Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information for "Solvation Entropy, Enthalpy and Free Energy Prediction using a Multi-task Deep Learning Functional in 1D-RISM"

Daniel J. Fowles, David S. Palmer*
* Department of Pure and Applied Chemistry, University of Strathclyde, Thomas Graham Building, 295 Cathedral Street, Glasgow, Scotland G1 1XL, U.K.; E-mail: david.palmer@strath.ac.uk

1 Violin Plots

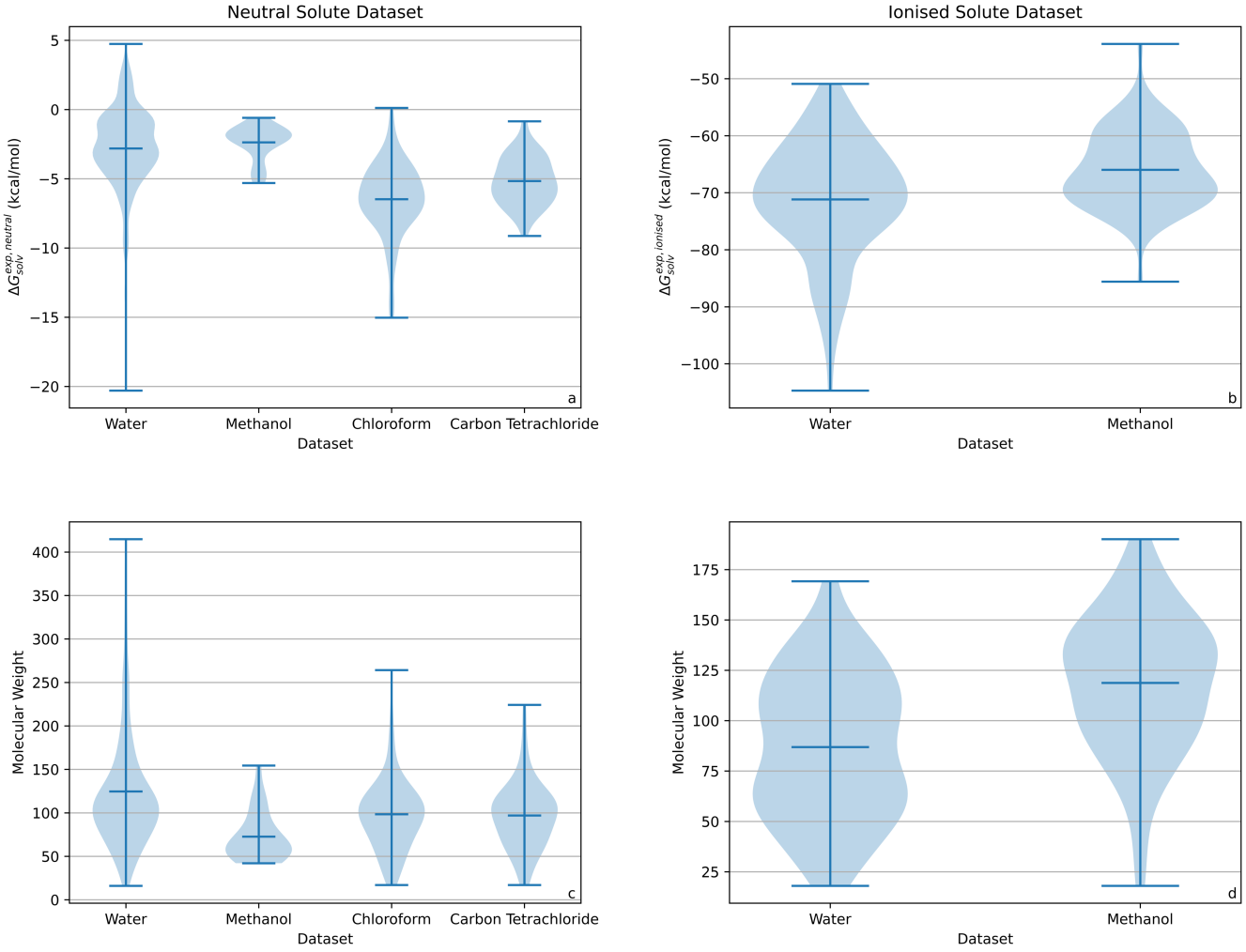

Figure 1 Violin plots showing solute molecule data by solvent for the neutral and ionised datasets. a = experimental solvation free energies for the neutral solute dataset, b = experimental solvation free energies for the ionised solute dataset, c = molecular weights for the neutral solute dataset, d = molecular weights for the ionised solute dataset.

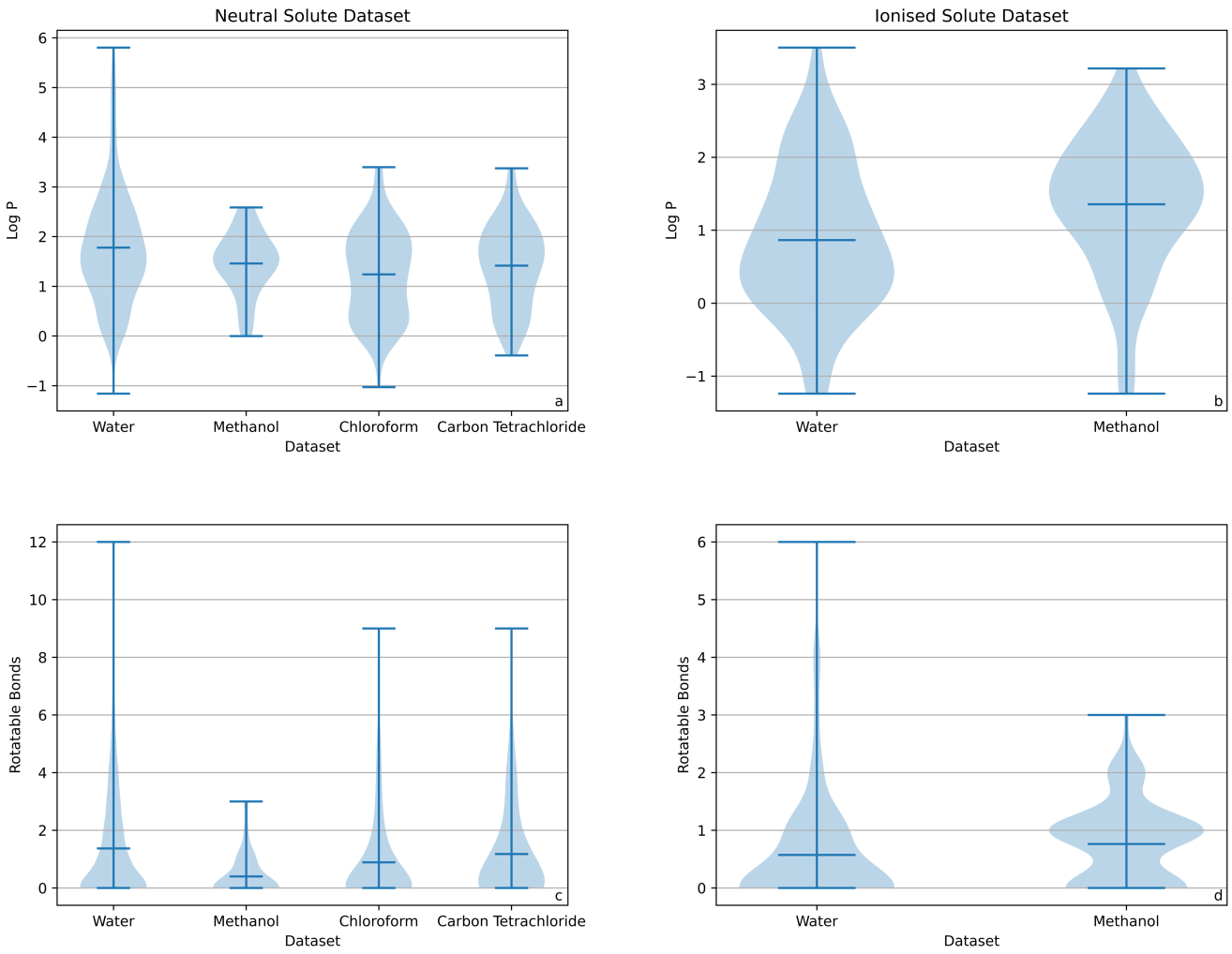

Figure 2 Violin plots showing solute molecule data by solvent for the neutral and ionised datasets. a = experimental log P for the neutral solute dataset, b = experimental log P for the ionised solute dataset, c = number of rotatable bonds per solute for the neutral solute dataset, d = number of rotatable bonds per solute for the ionised solute dataset.

Neutral Dataset			SFE		MW		Log P		Rot. Bonds	
Solvent	Temperature	Datapoints	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Carbon Tetrachloride	298K	79	-3.00	3.09	118.33	65.54	1.68	1.12	1.28	1.79
Chloroform	298K	109								
Methanol	298K	25								
Water	273-373K	3440								
Ionised Dataset			SFE		MW		Log P		Rot. Bonds	
Solvent	Temperature	Datapoints	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Methanol	298K	77	-68.98	9.24	100.53	38.64	1.08	1.00	0.65	0.93
Water	298K	103								

Table 1 Mean and standard deviation (SD) of each experimental property across the neutral and ionised datasets. Values associated with SFE and molecular weight (MW) are given in kcal/mol and g/mol, respectively.

Thermodynamics Dataset			Solvation Free Energy		Solvation Enthalpy		Solvation Entropy	
Solvent	Temperature	Datapoints	Mean	SD	Mean	SD	Mean	SD
Water	298K	139	-2.46	2.48	-10.17	3.70	-7.72	2.32

Table 2 Mean and standard deviation (SD) of each solvation thermodynamic property across neutral solutes with available experimental solvation enthalpy, entropy and free energy data. Values are given in kcal/mol.

2 Solvation Free Energy - Neutral & Ionised Datasets

KH: $\Delta G_{solv}^{exp,neutral}$ Dataset										
Neutral Solvation Free Energy			By Solvent				Full Dataset			
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
Carbon Tetrachloride	298K	79	0.75	0.83	0.48	0.63	0.90	1.14	0.22	1.11
Chloroform	298K	109	0.84	0.98	0.50	0.80				
Methanol	298K	25	0.21	0.96	0.24	0.84				
Water	273-373K	3440	0.91	1.21	0.10	1.19				
KH: $\Delta G_{solv}^{exp,ionised}$ Dataset										
Ionised Solvation Free Energy			By Solvent				Full Dataset			
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
Methanol	298K	77	0.42	3.40	0.62	2.81	0.70	3.66	0.84	3.26
Water	298K	103	0.73	3.81	1.05	3.33				
HNC: $\Delta G_{solv}^{exp,neutral}$ Dataset										
Neutral Solvation Free Energy			By Solvent				Full Dataset			
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
Carbon Tetrachloride	298K	79	0.82	0.71	0.35	0.58	0.93	0.96	0.29	0.91
Chloroform	298K	109	0.86	0.95	0.46	0.80				
Methanol	298K	25	0.52	0.62	0.17	0.55				
Water	273-373K	3440	0.94	1.01	0.24	0.97				
HNC: $\Delta G_{solv}^{exp,ionised}$ Dataset										
Ionised Solvation Free Energy			By Solvent				Full Dataset			
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
Methanol	298K	77	0.69	3.21	0.39	2.84	0.77	3.77	0.24	3.58
Water	298K	103	0.77	4.06	0.15	3.82				

Table 3 Breakdown of solvation free energy predictions for neutral and ionised solutes. Separate CNN were trained on either neutral or ionised solute SFED, which were generated using the KH or HNC free energy functionals. Statistics are given across each dataset, with model performance separated by solvent as well as across the full datasets of neutral or ionised solutes. Errors are given in kcal/mol.

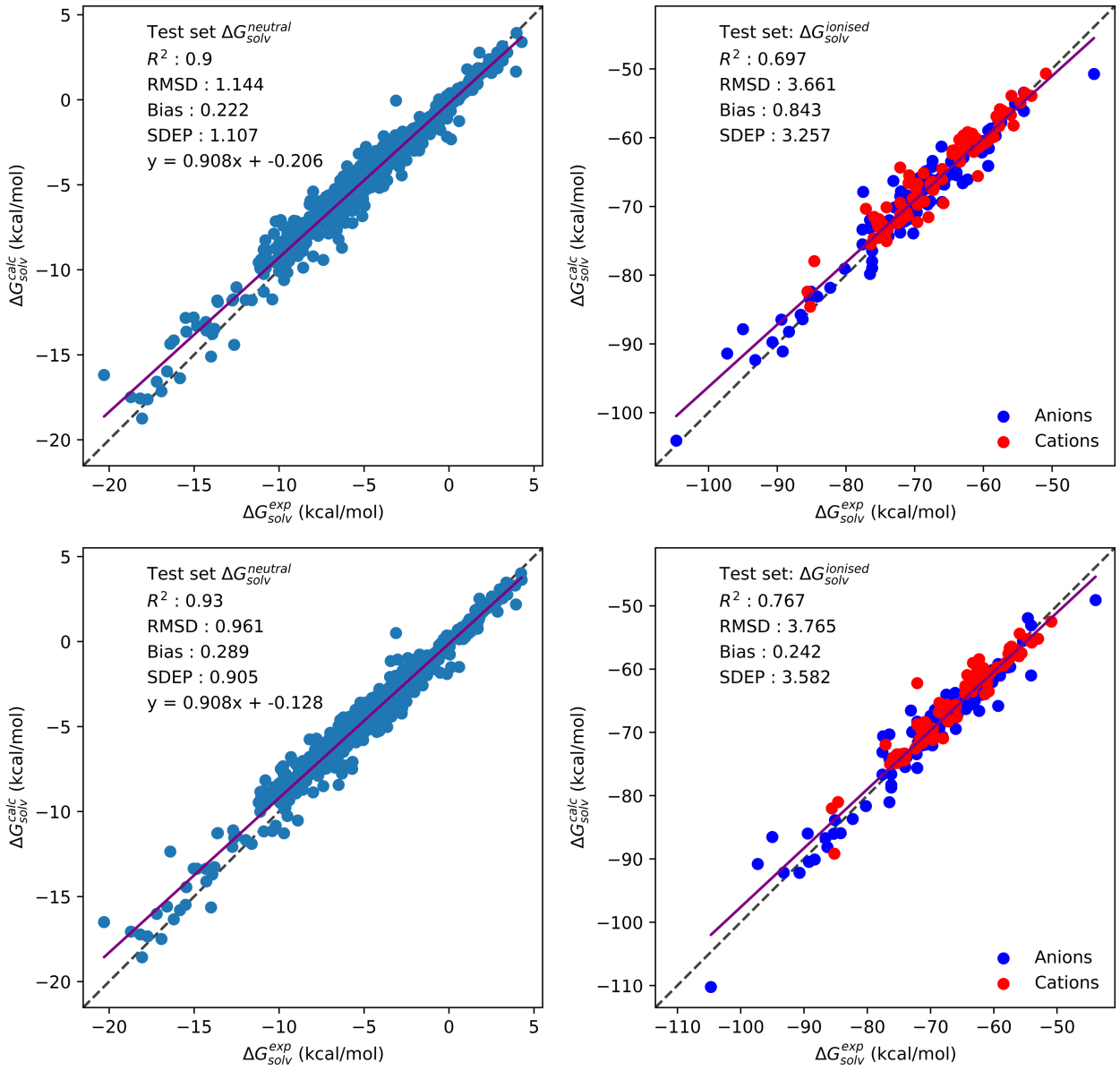

Figure 3 Correlation plots showing solvation free energy predictions from CNN trained on the individual neutral and ionised datasets. SFED datasets were generated using the KH and HNC functionals. By row, plots show solvation free energy predictions made using KH or HNC SFED, respectively. The ionised dataset correlation plots are further separated into cation and anion based predictions.

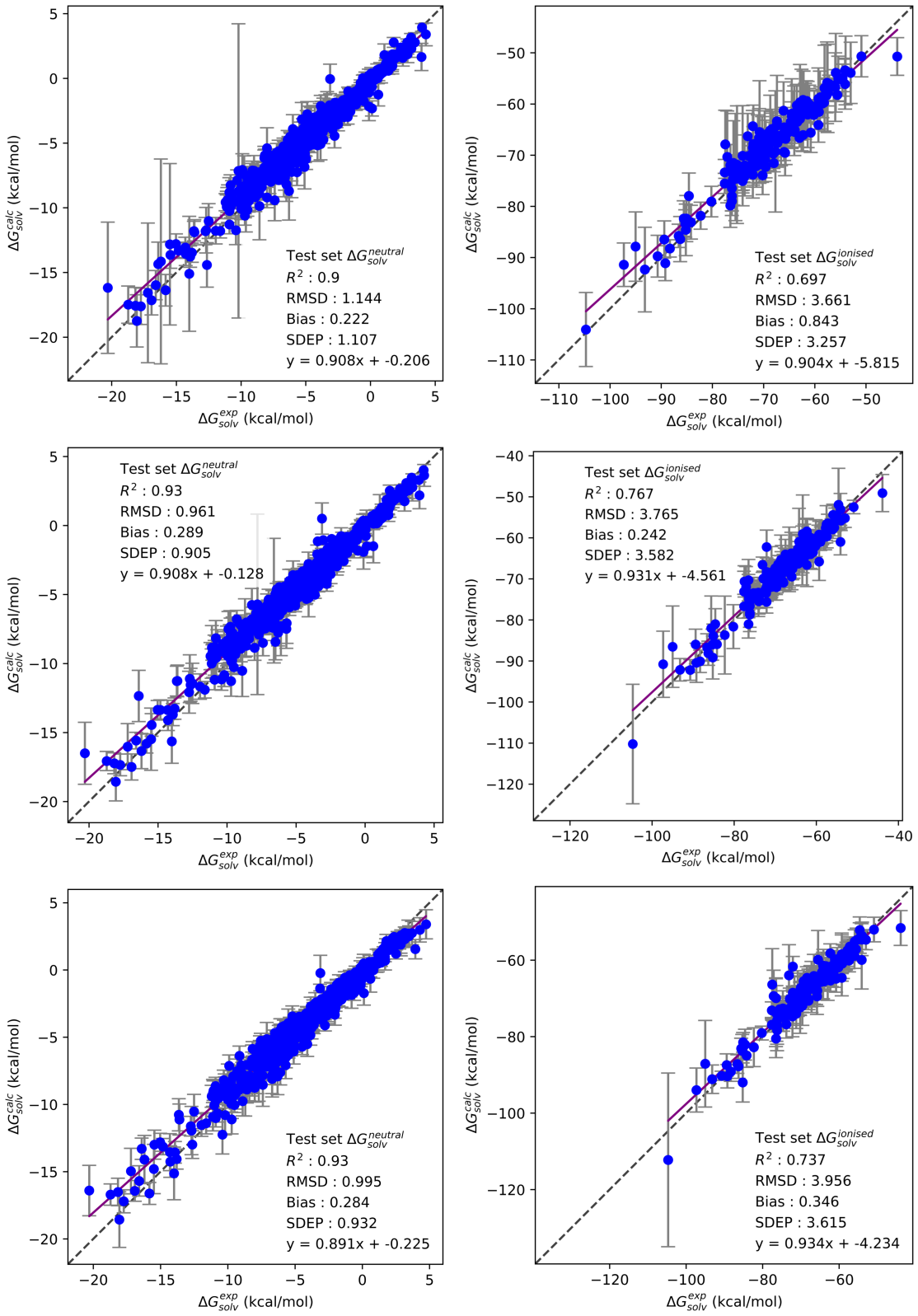

Figure 4 Correlation plots showing solvation free energy predictions from CNN trained on the individual neutral and ionised datasets. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation free energy predictions made using KH, HNC or GF SFED, respectively.

3 Solvation Free Energy - Combined Dataset

KH: $\Delta G_{solv}^{exp,neutral}$, $\Delta G_{solv}^{exp,ionised}$ Dataset										
Neutral Solvation Free Energy			By Solvent				Full Dataset			
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
Carbon Tetrachloride	298K	79	0.34	1.30	-0.03	1.08	0.98	3.43	-0.10	3.38
Chloroform	298K	109	0.45	1.78	-0.27	1.56				
Methanol	298K	25	-6.99	1.89	-0.70	1.53				
Water	273-373K	3440	0.74	2.12	-0.68	1.97				
Ionised Solvation Free Energy			By Solvent							
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP				
Methanol	298K	77	0.29	5.01	1.69	4.34				
Water	298K	103	0.50	6.75	0.95	6.41				
HNC: $\Delta G_{solv}^{exp,neutral}$, $\Delta G_{solv}^{exp,ionised}$ Dataset										
Neutral Solvation Free Energy			By Solvent				Full Dataset			
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
Carbon Tetrachloride	298K	79	0.15	1.40	-0.09	1.08	0.98	3.46	-0.22	3.40
Chloroform	298K	109	0.42	1.77	-0.33	1.48				
Methanol	298K	25	-3.01	1.62	-0.80	1.23				
Water	273-373K	3440	0.72	2.19	-0.57	2.04				
Ionised Solvation Free Energy			By Solvent							
Solvent	Temperature	Datapoints	R^2	RMSD	Bias	SDEP				
Methanol	298K	77	0.32	5.30	1.14	4.90				
Water	298K	103	0.51	6.53	0.27	6.22				

Table 4 Breakdown of solvation free energy predictions made using CNN trained on a combined neutral and ionised dataset. SFED were generated using the KH and HNC free energy functionals. Statistics are given across the full dataset, as well as separated by solvent and neutral/ionised data. Errors are given in kcal/mol.

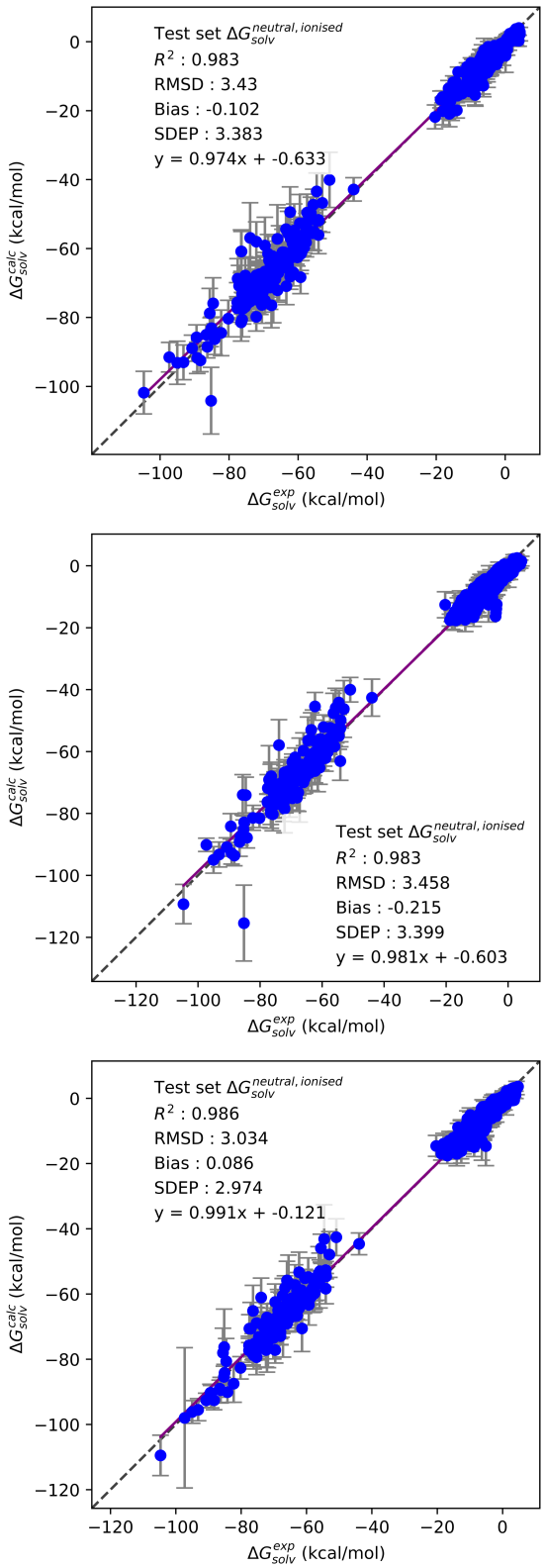

Figure 5 Correlation plots showing solvation free energy predictions from CNN trained on the combined neutral and ionised dataset. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation free energy predictions made using KH, HNC or GF SFED, respectively.

4 Solvation Thermodynamics - Multi-task CNN

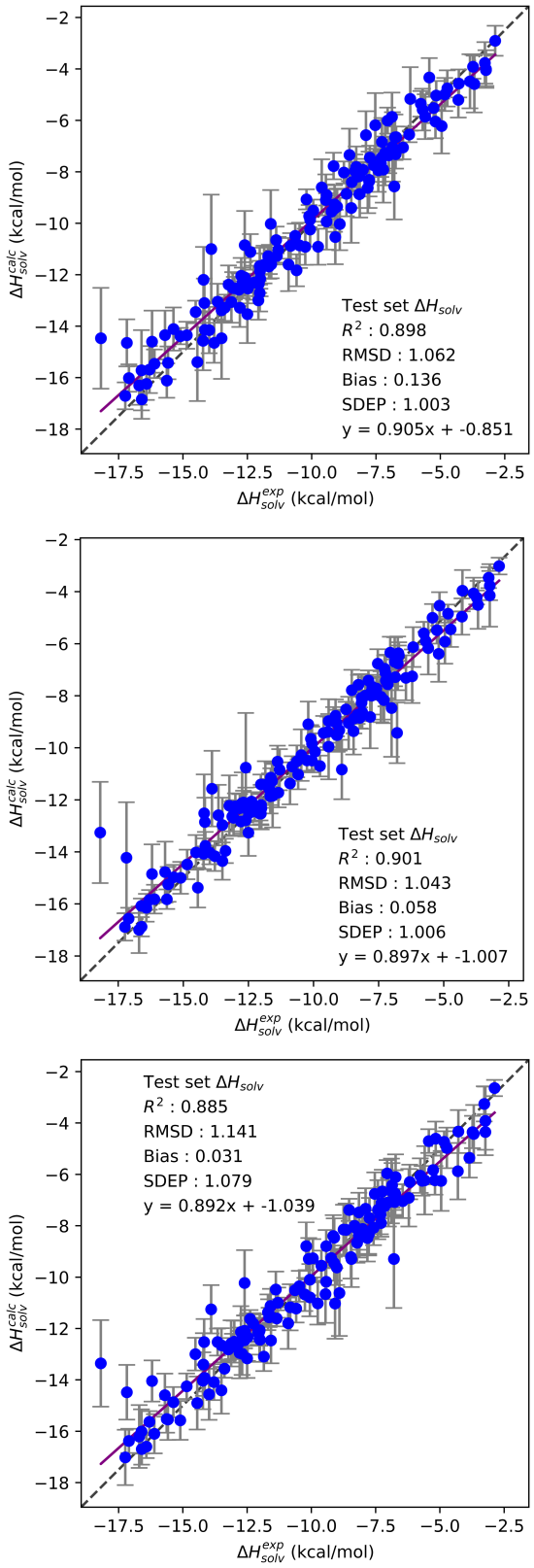

Figure 6 Correlation plots showing solvation enthalpy predictions from multi-output CNN trained on the solvation thermodynamics dataset. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation enthalpy predictions made using KH, HNC or GF SFED, respectively.

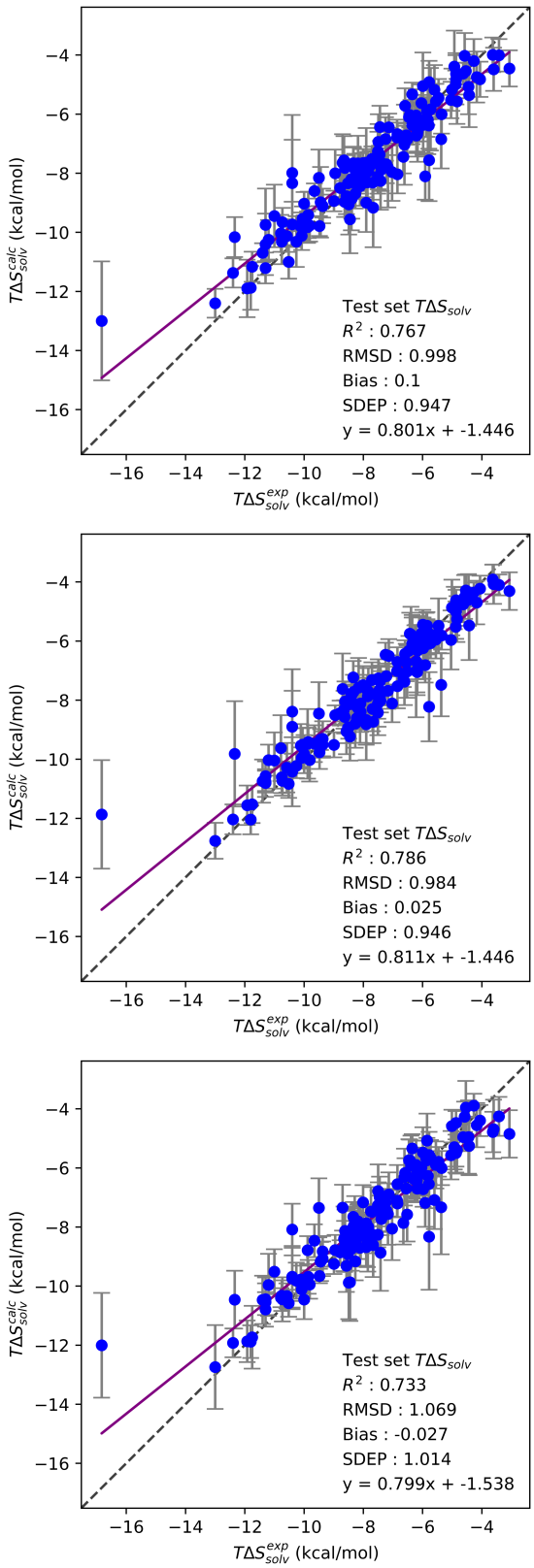

Figure 7 Correlation plots showing solvation entropy predictions from multi-output CNN trained on the solvation thermodynamics dataset. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation entropy predictions made using KH, HNC or GF SFED, respectively.

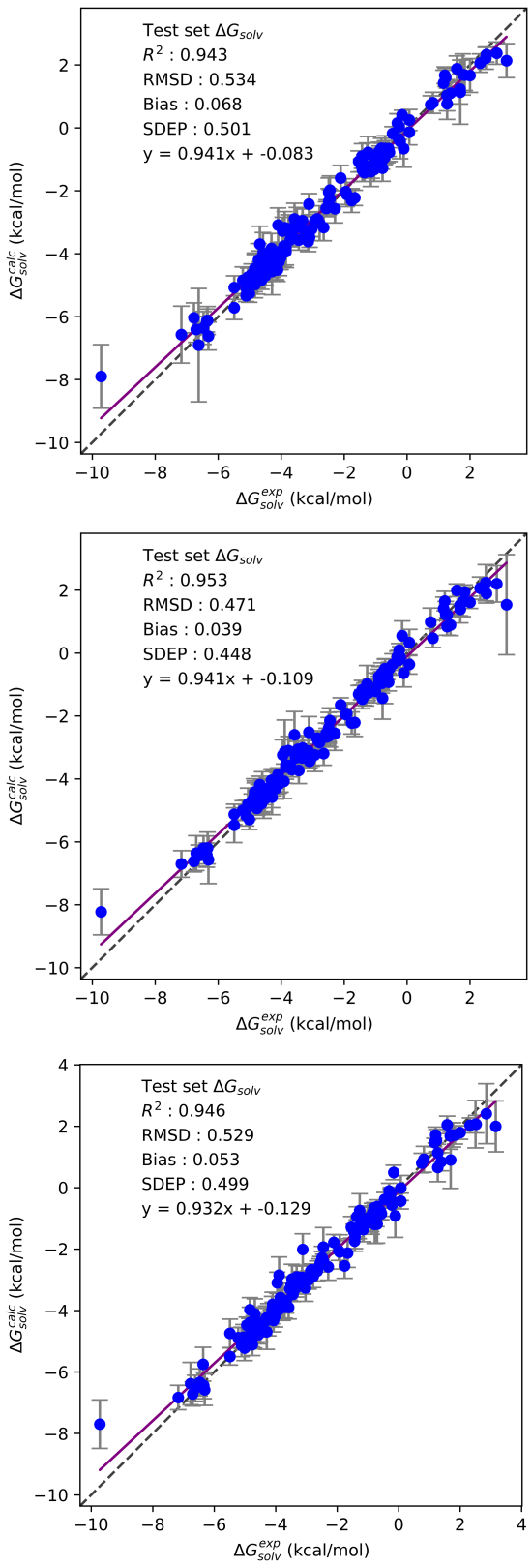

Figure 8 Correlation plots showing solvation free energy predictions from multi-output CNN trained on the solvation thermodynamics dataset. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation free energy predictions made using KH, HNC or GF SFED, respectively.

5 Solvation Thermodynamics - Single Task CNN

KH: Single output CNN - ΔH^{exp}_{solv}, $T\Delta S^{exp}_{solv}$, $\Delta G^{exp,neutral}_{solv}$ Dataset												
Water, 298K, 139 Datapoints	Solvation Enthalpy				Solvation Entropy				Solvation Free Energy			
	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
	0.89	1.09	0.13	1.04	0.76	1.02	0.03	0.98	0.96	0.42	0.02	0.4
HNC: Single output CNN - ΔH^{exp}_{solv}, $T\Delta S^{exp}_{solv}$, $\Delta G^{exp,neutral}_{solv}$ Dataset												
Water, 298K, 139 Datapoints	Solvation Enthalpy				Solvation Entropy				Solvation Free Energy			
	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
	0.89	1.12	-0.04	1.08	0.81	0.92	0.05	0.88	0.97	0.39	0.00	0.37
GF: Single output CNN - ΔH^{exp}_{solv}, $T\Delta S^{exp}_{solv}$, $\Delta G^{exp,neutral}_{solv}$ Dataset												
Water, 298K, 139 Datapoints	Solvation Enthalpy				Solvation Entropy				Solvation Free Energy			
	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP	R^2	RMSD	Bias	SDEP
	0.87	1.17	0.08	1.13	0.72	1.12	0.07	1.08	0.97	0.40	0.02	0.38

Table 5 Breakdown of solvation enthalpy, entropy and free energy predictions made using single output CNN trained on SFED generated using the KH, HNC and GF free energy functionals. A single CNN was trained on each individual dataset of experimental solvation enthalpy, entropy and free energy data. Errors are given in kcal/mol.

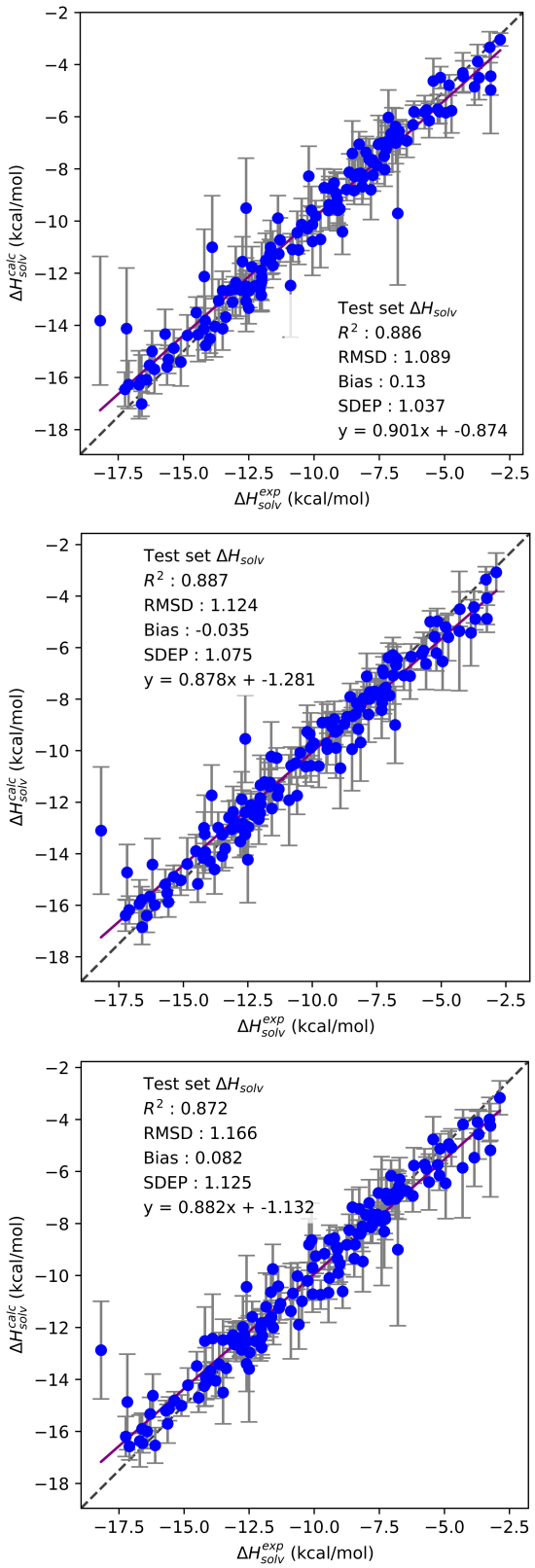

Figure 9 Correlation plots showing solvation enthalpy predictions from single output CNN trained on the solvation thermodynamics dataset. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation enthalpy predictions made using KH, HNC or GF SFED, respectively.

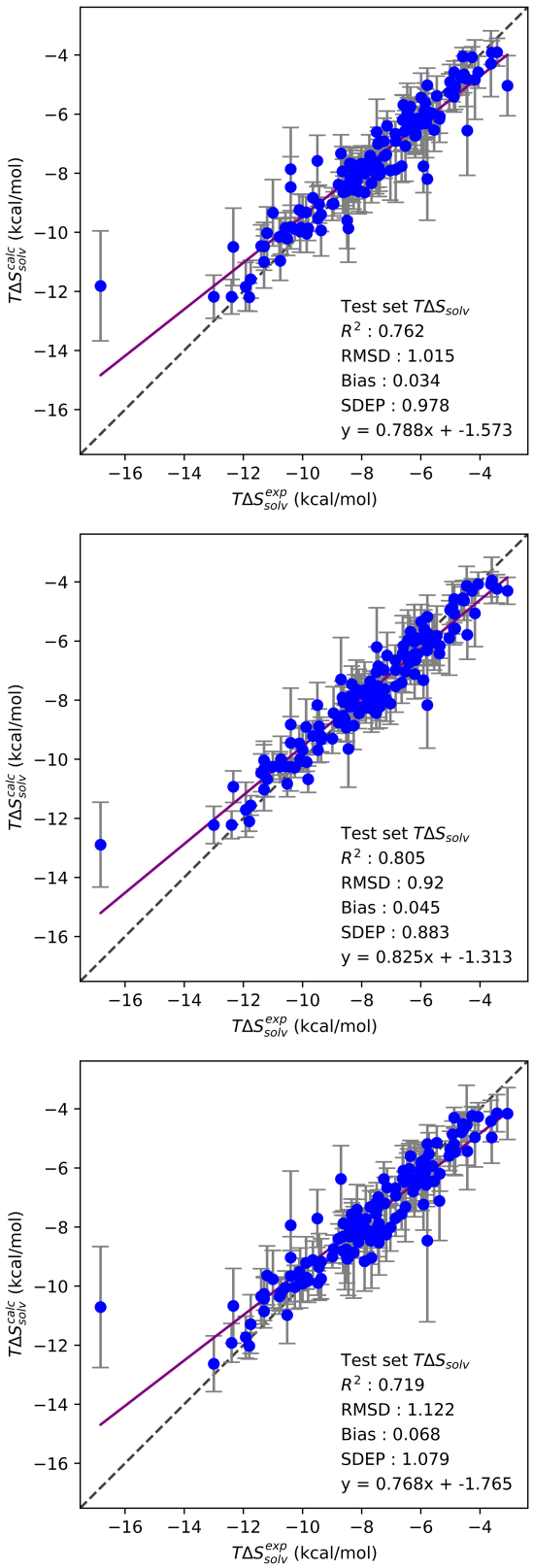

Figure 10 Correlation plots showing solvation entropy predictions from single output CNN trained on the solvation thermodynamics dataset. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation entropy predictions made using KH, HNC or GF SFED, respectively.

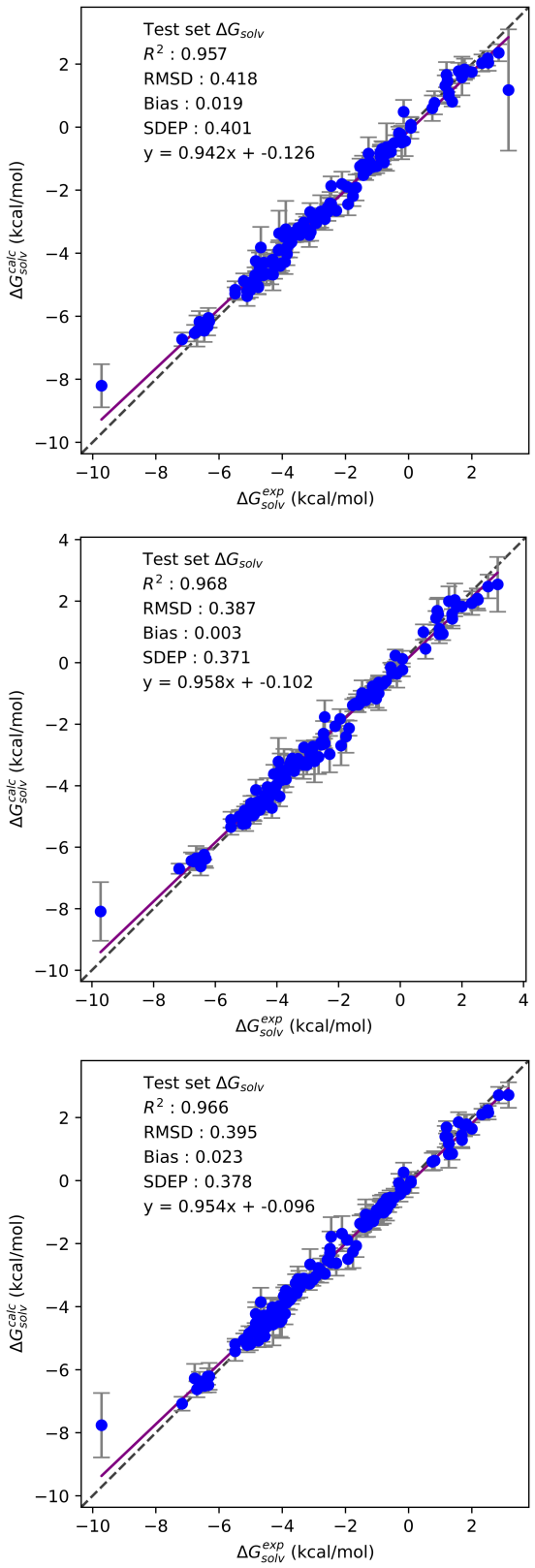

Figure 11 Correlation plots showing solvation free energy predictions from single output CNN trained on the solvation thermodynamics dataset. SFED datasets were generated using the KH, HNC and GF functionals. For each plot, error bars with the standard deviation of predictions for each solute are provided. By row, plots show solvation free energy predictions made using KH, HNC or GF SFED, respectively.

6 Solvation Thermodynamics - Correlation Between Parameters

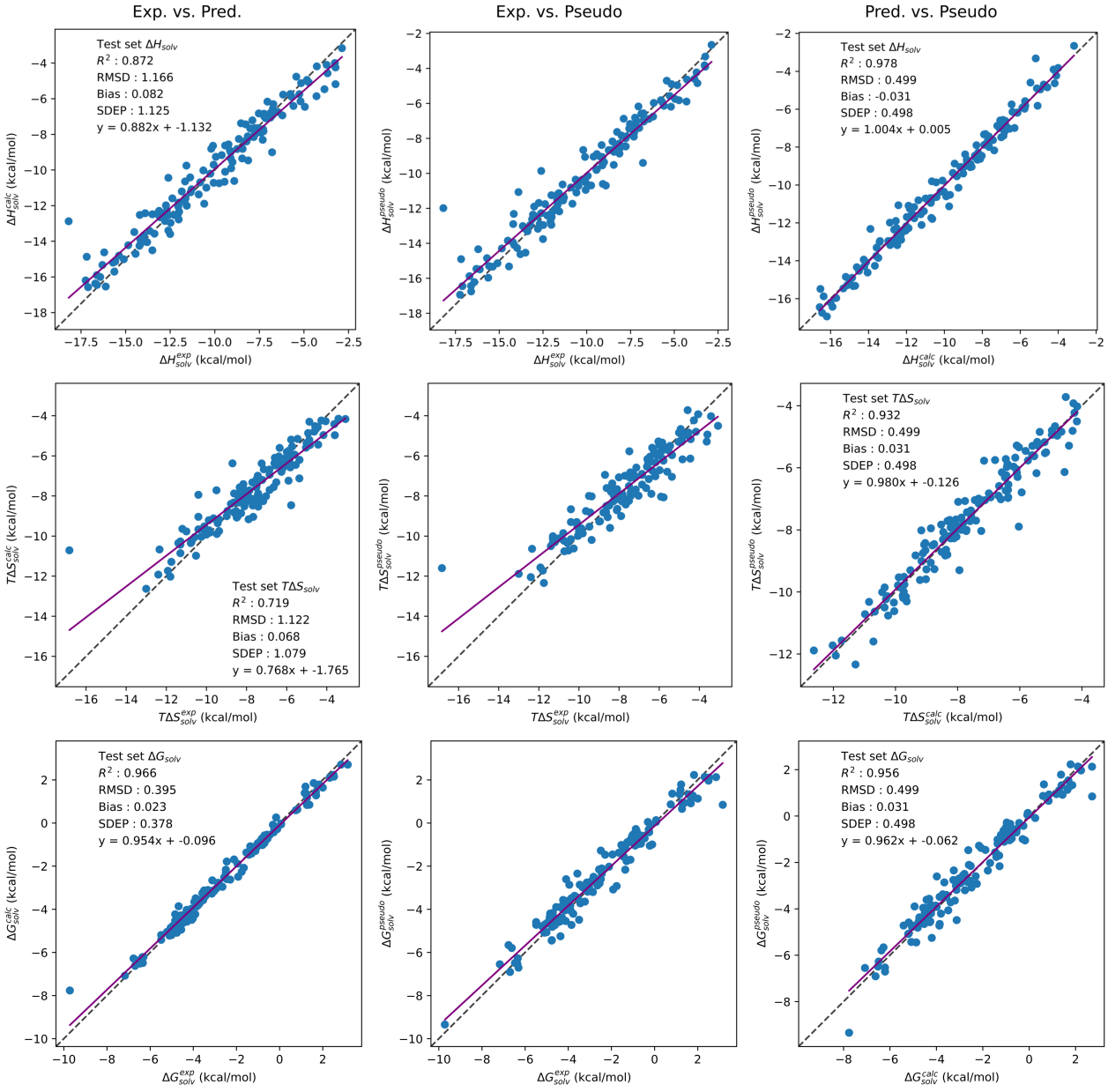

Figure 12 Correlation plots from a single output CNN trained on the solvation thermodynamics dataset. SFED were generated using the GF functional. From left to right, plots show correlation plots of experimental vs predicted values, experimental vs psuedo values and predicted vs pseudo values. Pseudo values are those calculated from the other two corresponding predicted values using $\Delta G = \Delta H - T\Delta S$. Statistics are given for experimental vs. predicted and predicted vs. pseudo values. Errors are shown in kcal/mol.

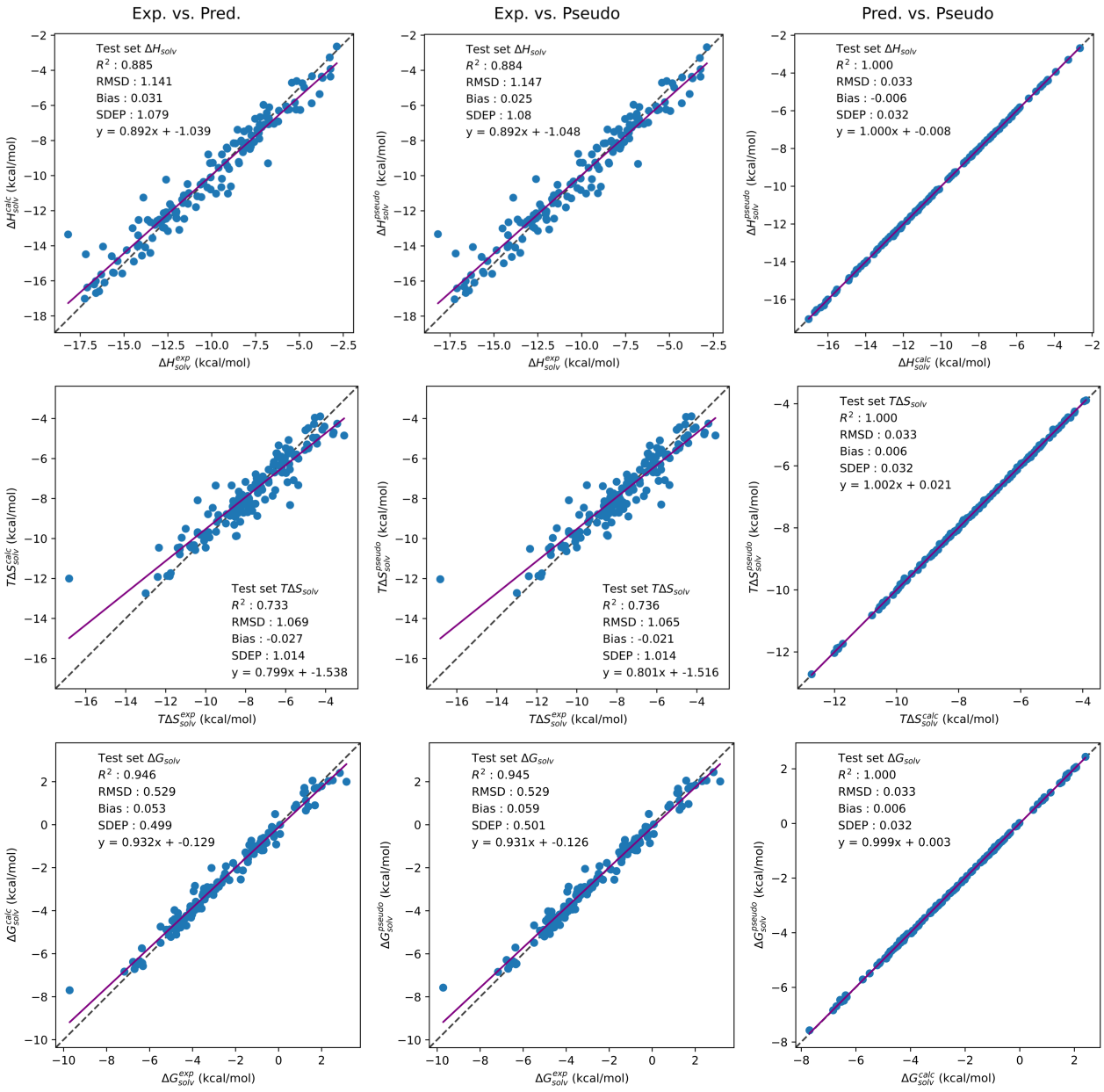

Figure 13 Correlation plots from a multi-output CNN trained on the solvation thermodynamics dataset. SFED were generated using the GF functional. From left to right, plots show correlation plots of experimental vs predicted values, experimental vs psuedo values and predicted vs pseudo values. Pseudo values are those calculated from the other two corresponding predicted values using $\Delta G = \Delta H - T\Delta S$. Errors are shown in kcal/mol.

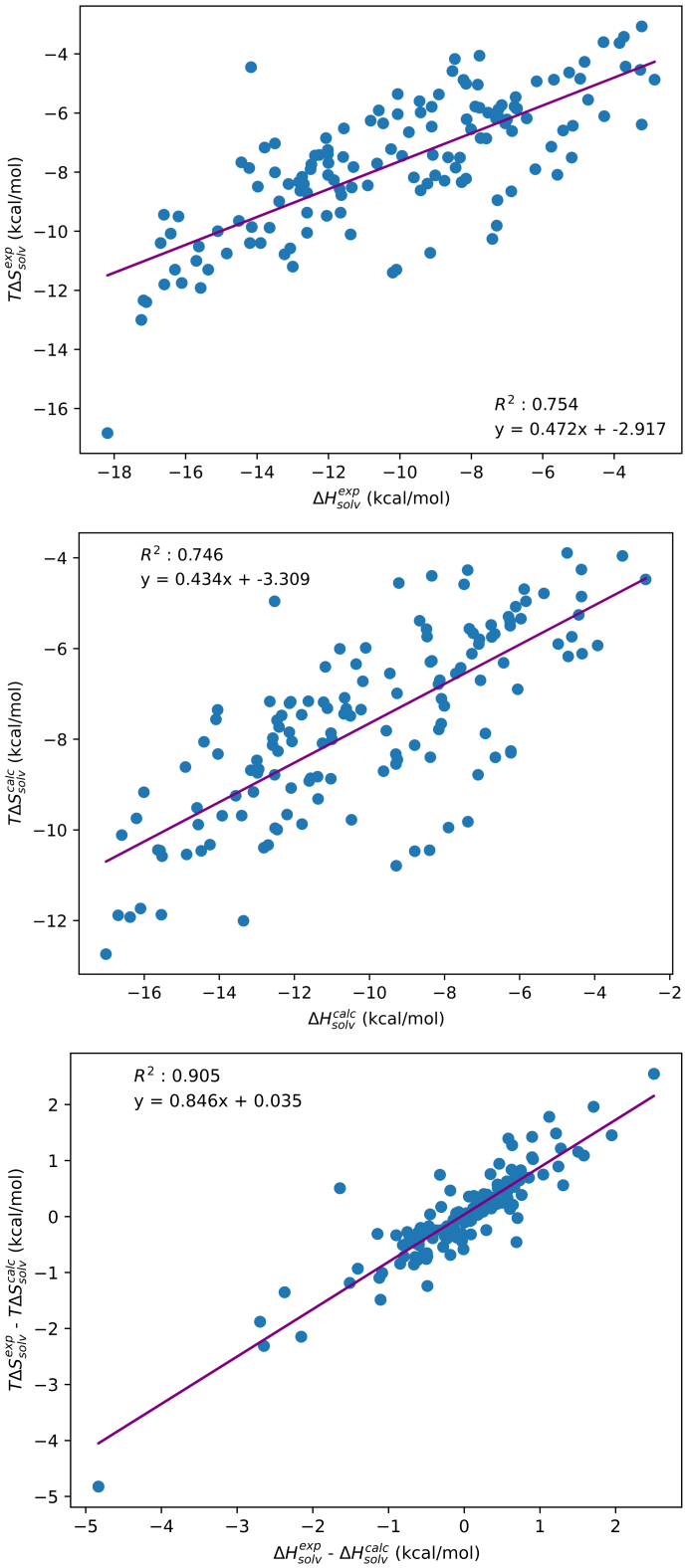

Figure 14 Correlation plots from a multi-output CNN trained on the solvation thermodynamics dataset. SFED were generated using the GF functional. From top to bottom, correlation plots of experimental solvation enthalpy vs experimental solvation entropy, predicted solvation enthalpy vs predicted solvation entropy, and the prediction errors for solvation enthalpy vs prediction errors for solvation entropy are shown. R^2 values show the Pearson correlation coefficient.