Electronic Supplementary Material (ESI) for Chemical Science.

Supporting Information:

Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy

Vytautas Gapsys[a,‡], *Laura Pérez-Benito*[b,‡], *Matteo Aldeghi*[a], *Daniel Seeliger*[c], *Herman van Vlijmen*[b], *Gary Tresadern*[b,*], *Bert L. de Groot*[a,*]

[a] Computational Biomolecular Dynamics Group, Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany. [b] Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., Turnhoutseweg 30, B-2340 Beerse, Belgium. [c] Medicinal Chemistry, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Strasse 65, D-88397 Biberach a.d. Riss, Germany.

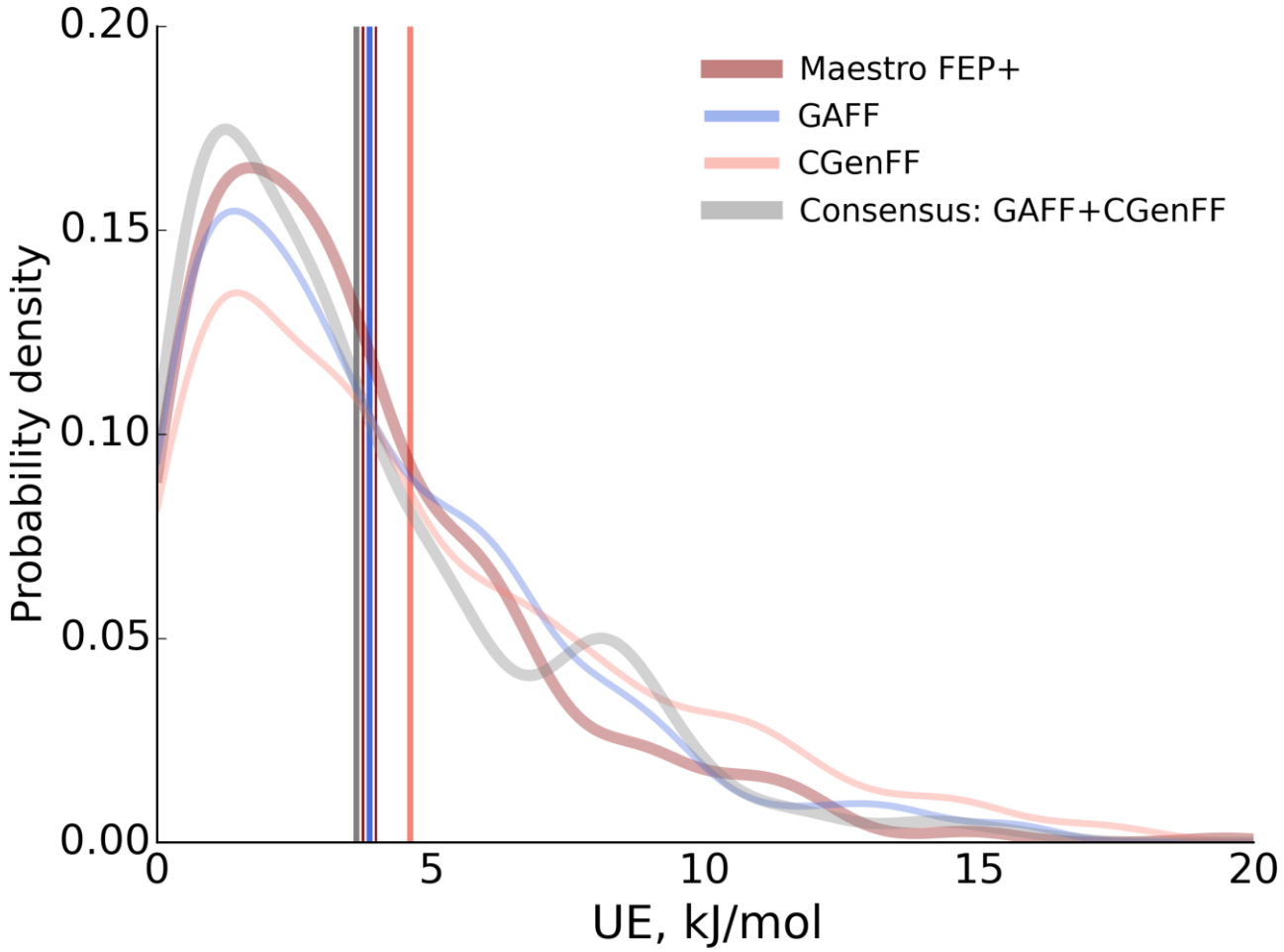

Figure S1: Distributions of the unsigned errors, i.e. unsigned differences from the experimental measurements. The vertical bars depict mean values. The whole dataset of 482 ligand modifications was considered.

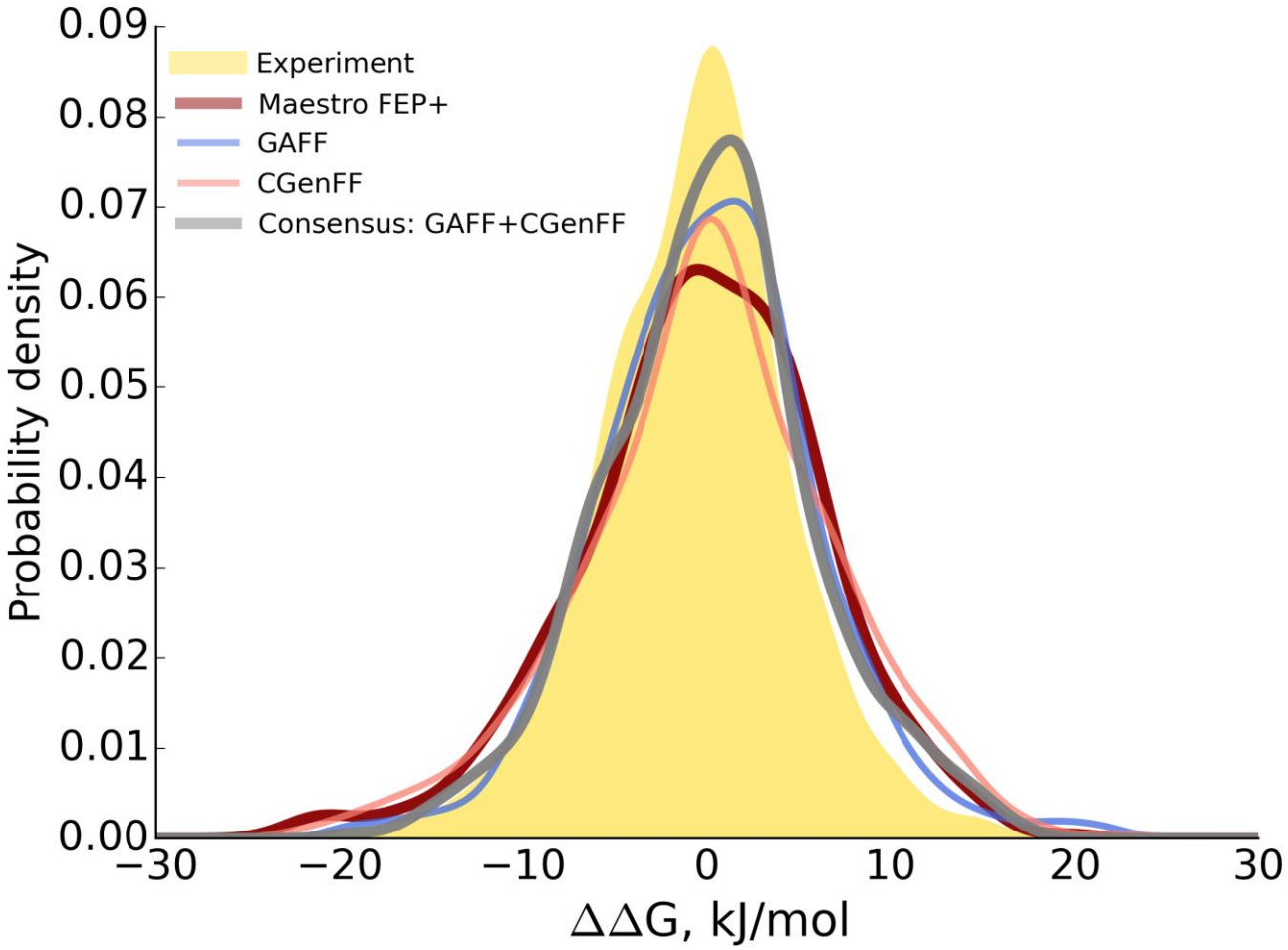

Figure S2: Distributions of the double free energy differences for the whole data set of 482 ligand modifications: experimental and calculated values.

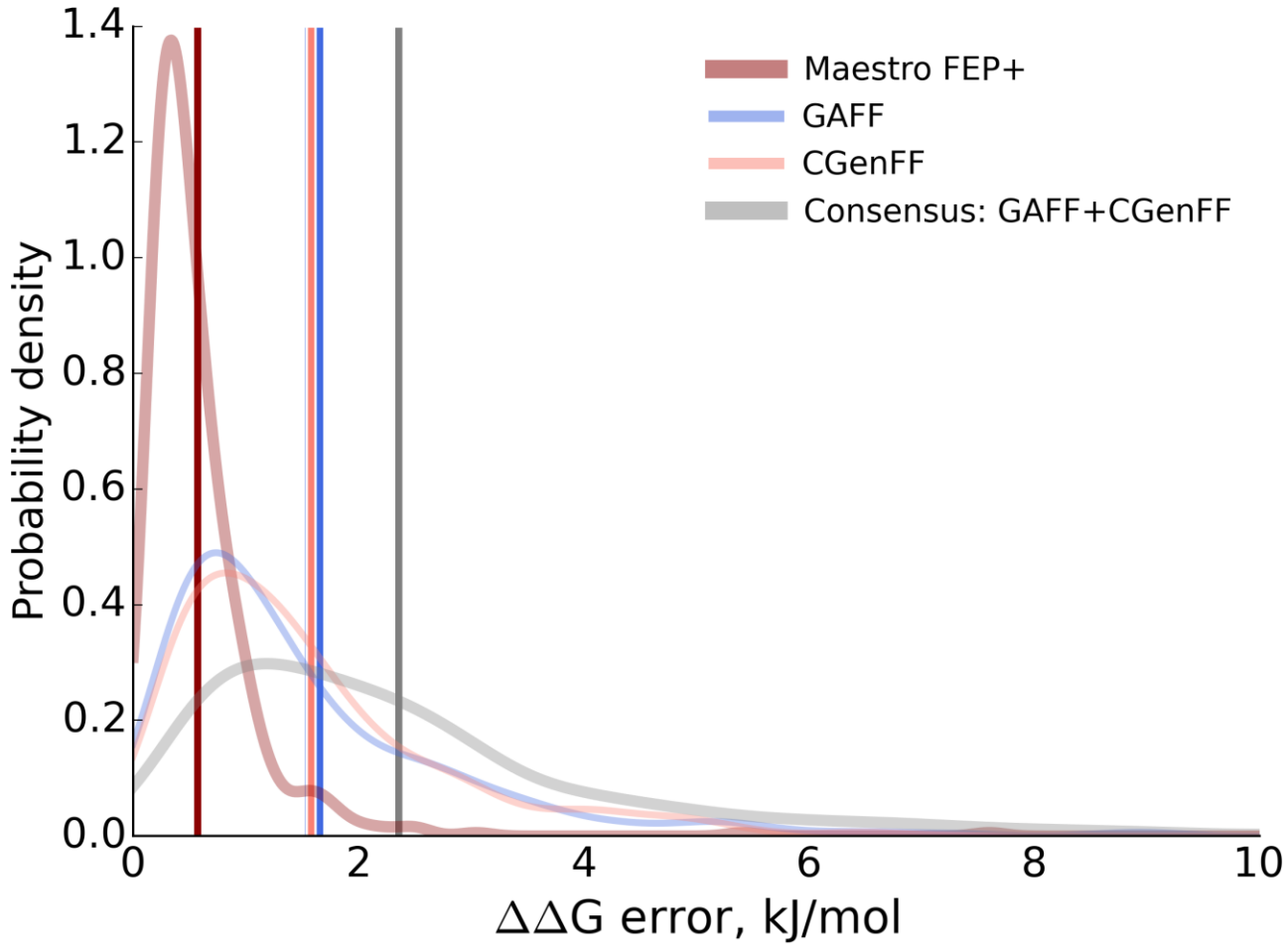

Figure S3: Distributions of the error estimates for the calculated double free energy differences. The vertical bars depict mean values. The whole dataset of 482 ligand modifications was considered.

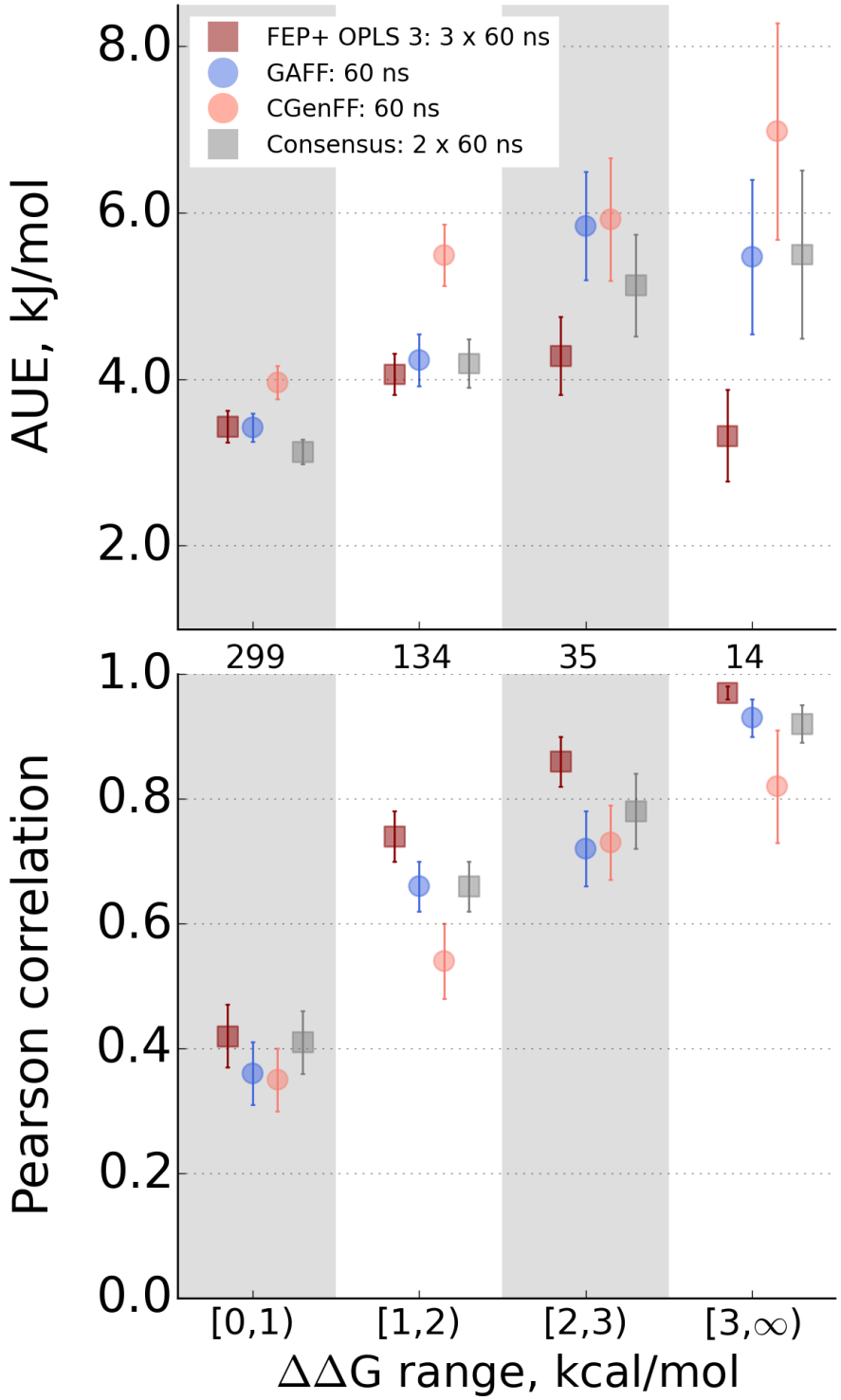

Figure S4: AUE and Pearson correlation for the ΔΔG estimates over four discrete ranges. In terms of AUE, consensus force field approach outperforms or performs on par with the FEP+ in the range of low double free energy differences, while for the changes showing larger differences FEP+ has a higher accuracy prediction. Similar trend holds for the Pearson correlation.

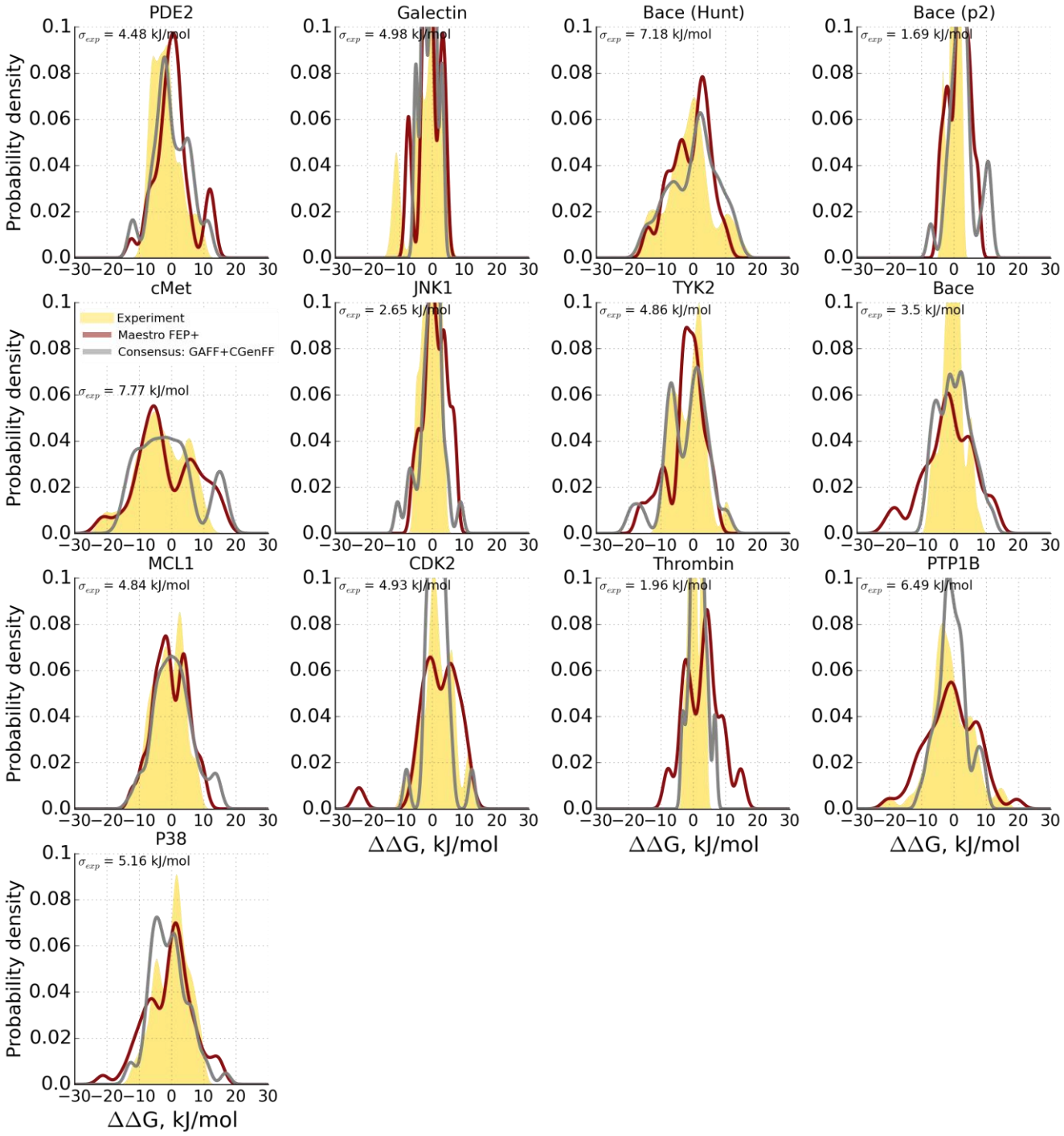

Figure S5: Distributions of the double free energy differences for every protein-ligand system: experimental and calculated values. The text in the panels reports on the standard deviation of the experimental distribution.

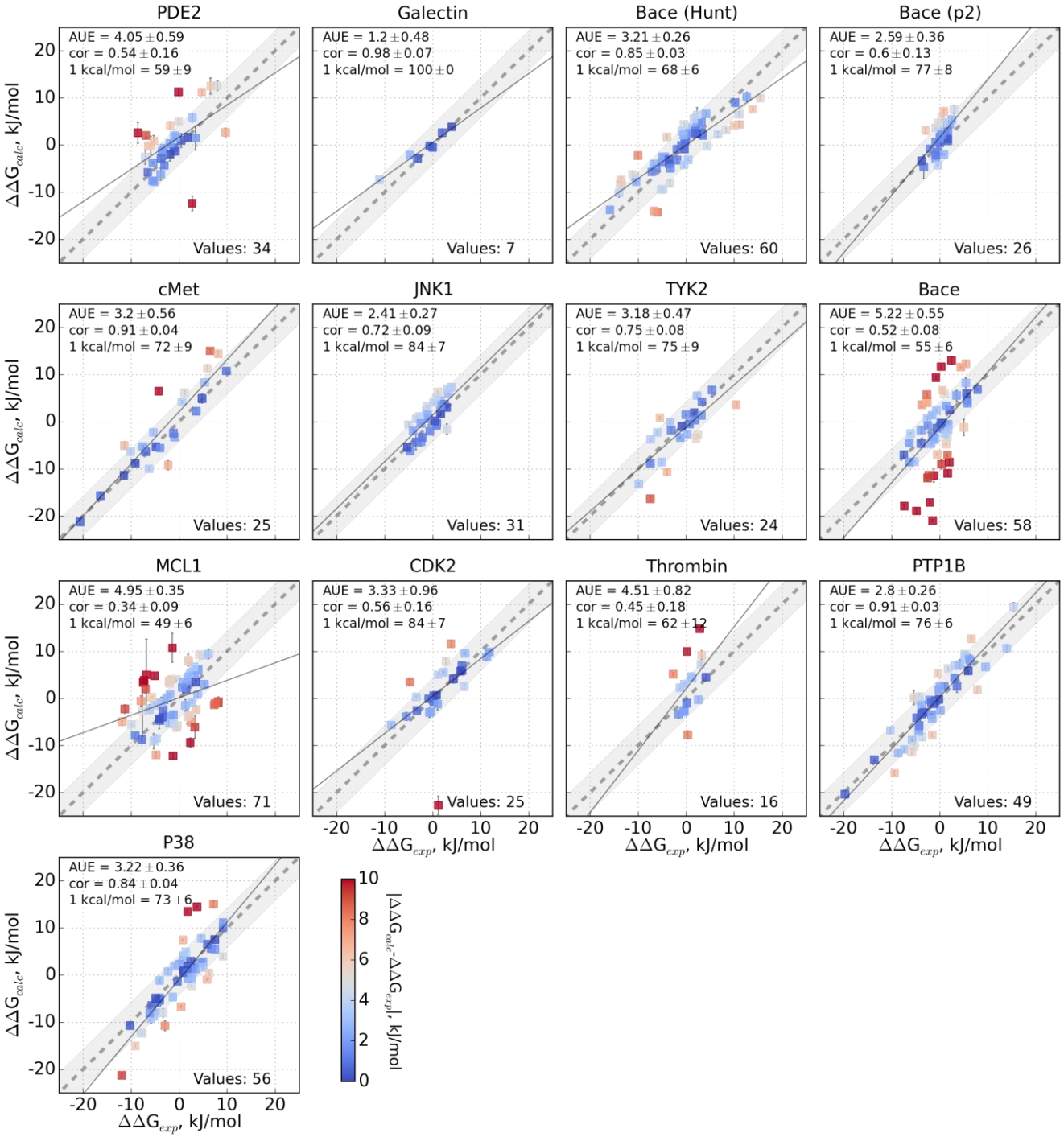

Figure S6: The FEP+ OPLS 3.1 calculations of the ΔΔG values plotted against the experimental measurements for every protein-ligand system separately. Text in the panels: average unsigned error (AUE) is in kJ/mol; cor is Pearson correlation; 1 kcal/mol denotes the percentage of the estimates that fall within 1 kcal/mol from the experimental measurement.

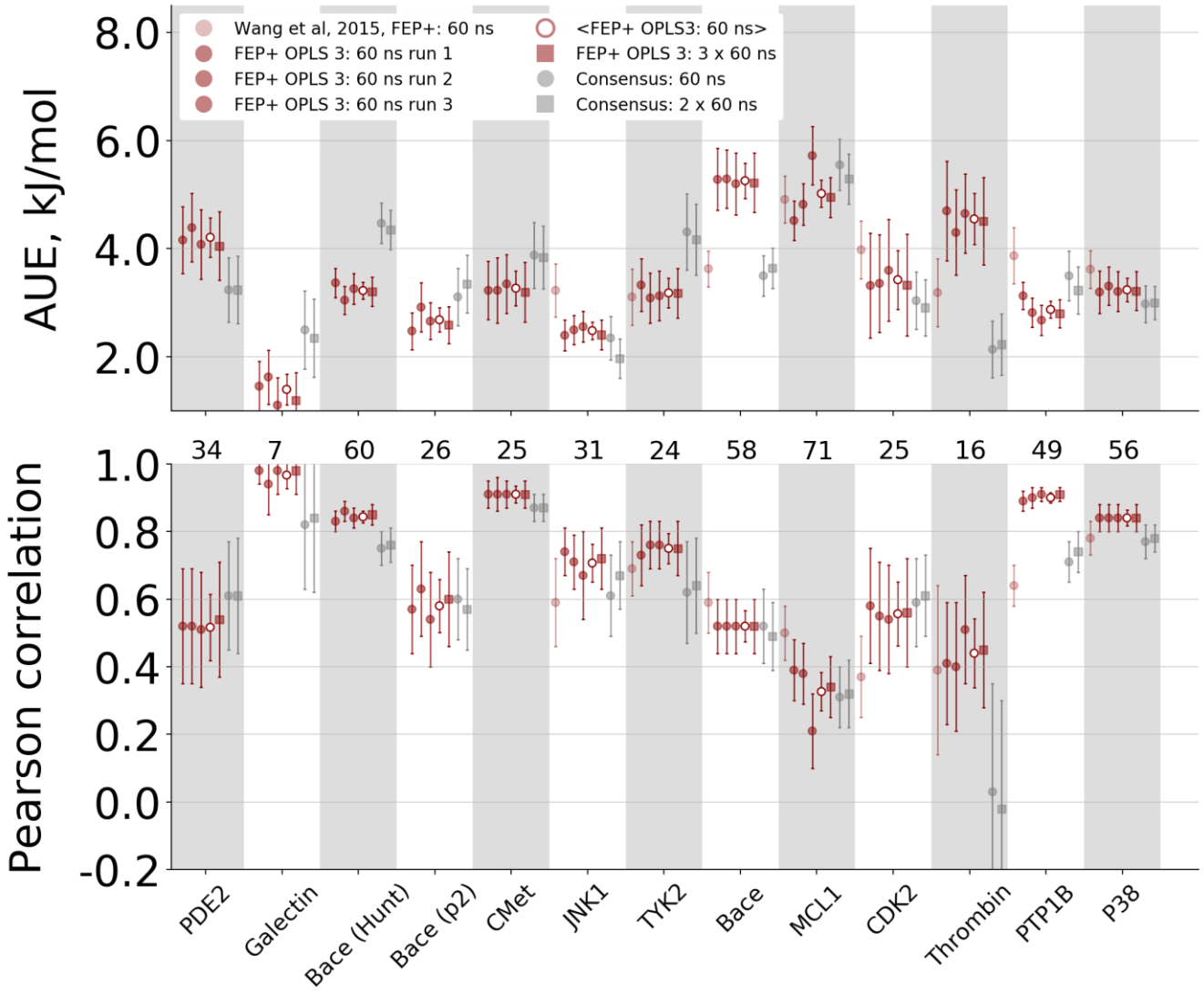

Figure S7: Average unsigned error (AUE) and Pearson correlation for the ΔΔG estimates split by protein-ligand system. The individual FEP+ runs are depicted as well as an average over the AUE and correlation of the three FEP+ replicas (white circle). An averaging over the ΔΔG values from three replicas is depicted as a red square (3 x 60 ns per ΔG estimate). The empty circle denotes average of three AUE estimates each from a 60 ns run. For the pmx based calculations two variants of the consensus force field approach are shown: one uses 60 ns per ΔG estimate, while another uses 2 x 60 ns. The numbers in between the top and bottom panels denote the number of ligand modifications considered for the corresponding system.

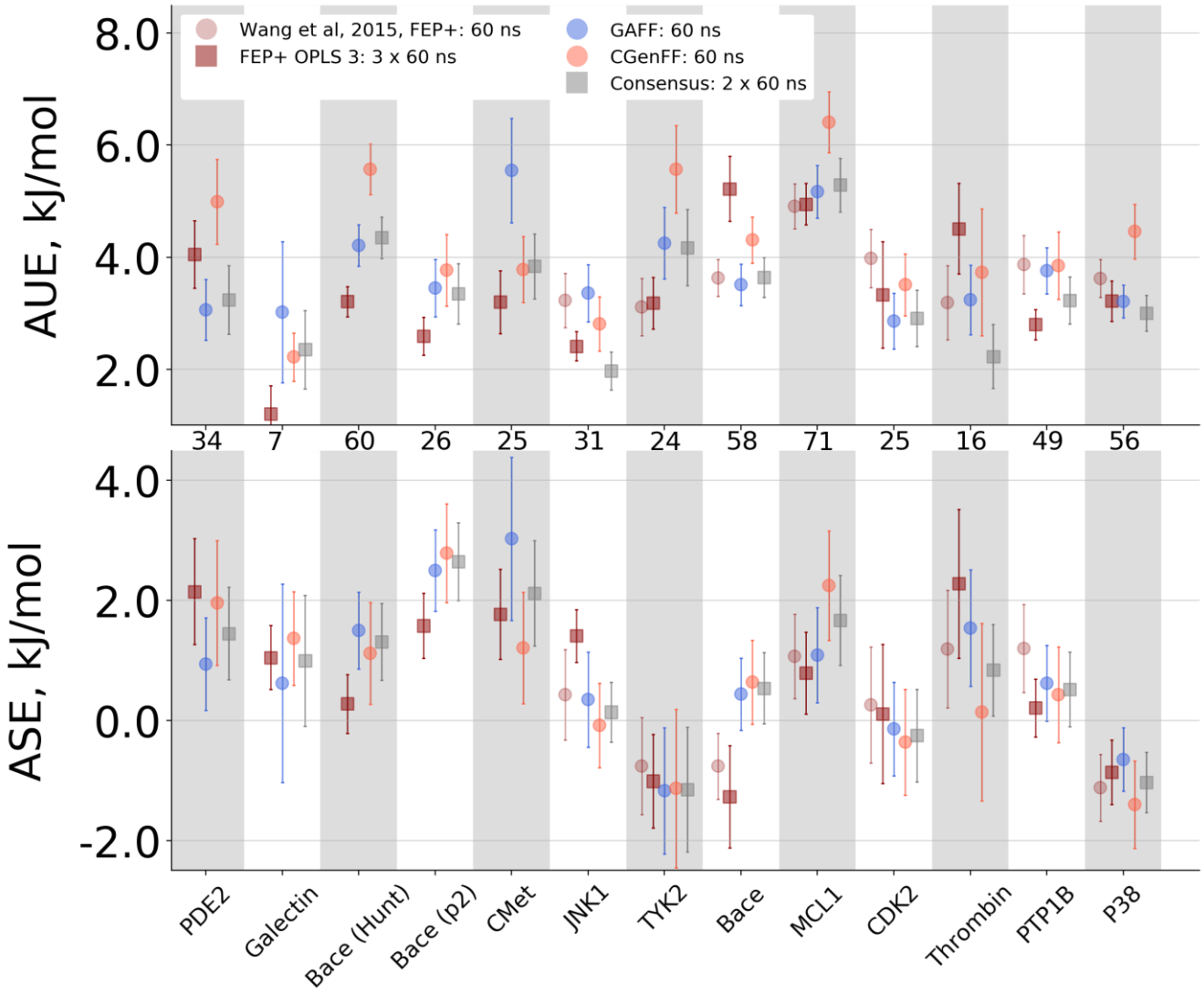

Figure S8: Average unsigned error (AUE) and average signed error (ASE) for the ΔΔG estimates split by protein-ligand system. The numbers in between the top and bottom panels denote the number of ligand modifications considered for the corresponding system.

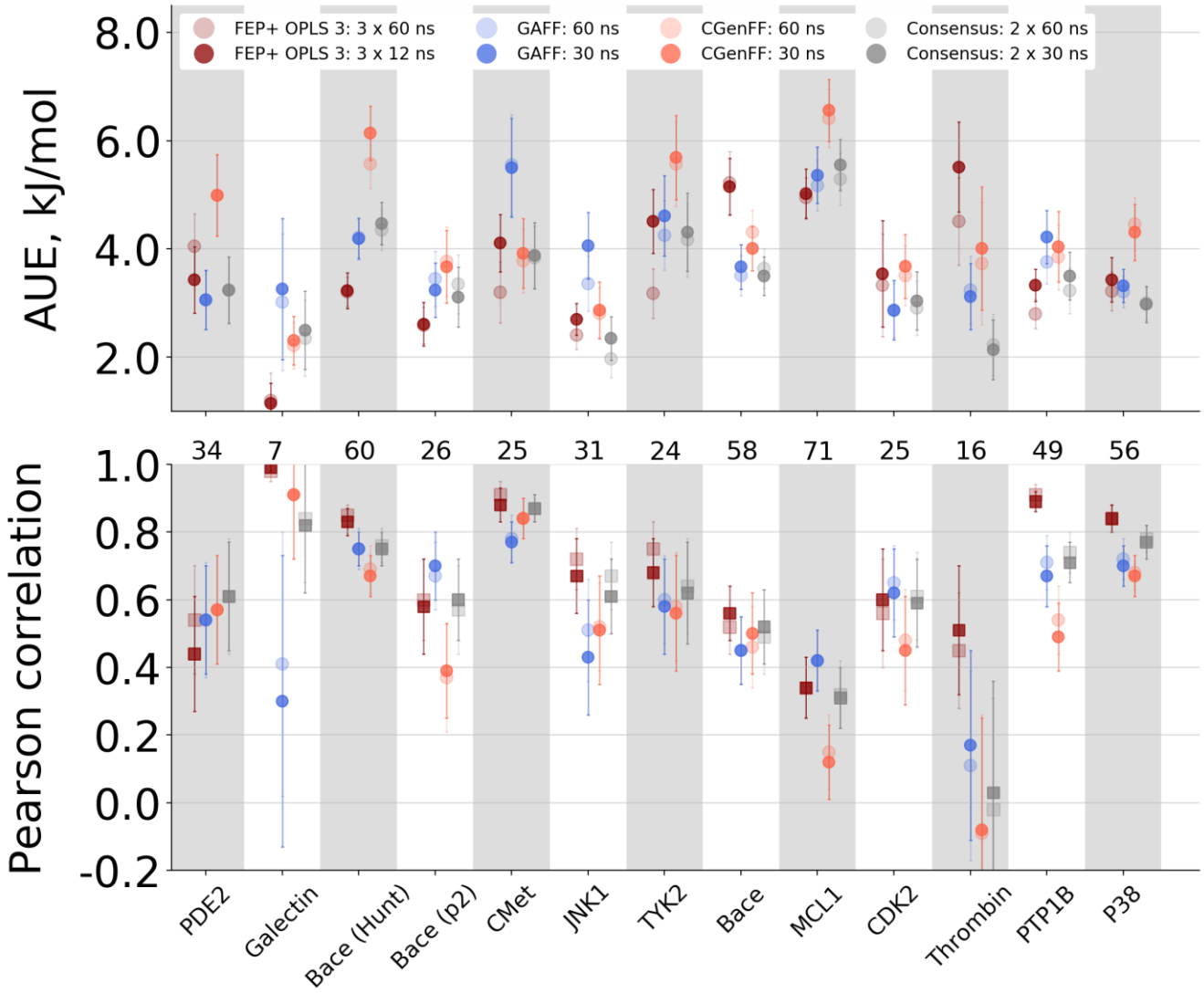

Figure S9: Average unsigned error (AUE) and Pearson correlation for the ΔΔG estimates split by protein-ligand system. The transparent symbols denote estimates obtained from the full simulation time, while opaque symbols depict estimates for which only a fraction of simulation time was used. The numbers in between the top and bottom panels denote the number of ligand modifications considered for the corresponding system.

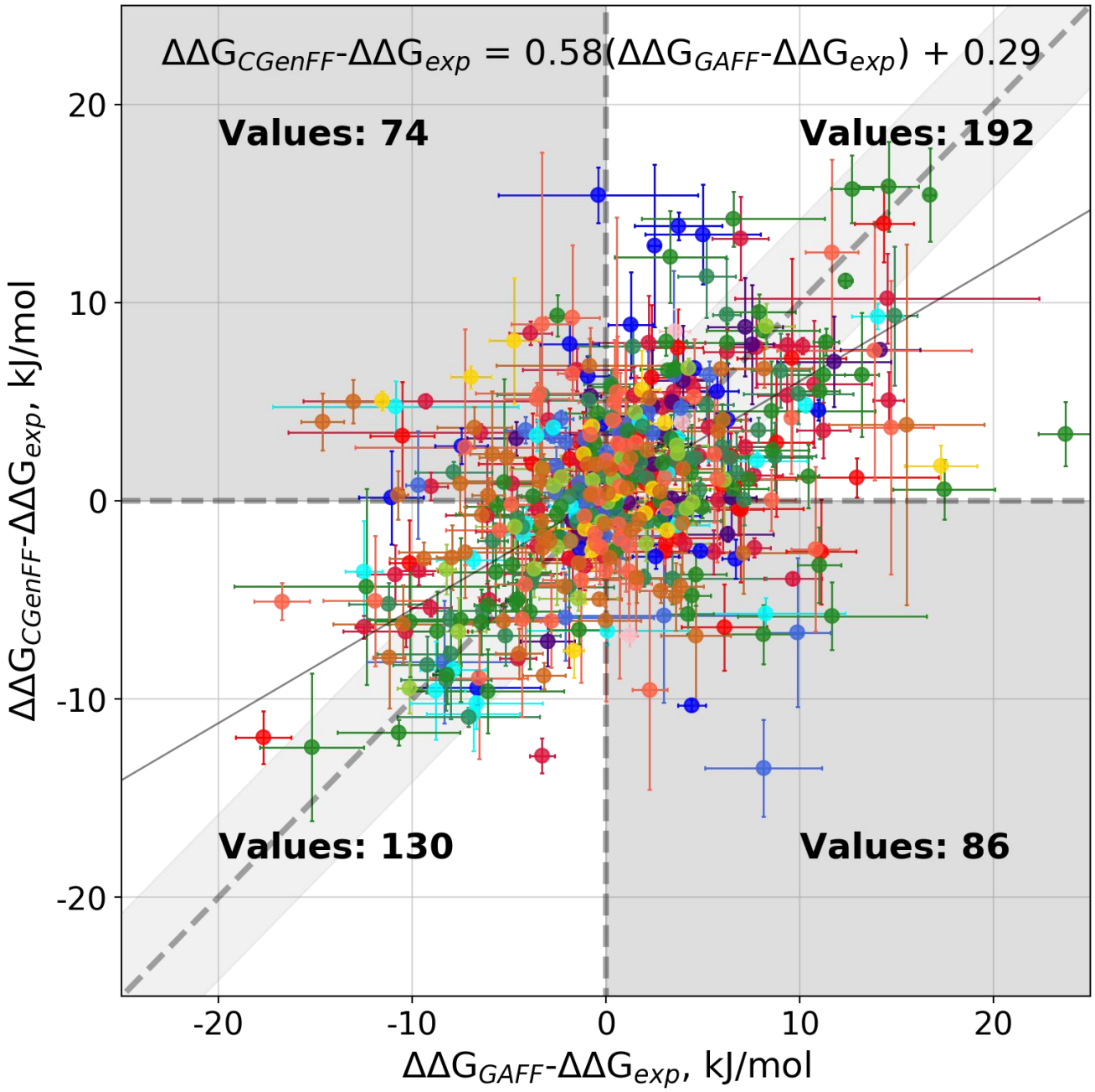

Figure S10: Signed deviations from experiment for the ΔΔG values calculated with the GAFF force field plotted against the deviations calculated with the CGenFF force field. Different colors are used for the protein-ligand datasets. The values shown in the plot list the number of points falling into disparate quadrants.

System	FEP+ (5 ns)	FEP+ (1 ns)	GAFF	CGenFF	Consensus
PDE2	4.05 ± 0.6	3.43 ± 0.62	3.06 ± 0.54	4.99 ± 0.77	3.24 ± 0.63
Galectin	1.2 ± 0.51	1.15 ± 0.37	3.02 ± 1.17	2.22 ± 0.44	2.5 ± 0.73
Bace (Hunt)	3.21 ± 0.26	3.23 ± 0.33	4.21 ± 0.38	5.57 ± 0.48	4.47 ± 0.37
Bace (p2)	2.59 ± 0.34	2.61 ± 0.4	3.45 ± 0.51	3.77 ± 0.66	3.11 ± 0.54
CMET	3.2 ± 0.55	4.11 ± 0.53	5.55 ± 0.95	3.78 ± 0.64	3.88 ± 0.6
JNK1	2.41 ± 0.27	2.7 ± 0.28	3.36 ± 0.51	2.81 ± 0.48	2.35 ± 0.41
TYK2	3.18 ± 0.47	4.51 ± 0.57	4.25 ± 0.68	5.57 ± 0.75	4.31 ± 0.74
BACE	5.22 ± 0.55	5.15 ± 0.51	3.51 ± 0.38	4.31 ± 0.41	3.5 ± 0.36
MCL1	4.95 ± 0.35	5.02 ± 0.43	5.17 ± 0.46	6.41 ± 0.57	5.55 ± 0.46
CDK2	3.33 ± 0.9	3.54 ± 1.03	2.86 ± 0.51	3.51 ± 0.56	3.04 ± 0.51
Thrombin	4.51 ± 0.82	5.51 ± 0.86	3.24 ± 0.63	3.73 ± 1.11	2.14 ± 0.57
PTP1B	2.8 ± 0.26	3.33 ± 0.31	3.76 ± 0.41	3.85 ± 0.63	3.5 ± 0.44
P38	3.22 ± 0.36	3.43 ± 0.41	3.21 ± 0.29	4.46 ± 0.5	2.98 ± 0.33

Table S1: AUE for the investigated protein-ligand systems. FEP+ (5 ns) indicates the case where 5 ns per λ window were used, while for FEP+ (1 ns) simulations of 1 ns per λ window were performed. The simulations underlying the data in the table in total used 180 ns per ΔG for FEP+ (5 ns), 36 ns for FEP+ (1 ns), 60 ns for GAFF, 60 for CGenFF and 120 ns for the Consensus results. Values are in kJ/mol.

System	FEP+ (5 ns)	FEP+ (1 ns)	GAFF	CGenFF	Consensus
PDE2	0.54 ± 0.16	0.44 ± 0.18	0.54 ± 0.16	0.57 ± 0.15	0.61 ± 0.15
Galectin	0.98 ± 0.03	0.99 ± 0.01	0.41 ± 0.38	0.91 ± 0.22	0.82 ± 0.24
Bace (Hunt)	0.85 ± 0.03	0.83 ± 0.04	0.75 ± 0.06	0.69 ± 0.06	0.75 ± 0.05
Bace (p2)	0.6 ± 0.14	0.58 ± 0.13	0.67 ± 0.1	0.37 ± 0.16	0.6 ± 0.12
CMET	0.91 ± 0.04	0.88 ± 0.05	0.78 ± 0.07	0.84 ± 0.06	0.87 ± 0.04
JNK1	0.72 ± 0.09	0.67 ± 0.1	0.51 ± 0.15	0.52 ± 0.12	0.61 ± 0.12
TYK2	0.75 ± 0.08	0.68 ± 0.11	0.6 ± 0.14	0.58 ± 0.16	0.62 ± 0.15
BACE	0.52 ± 0.08	0.56 ± 0.08	0.45 ± 0.1	0.46 ± 0.12	0.52 ± 0.11
MCL1	0.34 ± 0.09	0.34 ± 0.09	0.42 ± 0.09	0.15 ± 0.11	0.31 ± 0.09
CDK2	0.56 ± 0.16	0.6 ± 0.15	0.65 ± 0.11	0.48 ± 0.15	0.59 ± 0.13
Thrombin	0.45 ± 0.17	0.51 ± 0.19	0.11 ± 0.29	-0.09 ± 0.34	0.03 ± 0.32
PTP1B	0.91 ± 0.03	0.89 ± 0.03	0.71 ± 0.08	0.54 ± 0.1	0.71 ± 0.07
P38	0.84 ± 0.04	0.84 ± 0.04	0.72 ± 0.06	0.68 ± 0.05	0.77 ± 0.05

Table S2: Pearson's correlation for the investigated protein-ligand systems. FEP+ (5 ns) indicates the case where 5 ns per λ window were used, while for FEP+ (1 ns) simulations of 1 ns per λ window were performed. The simulations underlying the data in the table in total used 180 ns per ΔG for FEP+ (5 ns), 36 ns for FEP+ (1 ns), 60 ns for GAFF, 60 for CGenFF and 120 ns for the Consensus results.