

Solvent Effects on CO₂ Capture by Simple Amino Acids: An Integrated Density Functional Theory – Machine Learning Approach

Mukul^{a,b} & Sandhiya Lakshmanan^{*a,b}

^aCSIR-National Institute of Science Communication and Policy Research, New Delhi – 110012,
India

^bAcademy of Scientific and Innovative Research (AcSIR), Ghaziabad – 201002, India

*Email: sandhiya.niscpr@csir.res.in

Explicit Solvation study

Methodology: To validate implicit solvation results and show specific hydrogen-bonding interactions we performed explicit solvation calculations on monocarbamate species. Explicit solvation of monocarbamate was examined by optimizing a 1:1 complex with Water, DMSO and Glycerol at M06-2X/6-311++G(d,p) level. Natural Bond Orbital (NBO) analysis was performed on the optimized complex of glycine monocarbamate & glycerol complex to quantify hydrogen-bonding interactions and charge transfer effects.

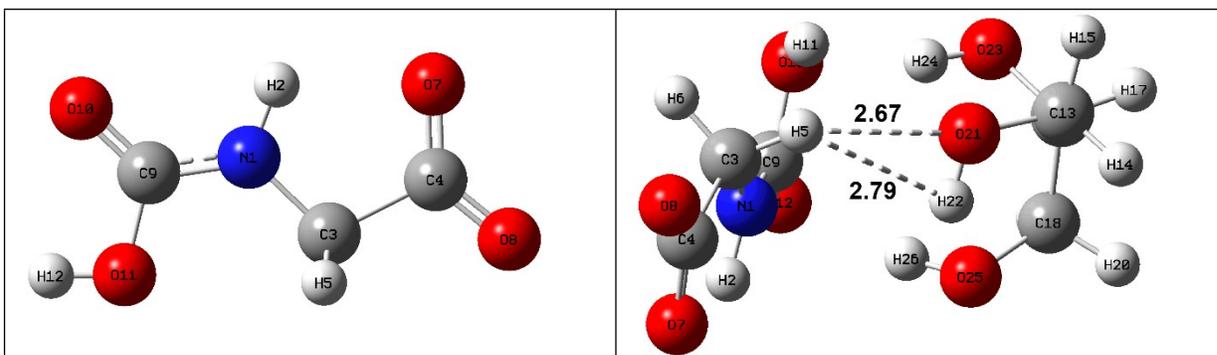


Figure S1: Optimized molecular structure of Glycine monocarbamate and Glycine monocarbamate & glycerol complex at M06-2X/6-311++G (d,p) level of theory. Dashed line indicates hydrogen bond with distance in Ångströms.

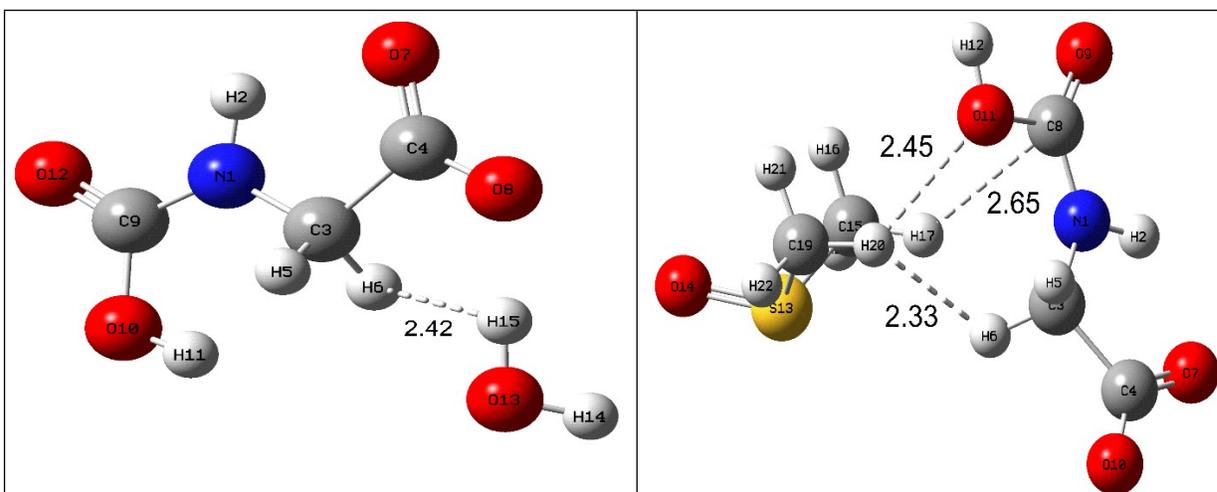


Figure S2: Optimized molecular structure of Glycine monocarbamate & water complex and Glycine monocarbamate & DMSO complex at M06-2X/6-311++G (d,p) level of theory. Dashed line indicates hydrogen bond with distance in Ångströms.

Table S1: NBO natural charges of Glycine monocarbamate and Glycine monocarbamate & glycerol complex from M06-2X/6-311++G(d,p) calculation.

Atoms	Monocarbamate	Monocarbamate Glycerol
-------	---------------	------------------------

		complex
N1	-0.67	-0.676
H2	0.443	0.452
C3	-0.295	-0.334
C4	0.784	0.793
H5	0.194	0.205
H6	0.194	0.195
O7	-0.807	-0.795
O8	-0.784	-0.77
C9	0.925	0.936
O10	-0.727	-0.733
O11	-0.734	-0.737
H12	0.475	0.499
C13		-0.044
H14		0.156
H15		0.177
C16		0.074
H17		0.17
C18		-0.047
H19		0.171
H20		0.163
O21		-0.783
H22		0.497
O23		-0.776
H24		0.495
O25		-0.797
H26		0.511

Table S2: Comparative explicit solvation benchmark of Glycine monocarbamate (GMC) 1:1 complexes with DMSO, water, and glycerol optimized at M06-2X/6-311++G(d,p) level of theory.

Parameters	Water	DMSO	Glycerol
Eps (Dielectric Constant)	78.4	46.7	42.5
H Bond Acidity	0.82	0	0.58
H Bond Basicity	0.35	0.8	0.78
H...A distances (Å)	2.42	2.33, 2.45, 2.65	2.67, 2.79
Number of H-bonds	1	3	2

Table S3: Glycine, Alanine and Serine anions values of enthalpy (ΔH_{298}) and free energy (ΔG_{298}) calculated at M06-2X/6-311+G(d,p) level of theory. The energy is given in kcal/mol.

Molecule	Medium	ΔH_{298}	ΔG_{298}
Glycine	Gas	-21.40	-11.22
Glycine	Water	-14.56	-4.37
Glycine	DMSO	-12.29	-1.55
Glycine	Glycerol	-53.75	-43.36
Glycine	Lactic Acid	-16.55	-6.03
Alanine	Gas	-16.05	-4.77
Alanine	Water	-14.50	-3.53
Alanine	DMSO	-12.03	-0.89
Alanine	Glycerol	-50.82	-38.60
Alanine	Lactic Acid	-13.15	-1.57
Serine	Gas	-8.31	2.74
Serine	Water	-10.42	0.62
Serine	DMSO	0.36	10.03
Serine	Glycerol	-50.82	-39.21
Serine	Lactic Acid	-13.67	-1.83

Table S4: Glycine, Alanine and Serine anions values of K_{eq} and CO_2 loading capacity (α) across all phases.

Molecule	Medium	K_{eq}	α
Glycine	Gas	1.70E+08	0.99
Glycine	Water	1.61E+03	0.99
Glycine	DMSO	1.38E+01	0.93
Glycine	Glycerol	6.12E+31	1
Glycine	Lactic Acid	2.66E+04	0.99
Alanine	Gas	3.17E+03	0.99
Alanine	Water	3.93E+02	0.99
Alanine	DMSO	4.50E+00	0.81
Alanine	Glycerol	2.00E+28	1
Alanine	Lactic Acid	1.43E+01	0.93
Serine	Gas	9.66E-03	0.00
Serine	Water	3.50E-01	0.25
Serine	DMSO	4.41E-08	4.40E-08
Serine	Glycerol	5.64E+28	1
Serine	Lactic Acid	2.20E+01	0.95

Table S5: Glycine anions values of enthalpy (ΔH_{298}) and free energy (ΔG_{298}) calculated at wB97XD/6-311+G(d,p) level of theory. The energy is given in kcal/mol.

Molecule	Medium	ΔH_{298}	ΔG_{298}
Glycine	Gas	-18.79	-8.49
Glycine	Water	-17.58	-7.19
Glycine	DMSO	-13.93	-3.76
Glycine	Glycerol	-19.50	-9.02
Glycine	Lactic Acid	-19.23	-8.72

Table S6: Free energies and reaction ΔG (kcal/mol) for amino acid + $\text{CO}_2(\text{solv}) \rightarrow$ monocarbamate anion(solv) across solvents. All values computed at M06-2X/6-311++G(d,p)/SMD. $G(\text{Reactants}) = G(\text{CO}_2) + G(\text{amino anion})$. $\Delta G = [G(\text{monocarbamate anion}) - G(\text{reactants})] \times 627.509 \text{ kcal/mol} \cdot \text{hartree}^{-1}$.

Molecule	Medium	G (CO_2)	G (Amino Anion)	G (Reactants)	G (Monocarbamate Anion)	ΔG (kcal/mol)
Glycine	Gas	-188.58356	-283.80635	-472.38991	-472.4078	-11.226145
Glycine	Water	-188.58038	-283.93207	-472.51245	-472.51942	-4.37374122
Glycine	DMSO	-188.58581	-283.90788	-472.49369	-472.49617	-1.55622356
Glycine	Glycerol	-188.51515	-283.92326	-472.43841	-472.50751	-43.3609065
Glycine	Lactic Acid	-188.57366	-283.91987	-472.49353	-472.50315	-6.03664139
Alanine	Gas	-188.58356	-323.09683	-511.68039	-511.688	-4.7753473
Alanine	Water	-188.58038	-323.21112	-511.7915	-511.79714	-3.53915358
Alanine	DMSO	-188.58581	-323.18786	-511.77367	-511.77509	-0.89106349
Alanine	Glycerol	-188.51515	-323.20335	-511.7185	-511.780023	-38.606267
Alanine	Lactic Acid	-188.57366	-323.2006	-511.77426	-511.77677	-1.57504885
Serine	Gas	-188.58356	-398.32618	-586.90974	-586.90536	2.74849161
Serine	Water	-188.58038	-398.43104	-587.01142	-587.01043	0.62123441
Serine	DMSO	-188.58581	-398.41311	-586.99892	-586.98293	10.0338769
Serine	Glycerol	-188.51515	-398.41983	-586.93498	-586.99748	-39.2193438
Serine	Lactic Acid	-188.57366	-398.41543	-586.98909	-586.99201	-1.83232774

Table S7: Hyperparameters used for the support vector regression (SVR), Gaussian process regression (GPR), and gradient boosting regression (GBR) models.

Model	Hyperparameter	Value
Support Vector Regression (SVR)	kernel	'rbf'
	C	100
	gamma	'auto'
	epsilon	0.01
Gaussian Process Regression (GPR)	kernel	Constant \times RBF + WhiteKernel
	Constant multiplier	1
	RBF length_scale	1
	WhiteKernelnoise_level	0.01
	n_restarts_optimizer	10
	random_state	42
Gradient Boosting Regression (GBR)	n_estimators	200
	learning_rate	0.05
	max_depth	4
	min_samples_split	5
	min_samples_leaf	2
	random_state	42

Table S8: Forecasting metrics coefficient of determination (R^2), mean absolute error (MAE), mean squared error (MSE) and root mean square error (RMSE) of different machine learning models.

Model	R^2	MAE	MSE	RMSE
SVR	-1.71	0.79	1.14	1.04
GPR	-3.11	0.72	1.06	0.91
GBR	0.85	0.13	0.08	0.22

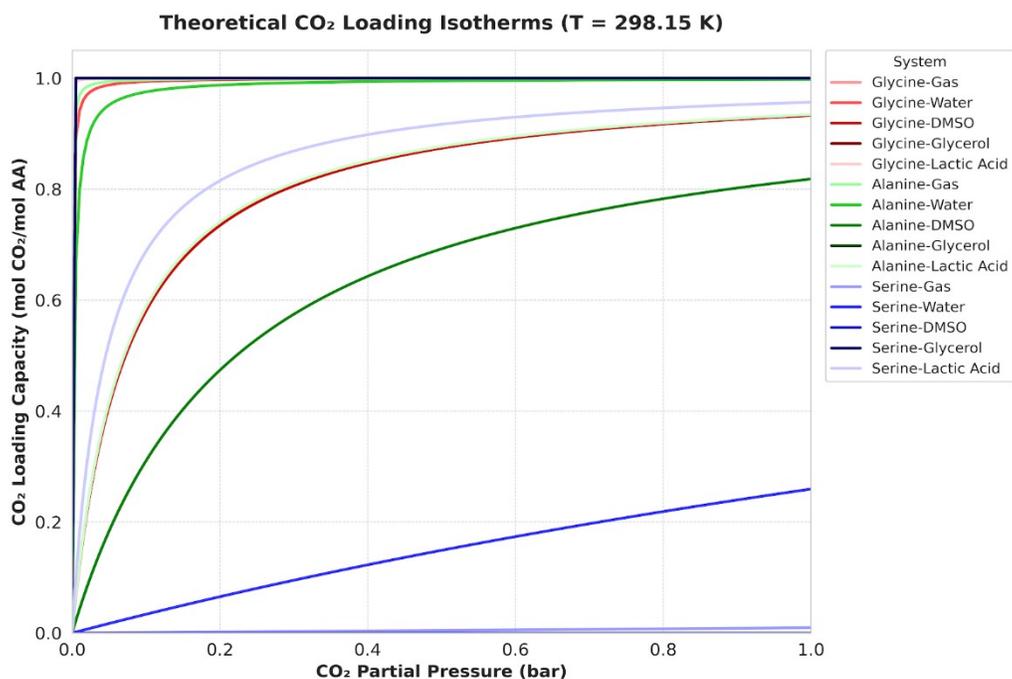


Figure S3: Theoretically CO₂ loading isotherms at 298.15 K for glycine, alanine and serine anions in five different media (gas phase, water, DMSO, glycerol, and lactic acid) using thermodynamically consistent Langmuir expression derived from DFT-calculated ΔG values (Eq. 2).

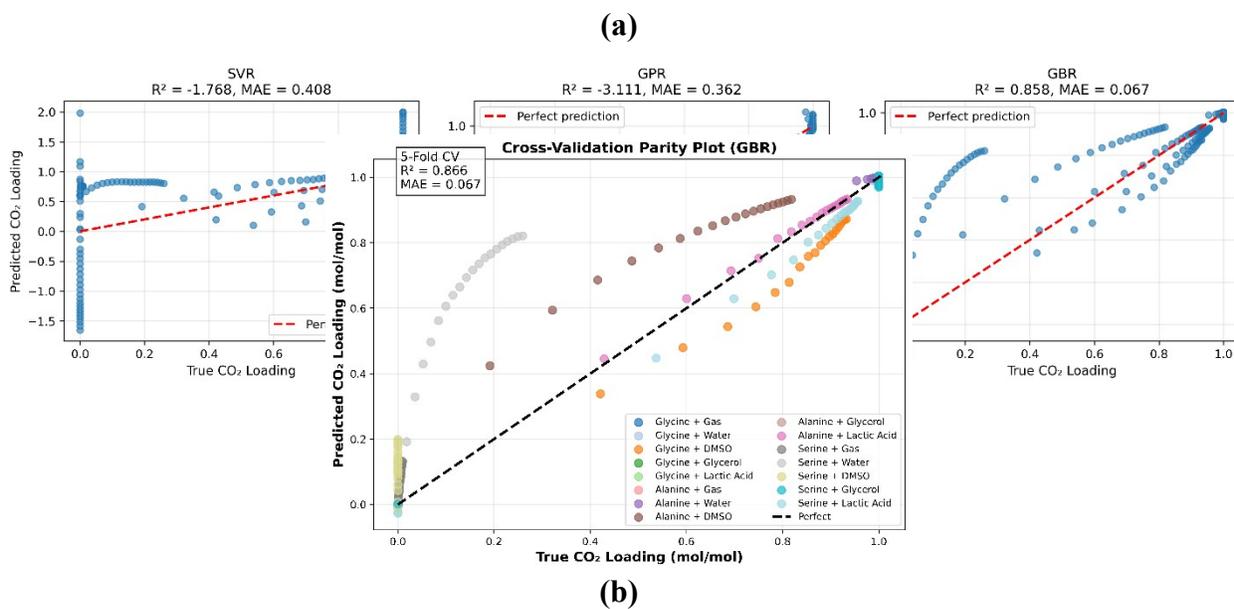


Figure S4: (a) Parity plots comparing the predictive performance of Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Gradient Boosting Regression (GBR) under system-level 5-fold cross-validation. (b) 5-fold cross-validation parity plot for GBR model using system-grouped splits. Overall performance: $R^2 = 0.866$, MAE = 0.067.

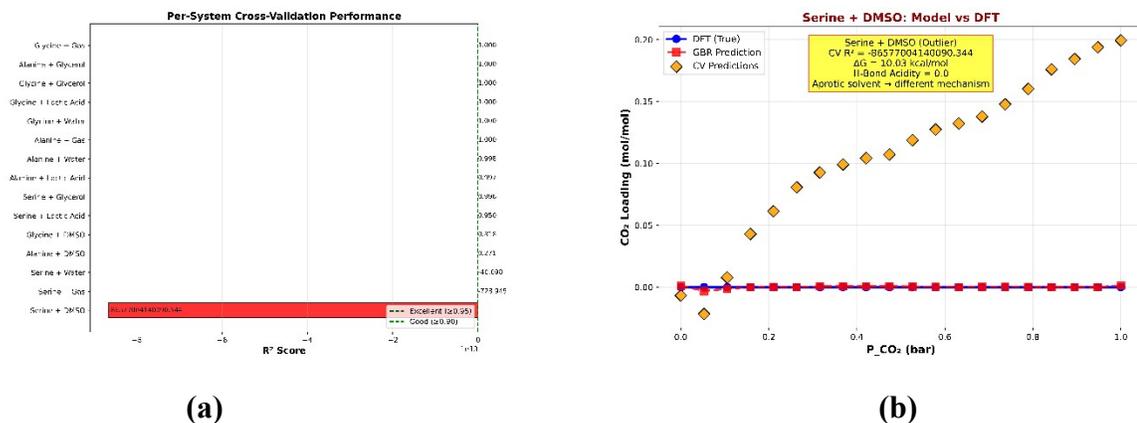


Figure S5: (a) Per-system cross-validation R^2 scores. Most systems achieve excellent performance ($R^2 \geq 0.95$) except Serine + DMSO. (b) CO_2 loading curve for Serine + DMSO. DFT-computed values (blue) vs GBR predictions (red) and individual fold CV predictions (orange diamonds). The large deviation confirms a distinct mechanism in this aprotic solvent (H-bond acidity = 0.0, $\Delta G = 10.03$ kcal/mol).

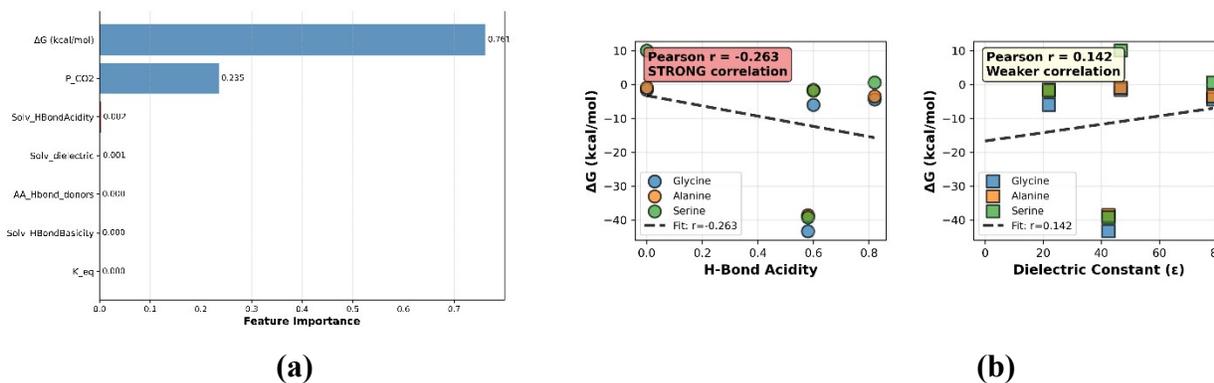


Figure S6: (a) Feature importance from GBR model. H-bond acidity slightly outweighs dielectric constant (ratio 1.6 \times) supporting its dominant role in reaction favorability. (b) Correlation of ΔG with solvent H-bond acidity and solvent dielectric constant.