3.37	13.81	[10]
TG**C**GAGCGCTC**G**CAATT	7.08 ± 4.62	21.43 ± 1.30	20.50	[10]
TG**AA**AGCGCT**TT**CAATT	-11.69 ± 5.22	17.13 ± 4.87	21.34	[10]
TG**CA**AGCGCT**TG**CAATT	9.92 ± 3.98	39.22 ± 4.19	28.45	[10]
ER				
GAGGTCACAGTGACCTG				
G**T**GGTCACAGTGACC**A**G	12.10 ± 2.58	2.75 ± 2.21	15.40	[11]
GA**C**GTCACAGTGAC**G**TG	41.32 ± 2.34	55.90 ± 3.70	17.88	[11]
GAG**A**TCACAGTGA**T**CTG	18.85 ± 2.70	39.39 ± 1.69	17.14	[11]
GAGG**A**CACAGTG**T**CCTG	31.77 ± 4.84	20.65 ± 2.30	16.87	[11]
GAGGT**G**ACAGT**C**ACCTG	24.82 ± 2.67	57.95 ± 2.47	21.02	[11]
GAGGTC**T**CAG**A**GACCTG	2.99 ± 4.39	8.53 ± 2.42	16.42	[11]
G**G**GG**C**C**GGGC**TGACC**C**G	4.33 ± 4.20	1.95 ± 3.14	13.55	[11]
CroR				
CTATCACCGCCAGAGGTAC				
C**A**ATCACCGCCAGAGGTAC	8.18 ± 1.36	3.05 ± 1.24	4.60	[12]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
C**G**ATCACCGCCAGAGGTAC	2.97 ± 1.44	7.90 ± 1.24	6.28	[12]
C**C**ATCACCGCCAGAGGTAC	3.10 ± 0.80	2.68 ± 0.91	5.44	[12]
CT**G**TCACCGCCAGAGGTAC	13.61 ± 0.90	15.71 ± 0.56	11.72	[12]
CT**T**TCACCGCCAGAGGTAC	-11.85 ± 2.20	13.49 ± 3.19	10.46	[12]
CT**C**TCACCGCCAGAGGTAC	2.74 ± 1.37	15.02 ± 1.01	12.97	[12]
CTA**G**CACCGCCAGAGGTAC	11.96 ± 1.22	11.84 ± 1.74	6.28	[12]
CTA**A**CACCGCCAGAGGTAC	4.98 ± 1.98	1.91 ± 2.40	2.51	[12]
CTA**C**CACCGCCAGAGGTAC	12.83 ± 1.01	8.24 ± 1.18	7.11	[12]
CTAT**G**ACCGCCAGAGGTAC	3.43 ± 2.20	-1.66 ± 1.79	11.30	[12]
CTAT**A**ACCGCCAGAGGTAC	2.50 ± 4.18	12.99 ± 1.47	10.04	[12]
CTAT**T**ACCGCCAGAGGTAC	1.82 ± 0.69	-0.18 ± 0.54	5.86	[12]
CTATC**C**CCGCCAGAGGTAC	1.20 ± 2.21	6.63 ± 2.08	6.69	[12]
CTATC**T**CCGCCAGAGGTAC	11.09 ± 1.23	8.74 ± 1.66	5.02	[12]
CTATC**G**CCGCCAGAGGTAC	6.79 ± 1.23	-0.08 ± 1.13	7.11	[12]
CTATCA**G**CGCCAGAGGTAC	7.72 ± 1.19	13.26 ± 1.56	9.20	[12]
CTATCA**A**CGCCAGAGGTAC	9.04 ± 2.01	8.26 ± 1.70	7.53	[12]
CTATCA**T**CGCCAGAGGTAC	4.74 ± 0.85	4.20 ± 0.66	9.62	[12]
CTATCAC**T**GCCAGAGGTAC	-0.63 ± 0.73	2.97 ± 0.67	1.67	[12]
CTATCAC**G**GCCAGAGGTAC	2.34 ± 1.19	4.29 ± 0.88	8.37	[12]
CTATCAC**A**GCCAGAGGTAC	3.79 ± 1.39	3.81 ± 0.93	10.88	[12]
CTATCACC**C**CCAGAGGTAC	4.85 ± 1.07	0.75 ± 1.26	0.21	[12]
CTATCACC**T**CCAGAGGTAC	-0.88 ± 1.41	-4.57 ± 1.03	-0.84	[12]
CTATCACC**A**CCAGAGGTAC	3.49 ± 0.75	-2.39 ± 0.63	-0.42	[12]
CTATCACCG**G**CAGAGGTAC	1.02 ± 1.11	-1.00 ± 0.90	0.42	[12]
CTATCACCG**T**CAGAGGTAC	4.58 ± 0.63	-5.22 ± 0.71	2.51	[12]
CTATCACCG**A**CAGAGGTAC	4.63 ± 1.04	-3.73 ± 1.06	2.72	[12]
CTATCACCGC**T**AGAGGTAC	-2.72 ± 0.95	-3.30 ± 0.66	-0.84	[12]
CTATCACCGC**A**AGAGGTAC	0.99 ± 2.50	-0.86 ± 0.83	-1.26	[12]
CTATCACCGC**G**AGAGGTAC	8.89 ± 1.99	0.72 ± 1.07	0.00	[12]
CTATCACCGCC**T**GAGGTAC	-7.23 ± 1.49	5.02 ± 1.19	7.95	[12]
CTATCACCGCC**C**GAGGTAC	-0.64 ± 1.09	0.97 ± 1.03	5.44	[12]
CTATCACCGCC**G**GAGGTAC	0.29 ± 0.71	-0.28 ± 0.57	-1.67	[12]
CTATCACCGCCA**A**AGGTAC	9.10 ± 0.75	6.50 ± 0.60	9.00	[12]
CTATCACCGCCA**T**AGGTAC	15.96 ± 1.58	11.18 ± 1.75	7.53	[12]
CTATCACCGCCA**C**AGGTAC	10.32 ± 1.08	12.98 ± 1.70	9.00	[12]
CTATCACCGCCAG**C**GGTAC	-4.20 ± 1.95	-5.34 ± 1.15	4.18	[12]
CTATCACCGCCAG**T**GGTAC	-7.73 ± 2.37	-5.03 ± 1.00	-2.51	[12]
CTATCACCGCCAG**G**GGTAC	-12.49 ± 1.69	-1.52 ± 1.39	3.35	[12]
CTATCACCGCCAGA**A**GTAC	9.37 ± 1.06	7.56 ± 0.74	7.11	[12]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
CTATCACCGCCAGA**T**GTAC	1.73 ± 2.87	4.78 ± 1.13	7.95	[12]
CTATCACCGCCAGA**C**GTAC	-0.93 ± 2.44	3.59 ± 1.33	4.18	[12]
CTATCACCGCCAGAG**A**TAC	-10.13 ± 1.15	-10.36 ± 0.75	-5.44	[12]
CTATCACCGCCAGAG**T**TAC	-6.85 ± 1.23	-4.11 ± 1.19	-2.93	[12]
CTATCACCGCCAGAG**C**TAC	-3.26 ± 1.57	1.70 ± 1.25	1.67	[12]
CTATCACCGCCAGAGG**G**AC	2.34 ± 1.46	5.62 ± 1.03	7.74	[12]
CTATCACCGCCAGAGG**A**AC	3.01 ± 3.54	11.32 ± 1.72	7.53	[12]
CTATCACCGCCAGAGG**C**AC	4.52 ± 1.04	4.62 ± 0.84	6.90	[12]
CTATCACCGCCAGAGGT**G**C	0.53 ± 0.63	1.79 ± 0.64	3.14	[12]
CTATCACCGCCAGAGGT**C**C	1.66 ± 1.25	-0.14 ± 1.21	2.72	[12]
CTATCACCGCCAGAGGT**T**C	0.91 ± 1.29	-2.93 ± 1.15	0.63	[12]
CTA**A**CACCG**TGC**GAG**T**T**G**C	22.07 ± 3.82	-8.48 ± 5.61	5.44	[12]
CTATCACCGC**A**AG**G**G**A**TAC	-8.79 ± 5.51	-7.47 ± 1.93	-4.60	[12]
CTATCACCGCCAG**T**GGTAC	-10.63 ± 1.58	-1.11 ± 1.05	-0.42	[12]
C**C**A**A**CACCGCCAGAG**A**TAC	-0.45 ± 3.65	5.06 ± 3.66	2.09	[12]
CTATCACCGC**AGAT**GGT**T**C	-6.39 ± 5.74	-7.85 ± 6.52	4.18	[12]
EcoRI				
GCAGAATTCTGC				
GCA**T**AATTCTGC	31.65 ± 4.27	-9.06 ± 3.44	17.15	[13]
GCA**A**AATTCTGC	7.19 ± 0.76	10.93 ± 1.48	18.41	[13]
GCA**C**AATTCTGC	21.85 ± 1.67	8.94 ± 1.47	21.76	[13]
GCAG**C**ATTCTGC	17.39 ± 1.69	26.37 ± 2.18	22.18	[13]
GCAG**G**ATTCTGC	20.89 ± 1.07	22.26 ± 1.48	23.43	[13]
GCAG**T**ATTCTGC	15.51 ± 2.22	21.12 ± 2.33	18.83	[13]
GCAGA**C**TTCTGC	11.24 ± 1.72	14.08 ± 1.45	21.34	[13]
GCAGA**G**TTCTGC	0.06 ± 1.18	13.25 ± 2.35	19.66	[13]
GCAGA**T**TTCTGC	5.32 ± 1.75	18.52 ± 3.74	22.59	[13]
GCAG**G**AT**C**CTGC	23.84 ± 2.01	27.10 ± 2.22	20.08	[13]
GCAGA**GC**TCTGC	2.44 ± 1.84	34.76 ± 1.69	22.18	[13]
GCA**A**AATT**A**TGC	25.39 ± 2.85	26.82 ± 5.06	23.85	[13]
GCA**CTTAAG**TGC	32.69 ± 12.91	75.82 ± 8.63	21.76	[13]
CAP				
CGAAAAATGTGATCTAGATCACATTTTTCG				
CGAAAAA**A**GTGATCTAGATCAC**T**TTTTTCG	1.18 ± 3.59	4.73 ± 2.15	10.46	[14]
CGAAAAA**C**GTGATCTAGATCAC**G**TTTTTCG	3.72 ± 1.55	7.12 ± 0.99	3.77	[14]
CGAAAAA**G**GTGATCTAGATCAC**C**TTTTTCG	-7.89 ± 2.48	0.85 ± 4.67	9.62	[14]
CGAAAAAT**A**TGATCTAGATCA**T**ATTTTTCG	9.60 ± 0.93	15.19 ± 1.69	19.25	[14]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
CGAAAAAT**C**TGATCTAGATCA**G**ATTTTTCG	20.96 ± 4.71	21.76 ± 2.85	21.76*	[14]
CGAAAAAT**T**TGATCTAGATCA**A**ATTTTTCG	22.11 ± 1.82	27.97 ± 2.61	12.55	[14]
CGAAAAATG**A**GATCTAGATC**T**CATTTTTCG	8.40 ± 4.60	15.62 ± 1.71	12.13	[14]
CGAAAAATG**C**GATCTAGATC**G**CATTTTTCG	10.97 ± 2.84	24.33 ± 1.75	11.72	[14]
CGAAAAATG**G**GATCTAGATC**C**CATTTTTCG	17.77 ± 5.00	13.56 ± 2.00	11.72	[14]
CGAAAAATGT**A**ATCTAGAT**T**ACATTTTTCG	21.76 ± 3.66	21.76 ± 1.58	21.76*	[14]
CGAAAAATGT**C**ATCTAGAT**G**ACATTTTTCG	21.76 ± 2.78	21.76 ± 5.35	21.76*	[14]
CGAAAAATGT**T**ATCTAGAT**A**ACATTTTTCG	23.05 ± 3.00	24.75 ± 4.35	19.66	[14]
CGAAAAATGTG**C**TCTAGA**G**CACATTTTTCG	-7.11 ± 3.00	5.94 ± 2.26	9.62	[14]
CGAAAAATGTG**G**TCTAGA**C**CACATTTTTCG	3.17 ± 2.60	2.67 ± 2.97	14.23	[14]
CGAAAAATGTG**T**TCTAGA**A**CACATTTTTCG	18.03 ± 2.62	-0.45 ± 2.23	12.97	[14]
PU.1 ETS				
ATAAGGGGAAGTGAA				
ATAA**AA**GGAAGTGAA	2.91 ± 1.47	-0.55 ± 1.05	0.71	[15]
ATA**CAA**GGAAGTGAA	8.56 ± 2.93	0.85 ± 2.18	5.19	[15]
ATA**GAA**GGAAGTGAA	4.76 ± 1.94	2.62 ± 1.82	6.11	[15]
ATA**TAA**GGAAGTGAA	9.65 ± 2.41	3.50 ± 1.80	1.09	[15]
ATAA**CA**GGAAGTGAA	-4.07 ± 2.62	-0.90 ± 1.41	-0.40	[15]
ATAAG**A**GGAAGTGAA	4.67 ± 0.91	5.94 ± 0.74	-2.85	[15]
ATAA**AC**GGAAGTGAA	-0.88 ± 1.71	2.34 ± 1.54	4.27	[15]
ATAA**A**GGGAAGTGAA	1.10 ± 0.79	-3.83 ± 0.62	6.02	[15]
ATAA**AT**GGAAGTGAA	14.25 ± 1.98	15.04 ± 1.38	10.46	[15]
ATAA**AA**GGAA**A**TGAA	10.70 ± 2.94	-6.33 ± 1.00	9.00	[15]
ATAA**AA**GGAAG**A**GAA	9.18 ± 5.58	-3.27 ± 2.32	6.74	[15]
ATAA**AA**GGAA**T**TGAA	12.52 ± 3.65	-4.70 ± 1.86	10.54	[15]
ATAA**AA**GGAAG**G**GAA	14.57 ± 2.18	-2.32 ± 1.46	9.37	[15]
ATAA**CC**GGAAGTGAA	-1.91 ± 2.76	3.72 ± 1.82	-1.76	[15]
ATAA**C**GGGAAGTGAA	2.96 ± 1.74	-1.10 ± 0.92	7.36	[15]
ATAA**CT**GGAAGTGAA	6.61 ± 1.78	7.37 ± 3.13	7.70	[15]
ATAAG**C**GGAAGTGAA	-2.30 ± 1.67	1.27 ± 1.22	-3.31	[15]
ATAAG**T**GGAAGTGAA	11.96 ± 1.31	12.62 ± 1.24	5.36	[15]
ATA**T**G**A**GGAAGTGAA	14.70 ± 4.50	10.42 ± 1.50	1.13	[15]
ATA**TAC**GGAAGTGAA	0.14 ± 2.50	11.67 ± 2.74	2.51	[15]
ATAA**AA**GGAA**AA**GAA	5.74 ± 2.73	0.73 ± 2.55	9.08	[15]
ATAA**AA**GGAA**TG**GAA	22.67 ± 3.53	-2.99 ± 2.79	11.84	[15]
ATA**T**G**C**GGAAGTGAA	2.84 ± 3.99	7.87 ± 2.03	1.30	[15]
ATAAG**A**GGAAG**G**GAA	11.96 ± 3.42	7.15 ± 1.51	8.16	[15]
ATAA**AC**GGAA**A**TGAA	4.29 ± 2.79	2.69 ± 2.01	9.33	[15]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
Ndt80				
GTGTCACAAATTA				
TTGTCACAAATTA	-0.01 ± 1.50	3.77 ± 1.49	1.68	[16]
CTGTCACAAATTA	2.35 ± 1.86	-3.78 ± 1.85	1.28	[16]
ATGTCACAAATTA	0.99 ± 0.73	1.61 ± 0.47	0.83	[16]
G**A**GTCACAAATTA	-7.46 ± 1.49	0.34 ± 1.21	0.45	[16]
G**G**GTCACAAATTA	-6.13 ± 1.50	5.14 ± 1.05	2.22	[16]
G**C**GTCACAAATTA	-7.16 ± 0.87	-1.22 ± 0.54	1.12	[16]
GT**A**TCACAAATTA	8.76 ± 0.94	19.55 ± 0.60	-0.25	[16]
GT**C**TCACAAATTA	2.39 ± 3.47	28.04 ± 1.62	3.09	[16]
GT**T**TCACAAATTA	11.99 ± 2.38	25.15 ± 1.90	1.41	[16]
GTG**C**CACAAATTA	0.15 ± 0.78	0.39 ± 0.57	-0.87	[16]
GTG**A**CACAAATTA	-2.47 ± 1.86	2.59 ± 1.22	-1.68	[16]
GTG**G**CACAAATTA	-7.22 ± 1.40	-0.48 ± 1.10	0.00	[16]
GTGT**T**ACAAATTA	10.12 ± 1.12	31.21 ± 0.77	3.46	[16]
GTGT**A**ACAAATTA	3.77 ± 2.83	26.22 ± 0.88	6.12	[16]
GTGTC**G**CAAATTA	1.08 ± 1.02	5.62 ± 0.61	3.00	[16]
GTGTC**C**CAAATTA	-5.22 ± 1.46	-1.85 ± 0.95	4.30	[16]
GTGTCA**T**AAATTA	8.17 ± 1.14	18.20 ± 0.65	9.48	[16]
GTGTCA**A**AAATTA	47.41 ± 2.49	28.00 ± 1.20	9.48	[16]
GTGTCAC**T**AATTA	-1.99 ± 1.53	-2.51 ± 1.05	5.14	[16]
GTGTCAC**G**AATTA	5.64 ± 0.98	4.25 ± 0.63	4.95	[16]
GTGTCACA**T**ATTA	3.76 ± 1.59	0.12 ± 1.04	4.16	[16]
GTGTCACA**G**ATTA	4.36 ± 0.68	-5.80 ± 0.53	4.77	[16]
GTGTCACAA**T**TTA	8.30 ± 1.12	0.45 ± 1.12	2.84	[16]
GTGTCACAA**G**TTA	6.04 ± 0.88	9.36 ± 0.67	2.10	[16]
GTGTCACAAA**G**TA	-5.66 ± 1.18	3.73 ± 1.24	-1.23	[16]
GTGTCACAAAT**G**A	3.02 ± 1.15	-0.61 ± 0.96	0.00	[16]
c-Myb				
CCTAACTGACA				
CC**G**AACTGACA	-8.46 ± 1.80	3.41 ± 2.42	3.75	[17,18]
CC**A**AACTGACA	-3.53 ± 1.52	-1.15 ± 3.19	2.42	[17]
CC**C**AACTGACA	-3.82 ± 1.12	-1.42 ± 1.43	0.35	[17]
CCT**G**ACTGACA	6.38 ± 1.06	-0.77 ± 0.75	9.23	[18–20]
CCT**T**ACTGACA	-13.30 ± 3.77	-0.38 ± 1.31	11.30	[19]
CCT**C**ACTGACA	-5.21 ± 1.61	2.49 ± 1.47	12.13	[19]
CCTA**G**CTGACA	2.77 ± 1.24	2.06 ± 0.76	7.26	[18–20]

Nucleotide sequence	$\Delta\Delta G$, kJ/mol Amber99sb*ILDN-BSC1	$\Delta\Delta G$, kJ/mol Charmm36m	$\Delta\Delta G$, kJ/mol Experimental	Reference
CCTA**T**CTGACA	1.89 ± 2.55	12.20 ± 2.91	6.69	[18,19]
CCTA**C**CTGACA	-1.54 ± 1.56	-8.38 ± 1.04	4.81	[18,19]
CCTAA**G**TGACA	6.39 ± 1.46	6.31 ± 1.80	12.97	[19]
CCTAA**A**TGACA	0.76 ± 1.90	1.92 ± 2.15	12.55	[19]
CCTAA**T**TGACA	7.81 ± 1.55	4.01 ± 0.99	10.20	[18–20]
CCTAAC**G**GACA	9.03 ± 2.08	7.10 ± 4.46	1.26	[19]
CCTAAC**A**GACA	14.37 ± 1.84	12.59 ± 2.29	4.60	[19]
CCTAAC**C**GACA	-1.71 ± 1.39	6.65 ± 0.93	0.00	[19]
CCTAACT**A**ACA	20.42 ± 1.36	13.88 ± 1.19	7.52	[18–20]
CCTAACT**T**ACA	19.48 ± 1.76	17.52 ± 2.29	7.95	[18,19]
CCTAACT**C**ACA	26.92 ± 1.44	31.69 ± 2.58	13.39	[19]
CCTAACTG**G**CA	-9.88 ± 1.49	-11.26 ± 0.90	2.51	[19]
CCTAACTG**T**CA	-3.58 ± 1.68	-16.52 ± 1.78	-0.84	[19]
CCTAACTG**C**CA	3.28 ± 2.20	-8.11 ± 1.97	-1.67	[19]
CCTAACTGA**G**A	7.73 ± 1.32	1.86 ± 0.97	7.95	[18,19]
CCTAACTGA**A**A	2.99 ± 1.61	5.75 ± 1.32	2.51	[19]
CCTAACTGA**T**A	2.18 ± 1.01	-0.00 ± 0.50	3.14	[18,19]
CCTAACTGAC**G**	-1.38 ± 0.77	-2.44 ± 0.44	1.42	[21]
CCTAACTGAC**T**	-1.40 ± 2.08	0.53 ± 2.59	0.17	[21]
CCTAACTGAC**C**	-3.77 ± 1.89	1.17 ± 1.51	1.84	[21]

* A lower bound for the $\Delta\Delta$G value is provided by an experiment. If a calculated value exceeds the experimental lower bound, the calculated $\Delta\Delta$G is capped at the experimental estimate.

Experimental values with several references were combined by taking an average $\Delta\Delta$G of the values reported in the literature sources.

References

(1) Hamilton, T. B.; Borel, F.; Romaniuk, P. J. Comparison of the DNA binding characteristics of the related zinc finger proteins WT1 and EGR1. *Biochemistry* **1998**, *37*, 2051–2058.

(2) Miller, J. C.; Pabo, C. O. Rearrangement of side-chains in a Zif268 mutant highlights the complexities of zinc finger-DNA recognition. *J. Mol. Biol.* **2001**, *313*, 309–315.

(3) Sarai, A.; Takeda, Y. Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 6513–6517.

(4) Grillo, A. O.; Brown, M. P.; Royer, C. A. Probing the physical basis for trp repressor-operator recognition. *J. Mol. Biol.* **1999**, *287*, 539–554.

(5) Hao, D.; Ohme-Takagi, M.; Sarai, A. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *J. Biol. Chem.* **1998**, *273*, 26857–26861.

(6) Fujimoto, S. Y.; Ohta, M.; Usui, A.; Shinshi, H.; Ohme-Takagi, M. Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box–mediated gene expression. *The Plant Cell* **2000**, *12*, 393–404.

(7) Jin, Y.; Zhong, H.; Vershon, A. K. The yeast a1 and α2 homeodomain proteins do not contribute equally to heterodimeric DNA binding. *Mol. Cell. Biol.* **1999**, *19*, 585–593.

(8) Engler, L. E.; Sapienza, P.; Dorner, L. F.; Kucera, R.; Schildkraut, I.; Jen-Jacobson, L. The energetics of the interaction of BamHI endonuclease with its recognition site GGATCC. *J. Mol. Biol.* **2001**, *307*, 619–636.

(9) Coskun-Ari, F. F.; Hill, T. M. Sequence-specific interactions in the Tus-Ter complex

and the effect of base pair substitutions on arrest of DNA replication in Escherichia coli. *J. Biol. Chem.* **1997**, *272*, 26448–26456.

(10) Frank, D. E.; Saecker, R. M.; Bond, J. P.; Capp, M. W.; Tsodikov, O. V.; Melcher, S. E.; Levandoski, M. M.; Record, M. T. Thermodynamics of the interactions of lac repressor with variants of the symmetric lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.* **1997**, *267*, 1186–1206.

(11) Boyer, M.; Poujol, N.; Margeat, E.; Royer, C. A. Quantitative characterization of the interaction between purified human estrogen receptor α and DNA using fluorescence anisotropy. *Nucleic Acids Res.* **2000**, *28*, 2494–2502.

(12) Takeda, Y.; Sarai, A.; Rivera, V. M. Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 439–443.

(13) Lesser, D. R.; Kurpiewski, M. R.; Jen-Jacobson, L. The energetic basis of specificity in the EcoRI endonuclease-DNA interaction. *Science* **1990**, *250*, 776–786.

(14) Gunasekera, A.; Ebright, Y. W.; Ebright, R. H. DNA sequence determinants for binding of the Escherichia coli catabolite gene activator protein. *J. Biol. Chem.* **1992**, *267*, 14713–14720.

(15) Poon, G. M. K.; Macgregor Jr, R. B. Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. *J. Mol. Biol.* **2003**, *328*, 805–819.

(16) Pierce, M.; Benjamin, K. R.; Montano, S. P.; Georgiadis, M. M.; Winter, E.; Vershon, A. K. Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. *Mol. Cell. Biol.* **2003**, *23*, 4814–4825.

(17) Oda, M.; Furukawa, K.; Ogata, K.; Sarai, A.; Ishii, S.; Nishimura, Y.; Nakamura, H. Investigation of the pyrimidine preference by the c-Myb DNA-binding domain at the initial base of the consensus sequence. *J. Biol. Chem.* **1997**, *272*, 17966–17971.

(18) Oda, M.; Furukawa, K.; Ogata, K.; Sarai, A.; Nakamura, H. Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. *J. Mol. Biol.* **1998**, *276*, 571–590.

(19) Ogata, K.; Kanei-Ishii, C.; Sasaki, M.; Hatanaka, H.; Nagadoi, A.; Enari, M.; Nakamura, H.; Nishimura, Y.; Ishii, S.; Sarai, A. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Mol. Biol.* **1996**, *3*, 178–187.

(20) Oda, M.; Furukawa, K.; Sarai, A.; Nakamura, H. Kinetic analysis of DNA binding by the c-Myb DNA-binding domain using surface plasmon resonance. *FEBS Lett.* **1999**, *454*, 288–292.

(21) Tanikawa, J.; Yasukawa, T.; Enari, M.; Ogata, K.; Nishimura, Y.; Ishii, S.; Sarai, A. Recognition of specific DNA sequences by the c-myb protooncogene product: role of three repeat units in the DNA-binding domain. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 9320–9324.