Electronic Supplementary Information

Designing solvent systems using self-evolving solubility databases and graph neural networks

Yeonjoon Kim,[a,b] Hojin Jung,[a] Sabari Kumar,[a] Robert S. Paton,[a] Seonah Kim[a]*

[a]Department of Chemistry, Colorado State University, Fort Collins, CO 80523, United States
[b]Department of Chemistry, Pukyong National University, Busan 48513, Republic of Korea

Corresponding Author: Seonah Kim (Email: seonah.kim@colostate.edu)

S1. Effects of functional group distributions of solutes/solvents in Aug-DBs on the Root-mean-square Errors (RMSEs) of SSD models

Table S1. The number of data points for each **Aug-DB**.

Aug-DB-#	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
# of data points	183,432	65,833	41,560	34,013	32,474	34,611	34,421	35,701	34,792	35,617	34,134	31,074	30,626	29,906
Aug-DB-#	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
# of data points	26,324	27,968	21,448	23,263	16,959	19,311	14,931	15,951	12,215	10,901	12,160	8,927	7,142	8,521
Aug-DB-#	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**		
# of data points	7,705	6,346	5,603	5,605	4,522	3,708	3,168	3,664	3,280	3,233	2,961	2,249		

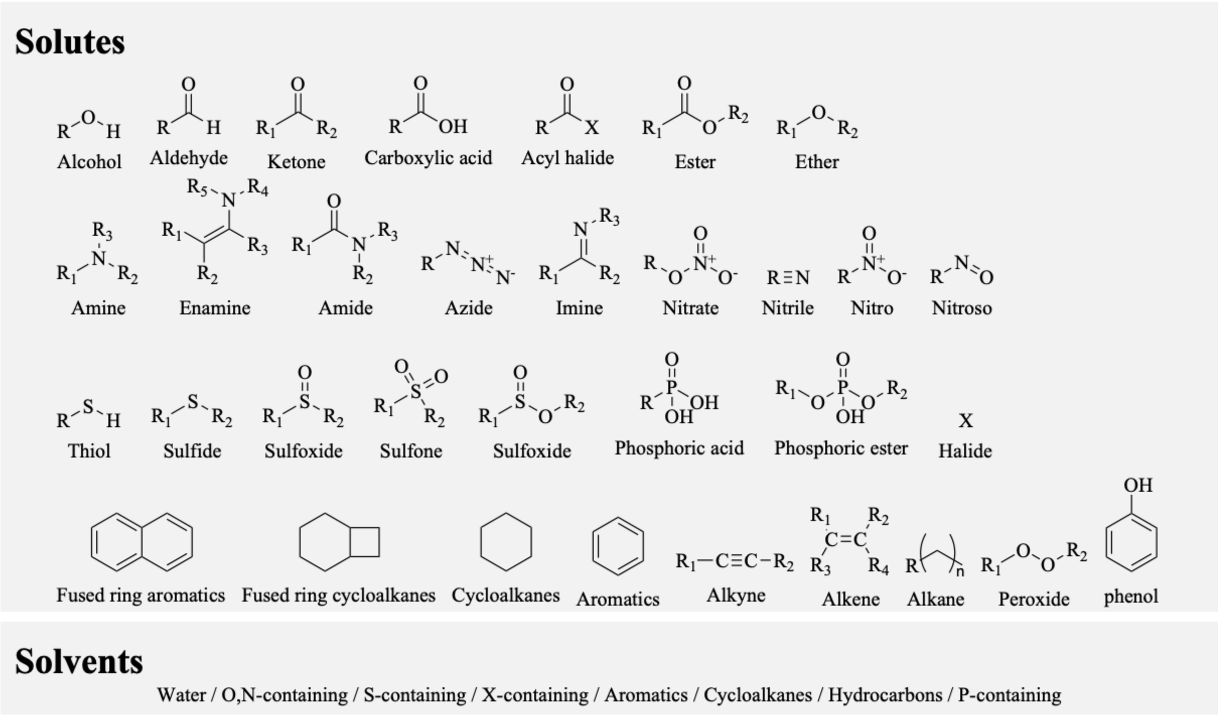


Fig. S1 Functional group categories of solutes and solvents in **Exp-DB** and **Aug-DB**s used for analyzing the effects of functional group distributions of databases on the RMSEs of SSD models.

Table S2. Solute-solvent functional groups that are overlapped between **Aug-DB** and **Exp-DB**, with the number of data points and mean absolute errors for each category, for three **Aug-DB**s.

FGs overlapped between Aug-DB-18 and Exp-DB (solute.solvent)	# of data points in Exp-DB	MAE
Thiol.Water	1	0.13
PhosphoricAcid.X-containing	1	1.65
Nitrile.Water	2	0.44
PhosphoricAcid.Aromatics	2	1.58
Cycloalkanes.Cycloalkanes	3	0.26
Total	**9**	**0.73**
FGs overlapped between Aug-DB-22 and Exp-DB (solute.solvent)	**# of data points in Exp-DB**	**MAE**
PhosphoricAcid.Hydrocarbons	3	1.13
Cycloalkanes.Aromatics	6	0.19
etc.X-containing	2	1.22
Peroxide.Water	1	0.13
PhosphoricAcid.X-containing	1	1.65
Thiol.Cycloalkanes	1	0.61
Cycloalkanes.S-containing	1	0.12
PhosphoricAcid.Aromatics	2	1.58
Total	**17**	**0.74**
FGs overlapped between Aug-DB-29 and Exp-DB (solute.solvent)	**# of data points in Exp-DB**	**MAE**
Sulfide.Cycloalkanes	1	1.21
Cycloalkanes.Aromatics	6	0.19
etc.X-containing	2	1.22
Ester.Water	26	1.16
Aldehyde.Water	7	0.43
Cycloalkanes.Hydrocarbons	13	0.31
Sulfide.Water	11	0.28
Arene.Cycloalkanes	4	0.23
Nitro.Cycloalkanes	1	0.65
Nitro.Water	12	1.13
Oxygen.O,N-containing	10	0.81
Total	**93**	**0.74**

The increasing/decreasing RMSEs for the **Exp-DB** test set (Fig. 3) are attributed to several factors. During the initial SSD cycles (~Cycle 15), more than 25,000 data points were added to the database (Table S1). In other words, the sizes of **Aug-DB**s until Cycle 16 are relatively more extensive than those after Cycle 16. Such larger **Aug-DB**s indicate that not sufficient data points have been added yet to accommodate new solutes/solvents having new functional groups (FGs) unseen by the model, presumably leading to the fluctuating RMSEs until Cycle 16 (Fig. 3). During the Cycles from 17 to 29, fewer data points were added (around 7,700 – 23,000, Table S1), and the RMSEs showed less fluctuation. However, an increase in RMSE compared to the previous cycle was still observed in Students 18 and 22 (Fig. 3). To explain these intermittent rises in RMSE, we hypothesized that RMSE increases if the added **Aug-DB** in the current cycle is less overlapped with **Exp-DB** in terms of FGs of solutes and solvents.

This hypothesis was verified by, first, assigning FGs of solutes and solvents in **Exp-DB** and **Aug-DB**s, based on the categories defined in Fig. S1. The assignment was performed automatically by matching substructure patterns between the simplified molecular-input line-entry system (SMILES) of a molecule and the SMILES arbitrary target specification (SMARTS) string of FGs. If a molecule has multiple FGs, the one with higher priority was assigned according to the priority order shown in Fig. S1. 33 categories were defined for solutes. For solvents, eight categories are sufficient to distinguish different FGs, according to the results of t-distributed stochastic neighbor embedding (t-SNE) clustering analysis shown in Fig. 7. After assigning FGs, we counted the number of data points in **Aug-DB**s whose FGs of solutes and solvents overlap with those in **Exp-DB**. This analysis was performed for **Aug-DB-18** and **Aug-DB-22** which led to an increase in RMSE for Students 18 and 22, and for **Aug-DB-29** which resulted in a decreasing RMSE for Student 29. Table S2 tabulates the number of data points whose solute-solvent FG pairs in **Aug-DB** overlap with those in **Exp-DB**, with the MAE between experimental and COSMO-RS solubilities. The ΔG_{solv} values from COSMO-RS were referred to here since it was used as a reference method during the SSD.

For **Aug-DB-18** and **Aug-DB-22**, FGs of only nine and 17 data points overlap with **Exp-DB**, whereas 93 data points in **Aug-DB-29** have common FGs with **Exp-DB**. As can be seen in Students 18 and 22, fewer data points having common FGs with **Exp-DB** leads to an increased RMSE, whereas Student 29 showed a decreasing trend because of a higher overlap of **Aug-DB-29** with **Exp-DB**. Of note, the accuracies of the reference COSMO-RS method are similar in all these three cases (MAEs of 0.73-0.74 kcal/mol). Therefore, it can be deduced that RMSEs of Student models are affected by the number of data points intersecting with the FGs of **Exp-DB** rather than the accuracy of the reference method.

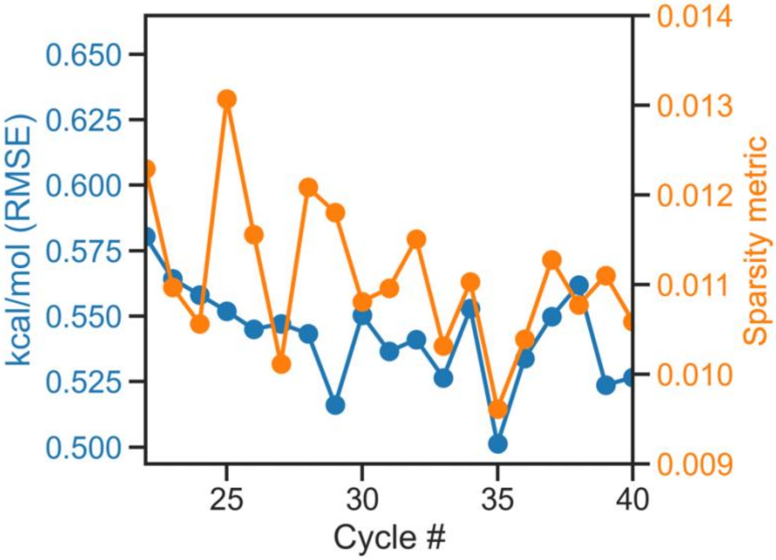
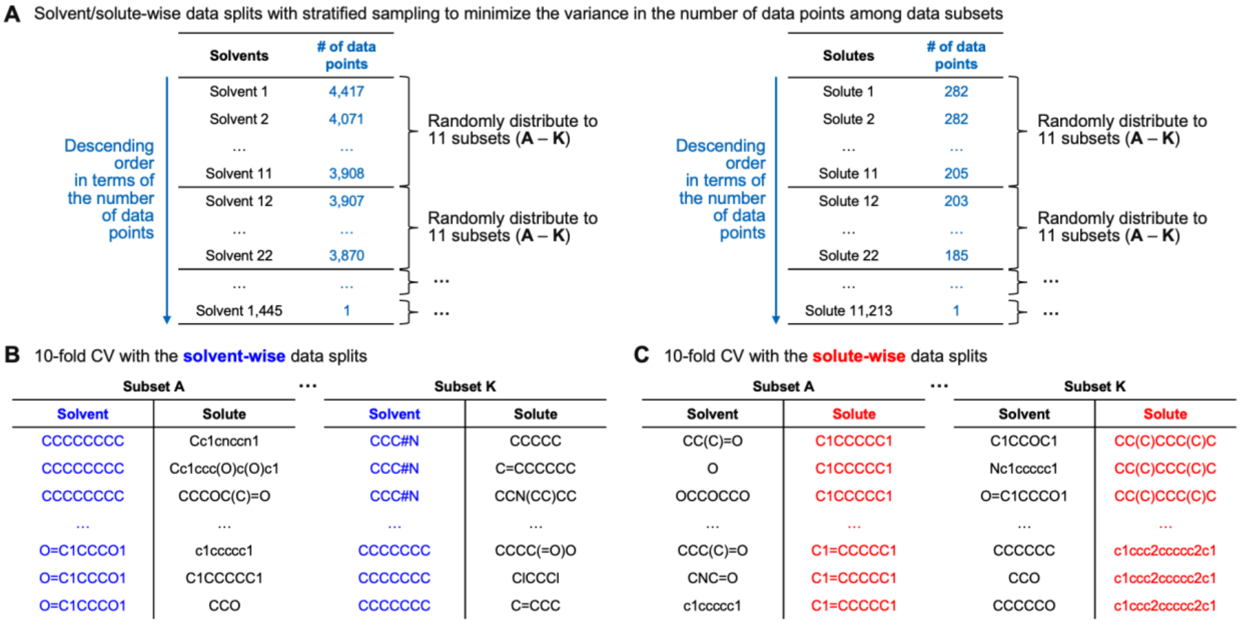


Fig. S2 The root-mean-square errors (RMSEs) vs. sparsity metric values in later SSD cycles (Cycle 22 – 40).

Meanwhile, after Cycle 30, the number of data points for the accumulated **Aug-DB**s exceeds 900,000, and the model is presumably prone to the ‘over-regularization’ by the solvents and solutes that frequently appeared in the accumulated **Aug-DB**s. In this case, the appearance of new solute-solvent pairs can improve the RMSEs of the **Exp-DB** test set for Students. For the quantitative validation of this hypothesis, we devised a metric for the sparsity in the distribution of augmented data points of each cycle as $\frac{\# \ of \ data \ points \ in \ the \ current \ Aug-DB}{\# \ unique \ solutes \times \# \ unique \ solvents}$. The metric value is 1.0 if all possible solute-solvent pairs exist in the accumulated database for the given solutes and solvents, and a value closer to 0.0 indicates more introduction of unseen solutes/solvents to the database. As shown in Fig. S2, from Cycle 30 to 37, the trend of the sparsity metric is mostly in line with the trend of RMSEs. This result demonstrates that adding unseen solute/solvent molecules (i.e., a decreased metric value) leads to SSD models that predict solubilities of problematic solute-solvent pairs in **Exp-DB** more accurately, and thus lower RMSEs.

S2. Detailed results of the 10-fold cross-validation using data subsets obtained from different data splitting methods



A
B
C
Descending order in terms of the number of data points
Randomly distribute to 11 subsets (A – K)
…

A Solvent/solute-wise data splits with stratified sampling to minimize the variance in the number of data points among data subsets

Solvents	# of data points
Solvent 1	4,417
Solvent 2	4,071
…	…
Solvent 11	3,908
Solvent 12	3,907
…	…
Solvent 22	3,870
…	…
Solvent 1,445	1

Solutes	# of data points
Solute 1	282
Solute 2	282
…	…
Solute 11	205
Solute 12	203
…	…
Solute 22	185
…	…
Solute 11,213	1

B 10-fold CV with the **solvent-wise** data splits

Subset A		…	Subset K	
Solvent	Solute		Solvent	Solute
CCCCCCCC	Cc1cnccn1		CCC#N	CCCCC
CCCCCCCC	Cc1ccc(O)c(O)c1		CCC#N	C=CCCCCC
CCCCCCCC	CCCOC(C)=O		CCC#N	CCN(CC)CC
…	…		…	…
O=C1CCCO1	c1ccccc1		CCCCCCC	CCCC(=O)O
O=C1CCCO1	C1CCCCC1		CCCCCCC	ClCCCl
O=C1CCCO1	CCO		CCCCCCC	C=CCC

C 10-fold CV with the **solute-wise** data splits

Subset A		…	Subset K	
Solvent	Solute		Solvent	Solute
CC(C)=O	C1CCCCC1		C1CCOC1	CC(C)CCC(C)C
O	C1CCCCC1		Nc1ccccc1	CC(C)CCC(C)C
OCCOCCO	C1CCCCC1		O=C1CCCO1	CC(C)CCC(C)C
…	…		…	…
CCC(C)=O	C1=CCCCC1		CCCCCC	c1ccc2ccccc2c1
CNC=O	C1=CCCCC1		CCO	c1ccc2ccccc2c1
c1ccccc1	C1=CCCCC1		CCCCCO	c1ccc2ccccc2c1

Fig. S3 (A) The stratified sampling scheme in solvent/solute-wise data splits to minimize the variance in the number of data points among different data subsets. Data subsets obtained from (B) solvent-wise data splitting, and (C) solute-wise data splitting, with SMILES strings of some example molecules.

For the stratified sampling, each bin of solvents/solutes was sorted in descending order with respect to the number of data points. Then, every 11 bins with a similar number of data points were grouped, and the bins in each group were distributed randomly to 11 subsets **A** – **K**. Such a stratified sampling results in a certain solvent/solute included in only one subset (Fig. S3). These are useful in assessing model accuracy when specific solvents or solutes in the validation/test set are unseen in the training set.

Table S3. Detailed information about the data set splits for the 10-fold cross-validation (**Fig. 6**) of the Student 35 models.

Data set	# of data points (**Exp-DB**)	# of data points (35 **Aug-DB**s)
Fold 1	1,048	82,893
Fold 2	1,048	82,888
Fold 3	1,048	82,886
Fold 4	1,047	82,882
Fold 5	1,047	82,877
Fold 6	1,047	82,874
Fold 7	1,047	82,871
Fold 8	1,047	82,868
Fold 9	1,047	82,865
Fold 10	1,047	82,861
Held-out test set	1,164	92,107

Table S4. Model accuracies of the 10-fold cross-validation with the random solute-solvent data set split for Student 35.

Exp-DB MAE (Validation sets from 10 folds, kcal/mol)[a]	**Exp-DB** MAE (Held-out test set, kcal/mol)[b]	**Exp-DB** RMSE (Validation sets from 10 folds, kcal/mol)[a]	**Exp-DB** RMSE (Held-out test set, kcal/mol)[b]
0.22	0.25±0.01	0.44	0.55±0.02

[a] Errors evaluated using all 10 validation sets collected from 10 folds. [b] Mean and standard deviation of MAEs/RMSEs from predicting solubilities of a held-out test set using 10 different models.

Table S3 lists the number of data points for each fold and held-out test set when the random solute-solvent sampling was performed. Each fold has almost the same number of data points so that the training:validation:test set ratio of 72:8:9 is maintained during each run of the 10-fold cross-validation. Table S4 summarizes the accuracies of the 10-fold cross-validation when the data splits in Table S3 were used. All MAEs and RMSEs of the 10 validation sets and held-out test set do not deviate significantly from **Exp-DB** test set MAEs (0.25 kcal/mol) and RMSEs (0.50 kcal/mol) evaluated when the best-case model was used. These results indicate the Student 35 models are scarcely susceptible to overfitting.

Table S5. Detailed information about the solvent-wise and solute-wise data set splits for the 10-fold cross-validation of the Student 35 models.[a]

Data set	Solvent-wise split				Solute-wise split			
	Exp-DB		35 **Aug-DB**s		**Exp-DB**		35 **Aug-DB**s	
	$N_{solvent}$	N_{data}	$N_{solvent}$	N_{data}	N_{solute}	N_{data}	N_{solute}	N_{data}
Fold 1	130	706	26	84,468	205	1,218	1,022	86,689
Fold 2	132	1,127	25	82,040	228	961	1,012	86,083
Fold 3	130	1,101	26	84,466	216	895	998	85,755
Fold 4	131	1,131	25	81,570	197	960	1,002	85,121
Fold 5	131	979	26	83,769	214	1,219	991	84,473
Fold 6	132	1,065	26	85,212	174	1,063	991	84,446
Fold 7	132	648	26	83,482	208	801	981	83,765
Fold 8	131	1,043	26	82,410	214	849	963	82,970
Fold 9	131	770	26	84,672	223	1,391	954	81,273
Fold 10	132[b]	2,079[b]	26[b]	84,771[b]	198	1,196	939	80,360
Held-out test set	131	988	26	84,012	198	1,084	938	79,937

[a]The number of data points varies among different split data sets because of different number of available data points for each unique solute/solvent. However, the stratified data set split was carried out so that the variance of the number of data points is minimized while being randomly sampled. [b]This split data set contains the water solvent.

Table S6. Model accuracies of the 10-fold cross-validation with solvent/solute-wise, stratified data set split for Student 35.

	Exp-DB MAE (Validation sets from 10 folds, kcal/mol)[a]	**Exp-DB** MAE (Held-out test set, kcal/mol)[b]	**Exp-DB** RMSE (Validation sets from 10 folds, kcal/mol)[a]	**Exp-DB** RMSE (Held-out test set, kcal/mol)[b]
Solvent-wise split	0.55 [0.27[c]]	0.28±0.02	1.23 [0.54[c]]	0.50±0.02
Solute-wise split	0.34	0.30±0.01	0.67	0.55±0.02

[a]Errors evaluated using all 10 validation sets collected from 10 folds. [b]Mean and standard deviation of MAEs/RMSEs from predicting solubilities of a held-out test set using 10 different models. [c]MAEs/RMSEs evaluated without the data points having the water solvent.

Similar analysis was carried out under the solvent/solute-wise data splits. This is to verify that the model accuracies are maintained when solubilities involving certain solutes and solvents are unseen in the training set and evaluated in the validation or test sets. Table S5 shows the number of data points in each subset of data when the solvent/solute-wise data splits were conducted. Although the differences in the number of data points are higher than those in Table S3, such deviations were minimized through the stratified sampling described in Fig. S3. Fold 10 in the solvent-wise split has particularly more data points, since it includes the water solvent for which more data points are available than for other solvents.

Table S6 summarizes the accuracies from the 10-fold cross-validation with the solvent/solute-wise data splits. In the solvent-wise splits, water solvent was included in one of the validation sets, leading to high errors than those without the water solvent. This result indicates that it is challenging to obtain desirable accuracies when the water solvent is unseen because the chemical behavior of water as a solvent is significantly different from other organic solvents. Similar trends were also observed in the literature for the water solvent.[1] The validation set MAE and RMSE without water (0.27 and 0.54 kcal/mol, respectively) is comparable to those from the best case model (0.25 and 0.50 kcal/mol, respectively), demonstrating the robustness of the model even in the solvent-wise data splits. Test set MAE and RMSE are 0.28 and 0.50 kcal/mol with the standard deviations of 0.02 kcal/mol when the prediction was carried out for 10 times, which also indicates the low overfitting tendency of the model. In the solute-wise splits, MAEs and RMSEs are slightly higher than the other above cases, but they showed an increment of errors only around 0.1 kcal/mol.

Table S7. Prediction accuracies from the 10-fold cross-validation of Student 35 with the solvent-wise data split, for the 10 representative solvents in the **Exp-DB** held-out test set.

Exp-DB test set solvent	RMSE (kcal/mol)	MAE (kcal/mol)	# of data points
Heptane	0.48	0.20	211
Diethyl ether	0.55	0.37	140
Butyl acetate	0.38	0.26	97
Hexadecene	0.62	0.28	90
Methyl acetate	0.46	0.34	87
N-Formylmorpholine	0.44	0.37	61
Cyclopropane	0.48	0.33	53
Methyl isobutyl ketone	0.50	0.31	23
Propanenitrile	0.30	0.25	20
Dipropyl ether	0.26	0.21	16

Table S7 shows the prediction errors for the 10 representative solvents in the **Exp-DB** test set when the solvent-wise data splits were performed. The MAEs and RMSEs range from 0.20 – 0.37 kcal/mol and 0.26 – 0.62 kcal/mol, respectively. These accuracies are comparable to those in the literature around 0.1 kcal/mol differences (MAEs: 0.10 – 0.27 kcal/mol, RMSEs: 0.13 – 0.51 kcal/mol), although the test set solvents are different.[1]

S3. Other control models to demonstrate the feasibility of the SSD shown in Fig. 2A

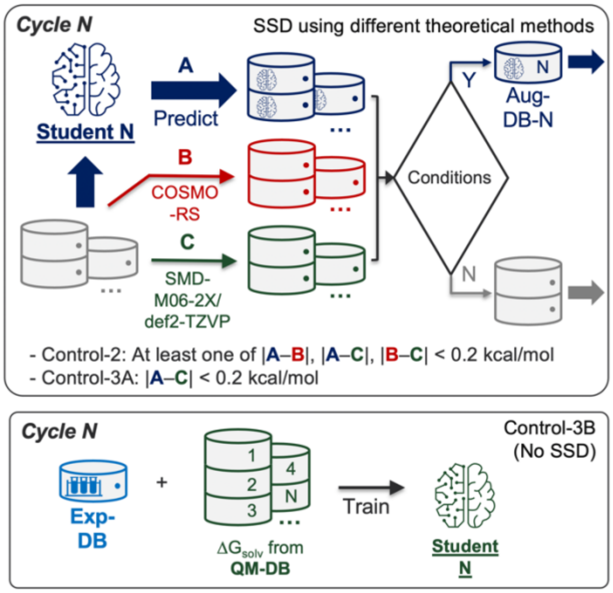


Fig. S4 Description of other Control models examined to demonstrate the feasibility of the SSD shown in **Fig. 2A**.

Table S8. Comparison of accuracies of Control models with the models from the SSD shown in **Fig. 2A**.

	SSD shown in **Fig. 2A**		Control-2		Control-3A		Control-3B	
Cycle #	**Exp-DB** test set RMSE (kcal/mol)	# of data points in the accumulated database	**Exp-DB** test set RMSE (kcal/mol)	# of data points in the accumulated database	**Exp-DB** test set RMSE (kcal/mol)	# of data points in the accumulated database	**Exp-DB** test set RMSE (kcal/mol)	# of data points in the accumulated database
0[a]	0.66	11,637						
1	0.59	195,069	0.64	228,488	0.68	33,290	0.68	33,290
2	0.60	260,902	0.65	286,984	0.65	41,417	0.71	41,417
3	0.60	302,462	0.65	319,845	0.70	45,757	0.68	45,757
35[b]	**0.50[b]**	**932,509[b]**						
36	0.53	936,173	0.54[c]	937,531[c]				
37	0.55	939,453	0.56[c]	941,048[c]				
38	0.56	942,686	0.53[c]	944,004[c]				

[a]The Teacher model trained using only **Exp-DB**. [b]The best-case model. [c]Control-2 scheme was applied, starting from the Student 35 model and 35 accumulated **Aug-DB**s obtained from the original SSD.

Fig. S4 shows the other Control models devised for comparing the accuracies of GNN models trained using both **CombiSolv-QM** and **QM-DB** (Control-2) or **QM-DB** only (Control-3A and Control-3B). **CombiSolv-QM** and **QM-DB** are described in Fig. 1A. In Control-2, three sets of solubility values were compared during the SSD cycles (**A**: Student-predicted, **B**: Calculated using COSMO-RS, **C**: Calculated using SMD-M06-2X/def2-TZVP). If at least one of |**A**–**B**|, |**A**–**C**|, or |**B**–**C**| is below the 0.2 kcal/mol threshold, the corresponding Student-predicted value is added to the Nth **Aug-DB** in Cycle N. Control-2 was to examine whether incorporating multiple theoretical methods can lead to more extensive databases with improved accuracies. On the other hand, Control-3A carried out the SSD using only the SMD-M06-2X/def2-TZVP solubility values (**C**) stored in **QM-DB** to investigate the feasibility of employing DFT with implicit solvation models during the SSD. The training using **QM-DB** ΔG_{solv} values without the SSD was also performed (Control-3B).

Table S8 summarizes the model accuracies and database sizes of the above three Control models and compares them with those obtained from the SSD depicted in Fig. 2A. For Control-2, three cycles were conducted, starting from the Teacher model. The number of data points in Cycle 3 is higher than that from the original SSD (319,845 vs. 302,462). All the resulting three models showed slightly lower **Exp-DB** test set RMSEs (0.64 – 0.65 kcal/mol) than that from Teacher (0.66 kcal/mol) but higher RMSEs than those from the original SSD (0.59 – 0.60 kcal/mol). We also applied the Control-2 scheme to the best-case Student 35 model and proceeded with three Control-2 cycles to seek possibilities of improving the accuracy from Student 35. More data points were added compared to the original SSD referencing only the COSMO-RS solubilities; however, no further decreases in RMSEs were observed from Student 35. The results from Control-2 indicate that referring to the single largest database, **CombiSolv-QM**, most effectively augments the database and improves the accuracy of models from the SSD.

Meanwhile, Control-3A and Control-3B were examined. In Control-3A, the sizes of **Aug-DB**s are less extensive than the SSD using **CombiSolv-QM**, and the lower accuracies were observed when **QM-DB** was employed for the SSD. Simply combining **Exp-DB** and **QM-DB** did not also improve the accuracy; the same trends were observed in the Control models (Fig. 3) where **Exp-DB** and **CombiSolv-QM** solubilities were combined without the SSD. These results demonstrate the importance of utilizing a large and comprehensive computational solubility database when conducting SSD. Although **QM-DB** did not show higher accuracies, possibly due to the less extensive database (220,332 vs. 1,000,000 for **QM-DB** vs. **CombiSolv-QM**), further SMD-DFT calculations can potentially lead to better accuracies. Moreover, the current **QM-DB** was useful in analyzing the strengths and weaknesses of COSMO-RS, SMD-DFT, and GNN models (Fig. 8 and **Error analysis of solubility prediction** Section in the Main Text). Such analysis informs the possible future work where heterogeneous data sources are adopted for developing predictive models.

S4. Results of training graph neural networks with noisy student self-distillation (NSSD)

Table S9. Comparison of accuracies for SSD and NSSD models.

Model description	Mean absolute error (kcal/mol, **Exp-DB**)			Mean absolute error (kcal/mol, **Aug-DB-1**)		
	Training	Validation	Test	Training	Validation	Test
Student 1, SSD	0.08	0.23	0.25	0.04	0.16	0.17
Student 1, 10% dropout, stochastic depth with survival probability of 90%	0.36	0.43	0.43	0.39	0.39	0.40
Student 1, 5% dropout, stochastic depth with survival probability of 20%	0.22	0.27	0.28	0.12	0.13	0.13

[a] Survival probability (p) indicates the probability of the 'final' layer. The probability decreases gradually throughout the five global update layers with the same interval, based on the given formula: $\mathbf{v}_{i+1,\text{updated}} = \mathbf{v}_i + p_i \mathbf{v}_{i+1}$ where $p_i = 1 – i[(1-p) / 4]$ $(i = 0, 1, 2, 3, 4)$, and $\mathbf{v}_i$ is the global feature vector at the i-th layer. For example, if p=0.8 (80%), The survival probabilities per each layer are: 1.0, 0.95, 0.9, 0.85, 0.8.

During the training using NSSD, noise is introduced to the model by applying dropout and stochastic depth methods to the hidden layers of the model. NSSD was effective in ML models for image classification because partially dropping the information from hidden layers would be helpful for handling the variance among different images with the same label. In this regard, we also tested multiple NSSD models in solubility predictions with different dropout rates (for all GNN layers and readout layers) and survival probabilities of stochastic depth (for the residual connection part of global feature vector). In all cases, NSSD showed higher prediction errors for **Exp-DB** than SSD (i.e., no noise was introduced to the model, Table S9). That is because dropout and stochastic depth can presumably cause errors in recognizing a molecule. The model can miss the information about key structural features related to solubility due to introducing noise to the model. In contrast, for images, if some part is lost, the model can still recognize and classify them. As a result, the SSD method was chosen throughout this study instead of NSSD for the development of self-evolving solubility databases and GNNs.

S5. Supplementary information for Application 1 – Linear Free Energy Relationships between solvation free energy and reaction rates of organic reactions

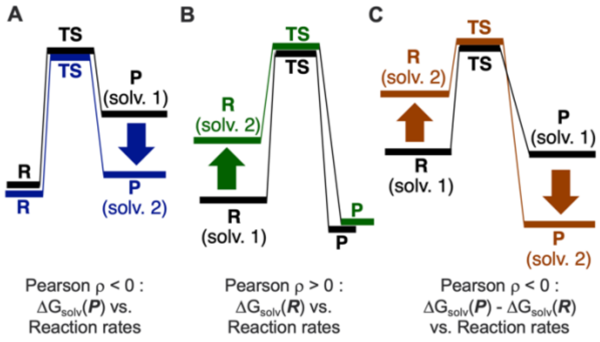


Fig. S5 Schematic energy diagrams of the reactions where (A) higher solvation stabilization of the product leads to more product formation, and (B) higher solvation destabilization of the reactant leads to faster reaction. (C) Designing new solvents can also affect both relative free energies of both reactant and product and thus the reaction rates.

Finding linear free energy relationships between solvation free energy and reaction rates based on the Hammond Postulate

In principle, a negative Pearson correlation should be found for $\Delta G_{solv}(\boldsymbol{P})$ vs. reaction rates for reactions in which the preferential stabilization of the product(s) plays a key role in lowering the activation barrier by increasing the thermodynamic driving force (Fig. S5A). Meanwhile, for some reactions, higher reaction rates can be achieved by selective destabilization of reactants by a solvent (Fig. S5B). Such a correlation leads to a positive Pearson ρ between $\Delta G_{solv}(\boldsymbol{R})$ and reaction rates. The cases in Figs. S5A and S5B mainly occur when the structure of a transition state is analogous to that of reactant(s) and product(s) of an elementary reaction, respectively, according to the Hammond Postulate. In addition, product stabilization and reactant destabilization can be considered together by using $\Delta G_{solv}(\boldsymbol{P}) - \Delta G_{solv}(\boldsymbol{R})$ as a descriptor (Fig. S5C). Changing the sign of $\Delta G_{solv}(\boldsymbol{R})$ and adding to $\Delta G_{solv}(\boldsymbol{P})$ enables the quantification of the influences on the reaction rates by both reactants and products, resulting in a negative Pearson ρ with reaction rates.

Chemical explanation of the linear relationships shown in Fig. 9

The reaction **I** is the dissociation of tert-butylperoxyaldehyde into tert-butyl alcohol and CO_2. Our results suggest that solvents with higher polarity (nitrobenzene, nitromethane, and chloroform) better stabilize the tert-butyl alcohol product than non-polar solvents (benzene, tetrachloromethane and heptane) and thus show higher rates. Lower Pearson ρ values were obtained from the other two reactions compared to **I**, but they display good negative correlations except for one solvent (MeCN and toluene for **II** and **III**, respectively). Of note, reaction **V** is analogous to **II** except for having more polar reactants than **II**. In this case, using non-polar solvents such as toluene show the most reactant destabilization and the highest reaction rate. The effect of different functional groups for the same reaction was captured by our GNN model, leading to the identification of a strong positive correlation (ρ=0.99). In contrast, high reaction rates were achieved when polar solvents such as water or ethanol were used with non-polar reactants (Br_2, pentene, and cyclopentadiene) for reactions **VI** and **VII**, respectively.

Reaction **VIII** is a ring opening to decarboxylate the reactant and form an alkene whose reaction rates were measured in five solvents. A non-polar solvent, decalin, shows the lowest reaction rate, whereas the fastest reaction was observed in a polar N-phenylforamide solvent. This is consistent with the fact that the zwitterionic product (***P***) is more polar than the reactant (***R***), so a polar solvent would be favorable to stabilize the product more than the reactant. The Cope rearrangement (**IX**), in five different solvents was also investigated. Two solvents with hydroxyl groups (ethylene glycol and phenol) showed higher reaction rates than other solvents. This is because the ketone group in the product can form hydrogen bonds with alcoholic solvents, leading to product stabilization and faster reactions.

Table S10. Pearson correlation coefficients between experimental reaction rates vs. all three ΔG_{solv} descriptors for 11 organic reactions, with the reason of choosing one descriptor.

Reaction #	**Descriptors**			**Explanation**
	$\Delta G_{solv}(\boldsymbol{P})$[a]	$\Delta G_{solv}(\boldsymbol{R})$[b]	$\Delta G_{solv}(\boldsymbol{P}) - \Delta G_{solv}(\boldsymbol{R})$[c]	
I	**-0.95**	-0.66	-0.63	Reactions can be driven by the thermodynamic stability of products and can be either single or multiple steps depending on the solvent.
II	**-0.90**	-0.95	0.99	
III	**-0.68**	0.50	-0.76	$\Delta G_{solv}(\boldsymbol{P}) - \Delta G_{solv}(\boldsymbol{R})$ is the best descriptor, but $\Delta G_{solv}(\boldsymbol{P})$ was chosen because the reactant contains nitroso group (R–N=O) which rarely appears in the database used for model training.
IV	0.94	**0.94**	0.82	Not a single-step reaction, so $\Delta G_{solv}(\boldsymbol{P}) - \Delta G_{solv}(\boldsymbol{R})$ is less reliable.
V	0.97	**0.99**	-0.95	$\Delta G_{solv}(\boldsymbol{P}) - \Delta G_{solv}(\boldsymbol{R})$ also shows a comparably strong correlation, but the reactant destabilization by solvents may be a key factor.
VI	0.91	**0.91**	-0.59	Not a single-step reaction, so $\Delta G_{solv}(\boldsymbol{P}) - \Delta G_{solv}(\boldsymbol{R})$ is less reliable.
VII	0.93	**0.80**	0.71	The reactant rapidly dimerizes even in mild conditions (room temperature), and the product is stable in most conditions except at high temperatures (> 150°C). Therefore, the reactant destabilization by solvents may be a key factor.
VIII	-0.49	-0.39	**-0.99**	Reactions where the kinetics are highly consistent with the Hammond postulate, and thus the generic $\Delta G_{solv}(\boldsymbol{P}) - \Delta G_{solv}(\boldsymbol{R})$ descriptor shows the best correlation. Also, all of them are elementary reactions.
IX	-0.61	0.09	**-0.95**	
X	-0.07	0.64	**-0.80**	
XI	-0.68	-0.59	**-0.84**	

[a] Close to -1 if product stabilization plays a key role for higher reaction rates.

[b] Close to 1 if reactant destabilization plays a key role for higher reaction rates.

[c] Close to -1 if both product stabilization and reactant destabilization play a key role for higher reaction rates.

References

1. F. H. Vermeire and W. H. Green, *Chem. Eng. J.*, 2021, **418**, 129307.