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

Combination of Explainable Machine Learning and Conceptual Density Functional Theory: Applications for the Study of Key Solvation Mechanisms

I-Ting Ho, Milena Matysik, Liliana Montano Herrera, Jiyoung Yang, Ralph Joachim Guderlei, Michael Laussegger, Bernhard Schrantz, Regine Hammer, Ramón Alain Miranda-Quintana, and Jens Smiatek*

E-mail: jens.smiatek@boehringer-ingelheim.com

Correlation Coefficients - Numerical Values

	ΔE_{CS}	ΔE_{AS}	ΔE_{CA}	$\Delta\Delta E_{sol}$	$\Sigma\Delta E_{ASCS}$	$\Delta\Delta E_{ASCS}$	E^{S}_{HOMO}	E^{S}_{LUMO}	χ_S	η_S	E^{C}_{HOMO}	E^{C}_{LUMO}	χ_C	η_C	E^{A}_{HOMO}	E^{A}_{LUMO}	χ_A	η_A
ΔE_{CS}	1.00	0.02	0.63	-0.40	0.99	-0.99	-0.13	-0.13	0.13	0.12	0.99	0.82	-0.99	-0.99	-0.03	-0.06	0.05	-0.03
ΔE_{AS}	0.02	1.00	0.71	-0.82	0.18	0.14	0.23	0.14	-0.19	-0.32	0.05	0.07	-0.05	-0.05	-0.88	-0.77	0.93	-0.02
ΔE_{CA}	0.63	0.71	1.00	-0.97	0.73	-0.51	0.01	-0.01	-0.00	-0.04	0.63	0.57	-0.63	-0.62	-0.64	-0.70	0.76	-0.17
$\Delta\Delta E_{sol}$	-0.40	-0.82	-0.97	1.00	-0.53	0.27	-0.05	-0.02	0.04	0.07	-0.41	-0.40	0.41	0.40	0.73	0.80	-0.87	0.20
$\Sigma\Delta E_{ASCS}$	0.99	0.18	0.73	-0.53	1.00	-0.95	-0.09	-0.10	0.10	0.07	0.99	0.82	-0.99	-0.98	-0.17	-0.18	0.20	-0.03
$\Delta\Delta E_{ASCS}$	-0.99	0.14	-0.51	0.27	-0.95	1.00	0.17	0.15	-0.16	-0.17	-0.98	-0.80	0.98	0.97	-0.11	-0.06	0.10	0.03
E^{S}_{HOMO}	-0.13	0.23	0.01	-0.05	-0.09	0.17	1.00	0.96	-0.99	-0.91	-0.07	-0.11	0.08	0.07	-0.05	-0.07	0.07	-0.03
E^{S}_{LUMO}	-0.13	0.14	-0.01	-0.02	-0.10	0.15	0.96	1.00	-0.98	-0.75	-0.09	-0.12	0.10	0.09	-0.04	-0.06	0.06	-0.02
χ_S	0.13	-0.19	-0.00	0.04	0.10	-0.16	-0.99	-0.98	1.00	0.86	0.08	0.12	-0.09	-0.08	0.05	0.07	-0.06	0.03
η_S	0.12	-0.32	-0.04	0.07	0.07	-0.17	-0.91	-0.75	0.86	1.00	0.03	0.07	-0.03	-0.03	0.05	0.08	-0.07	0.04
E^{C}_{HOMO}	0.99	0.05	0.63	-0.41	0.99	-0.98	-0.07	-0.09	0.08	0.03	1.00	0.79	-1.00	-1.00	-0.04	-0.06	0.05	-0.03
E^{C}_{LUMO}	0.82	0.07	0.57	-0.40	0.82	-0.80	-0.11	-0.12	0.12	0.07	0.79	1.00	-0.81	-0.76	-0.06	-0.10	0.09	-0.06
χ_C	-0.99	-0.05	-0.63	0.41	-0.99	0.98	0.08	0.10	-0.09	-0.03	-1.00	-0.81	1.00	1.00	0.04	0.06	-0.06	0.03
η_C	-0.99	-0.05	-0.62	0.40	-0.98	0.97	0.07	0.09	-0.08	-0.03	-1.00	-0.76	1.00	1.00	0.03	0.05	-0.05	0.03
E^{A}_{HOMO}	-0.03	-0.88	-0.64	0.73	-0.17	-0.11	-0.05	-0.04	0.05	0.05	-0.04	-0.06	0.04	0.03	1.00	0.54	-0.86	-0.35
E^{A}_{LUMO}	-0.06	-0.77	-0.70	0.80	-0.18	-0.06	-0.07	-0.06	0.07	0.08	-0.06	-0.10	0.06	0.05	0.54	1.00	-0.90	0.60
χ_A	0.05	0.93	0.76	-0.87	0.20	0.10	0.07	0.06	-0.06	-0.07	0.05	0.09	-0.06	-0.05	-0.86	-0.90	1.00	-0.19
η_A	-0.03	-0.02	-0.17	0.20	-0.03	0.03	-0.03	-0.02	0.03	0.04	-0.03	-0.06	0.03	0.03	-0.35	0.60	-0.19	1.00

Figure 1: Correlation coefficients for feature set.

Correlation Coefficients - Heatmap

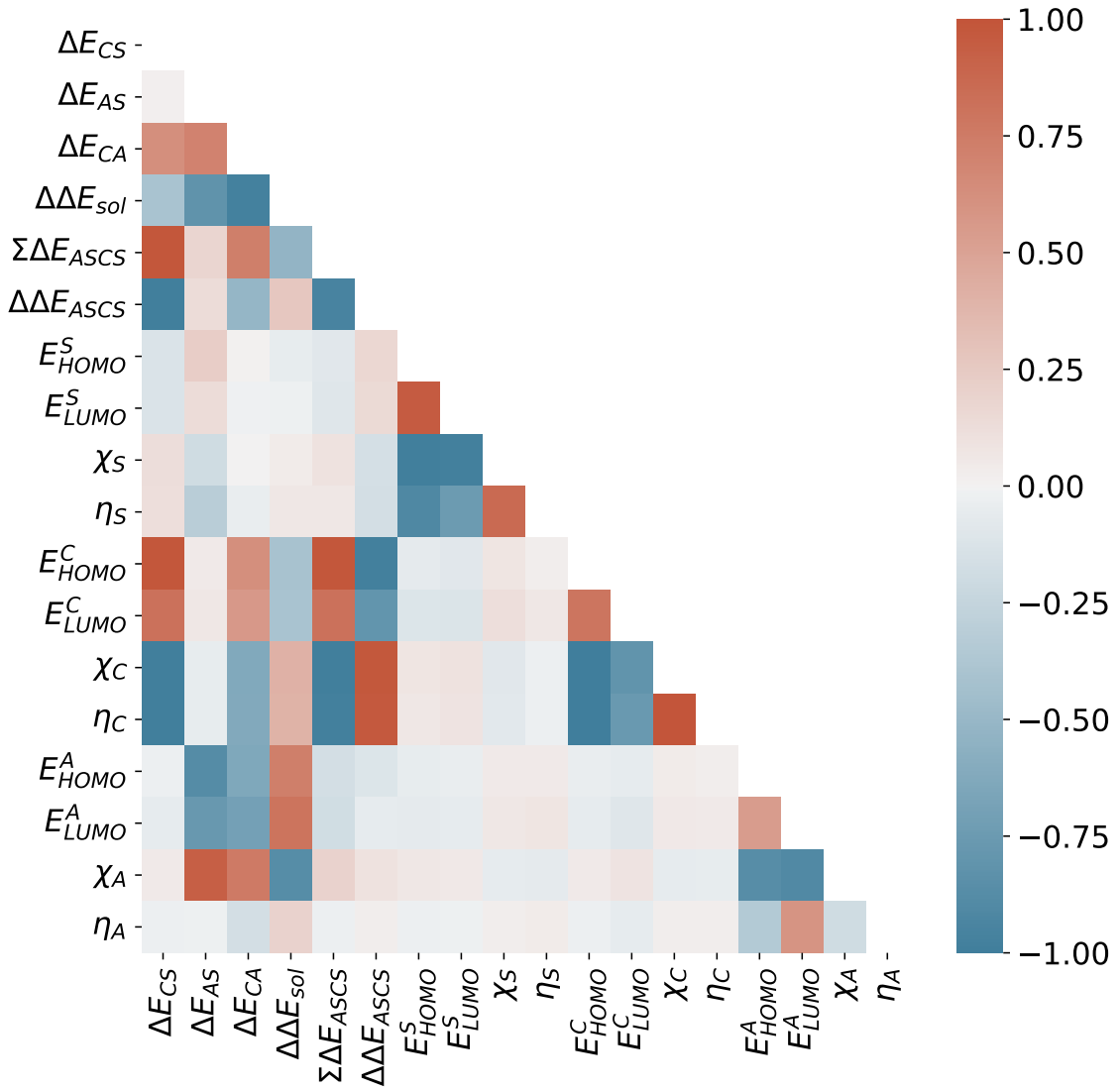

Figure 2: Correlation coefficients for feature set as heatmap.

Free Energy Calculations - Reduced XGB Model

Goodness of fit plot

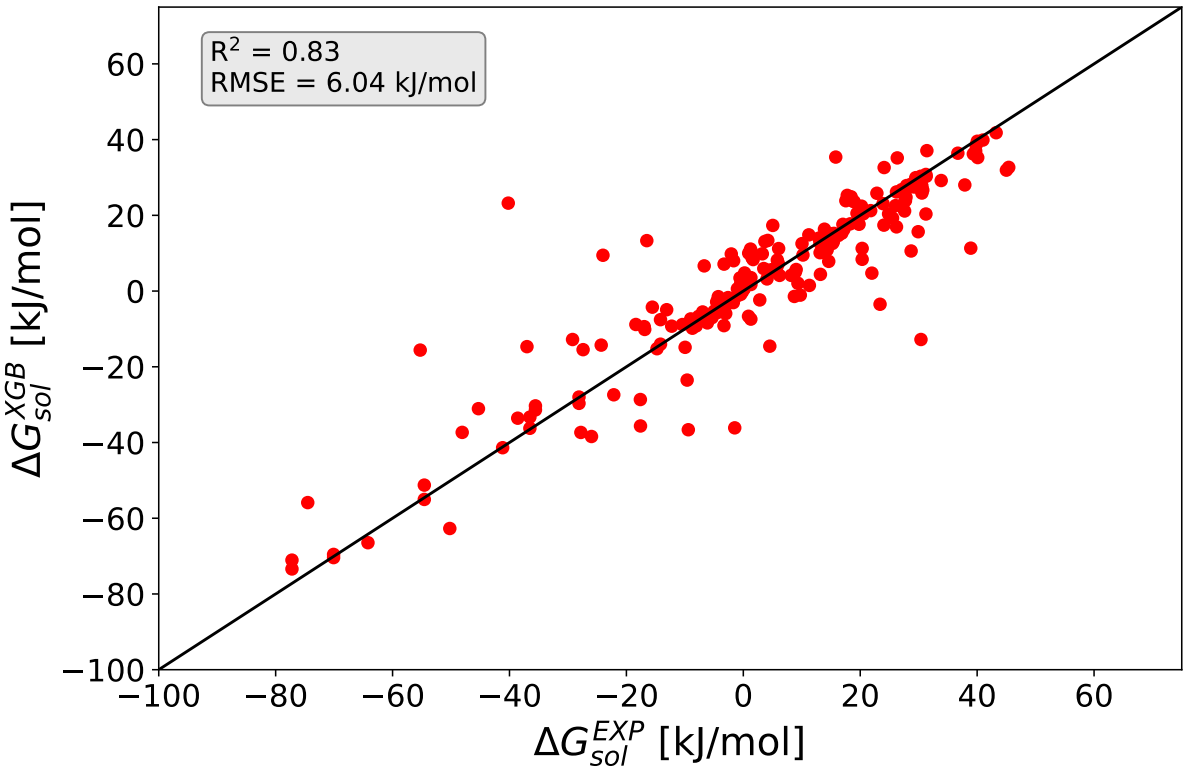

Figure 3: Goodness of fit plots for predicted (subscript XGB) and experimental values (subscript EXP) for ΔG_{sol} regarding the reduced XGB model. The straight black lines highlight a full coincidence.

SHAP analysis

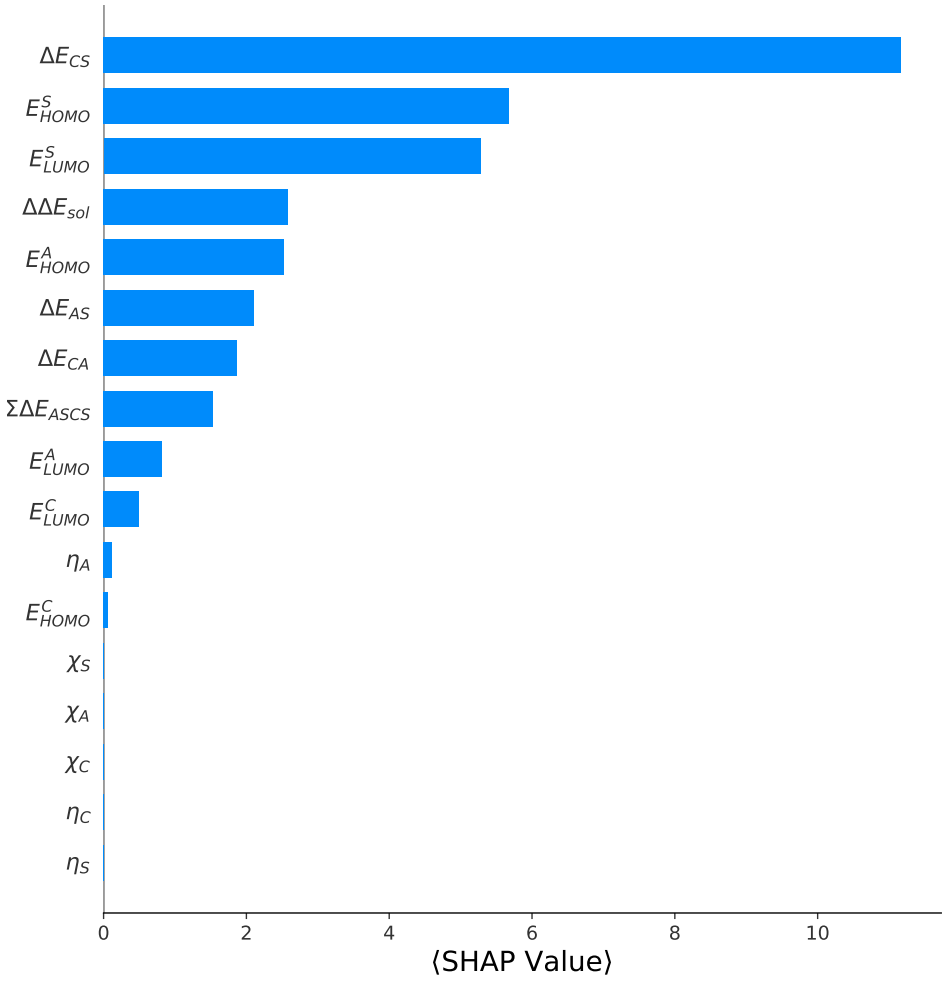

Figure 4: Net SHAP values for feature ranking of the reduced XGB model for the prediction of ΔG_{sol}.

Free Enthalpy Calculations - Reduced XGB Model

Goodness of fit plot

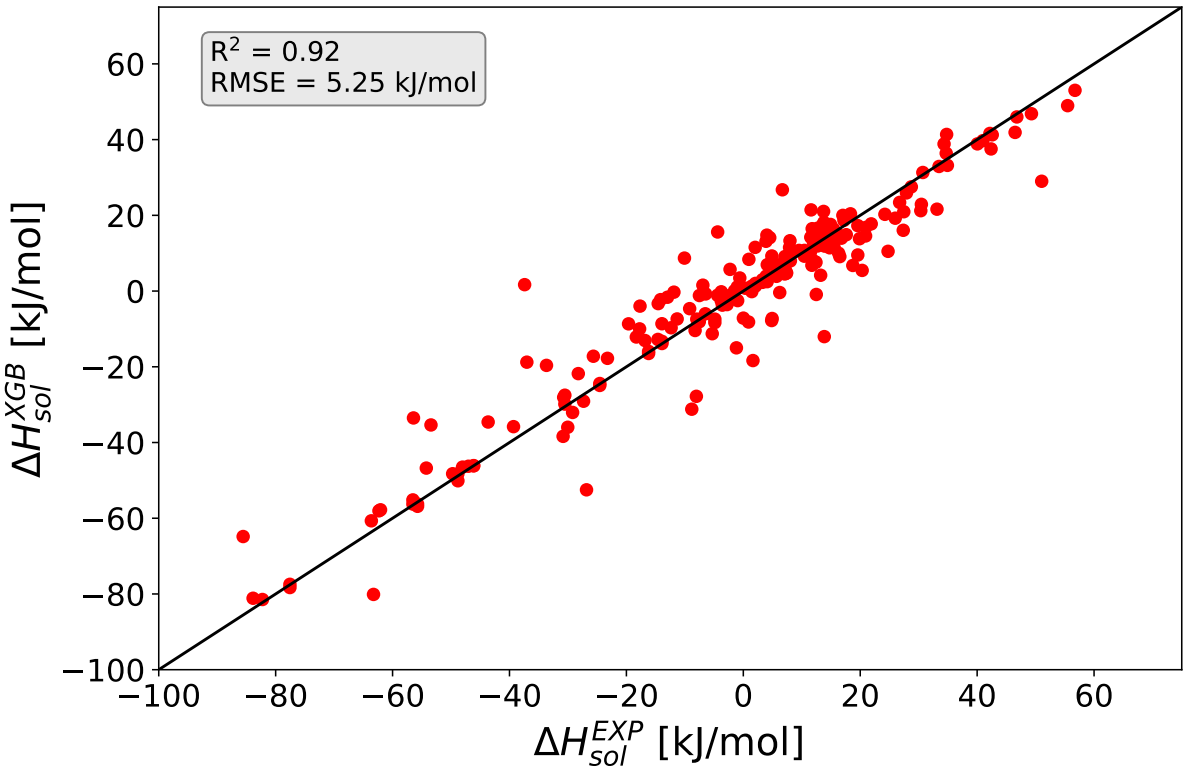

Figure 5: Goodness of fit plots for predicted (subscript XGB) and experimental values (subscript EXP) for ΔH_{sol} regarding the reduced XGB model. The straight black lines highlight a full coincidence.

SHAP analysis

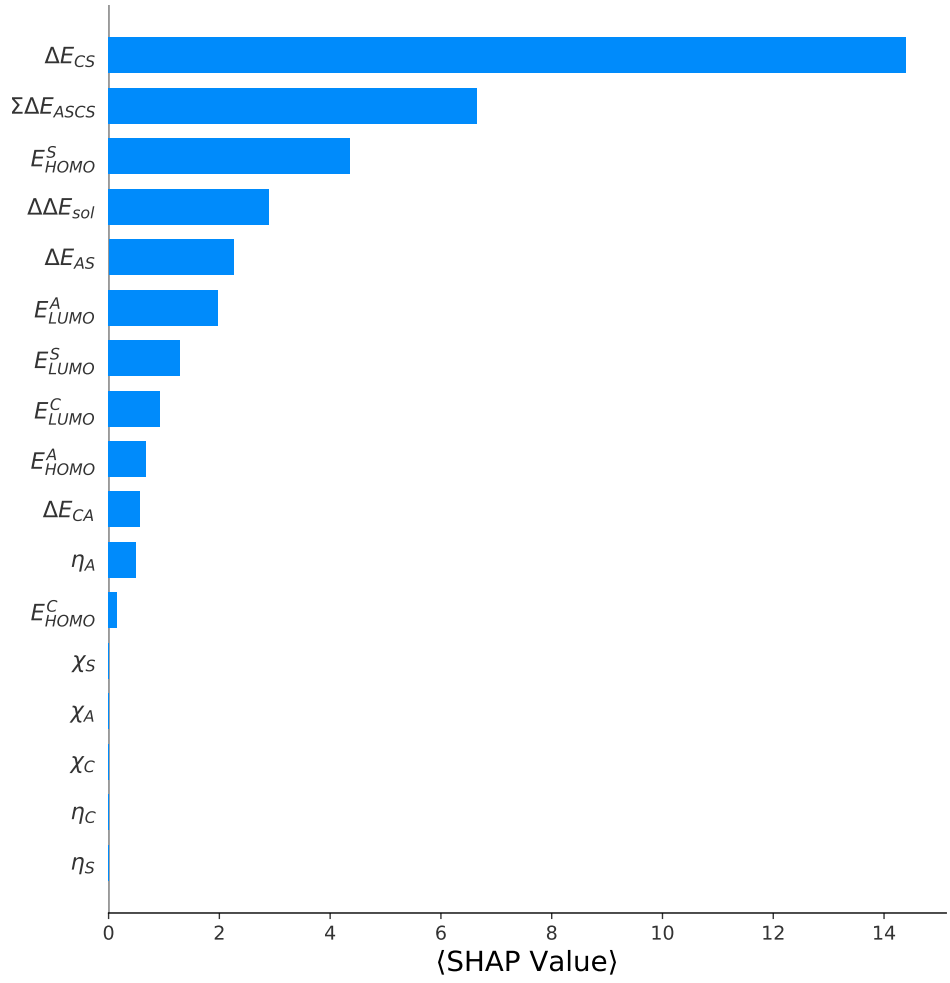

Figure 6: Net SHAP values for feature ranking of the reduced XGB model for the prediction of ΔH_{sol}.

Free Entropy Calculations - Reduced ET Model

Goodness of fit plot

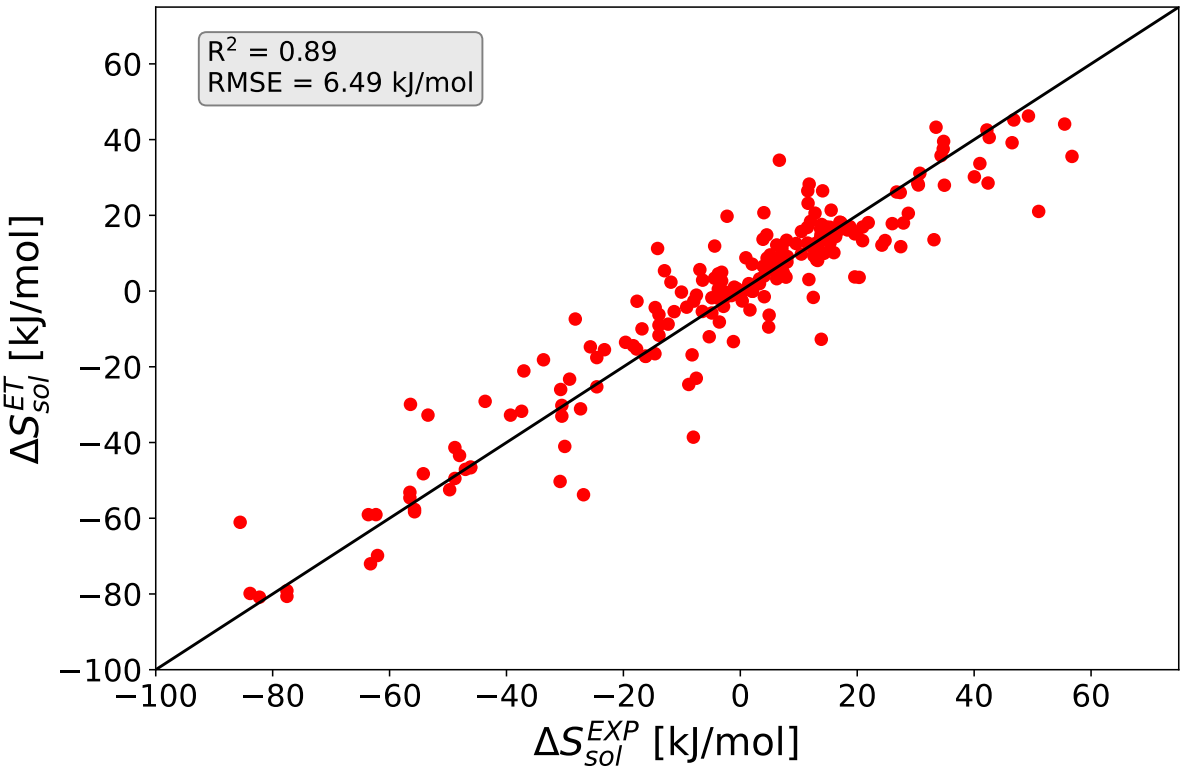

Figure 7: Goodness of fit plots for predicted (subscript XGB) and experimental values (subscript EXP) for ΔS_{sol} regarding the reduced ET model. The straight black lines highlight a full coincidence.

SHAP analysis

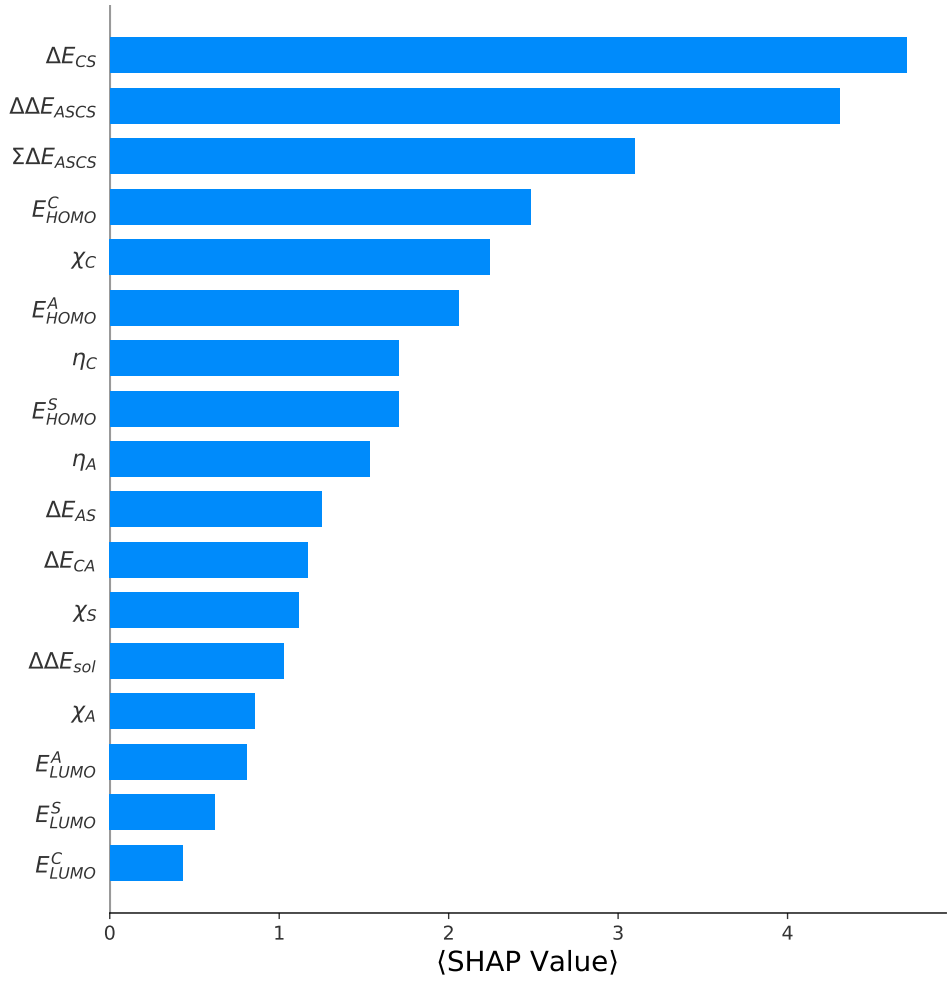

Figure 8: Net SHAP values for feature ranking of the reduced ET model for the prediction of ΔS_{sol}.