Supporting information

Prediction of biphasic separation on CO_2 absorption using molecular surface information-based machine learning model

*Taishi Kataoka, Yingquan Hao, Ying Chieh Hung, Yasuhiko Orita, Yusuke Shimoyama**

Department of Chemical Science and Engineering, Tokyo Institute of Technology, 2-12-1 S1-33, Ookayama, Meguro-ku, Tokyo 152-8550, Japan

*For correspondence: yshimo@chemeng.titech.ac.jp, TEL & FAX : +81 3 5734 3285

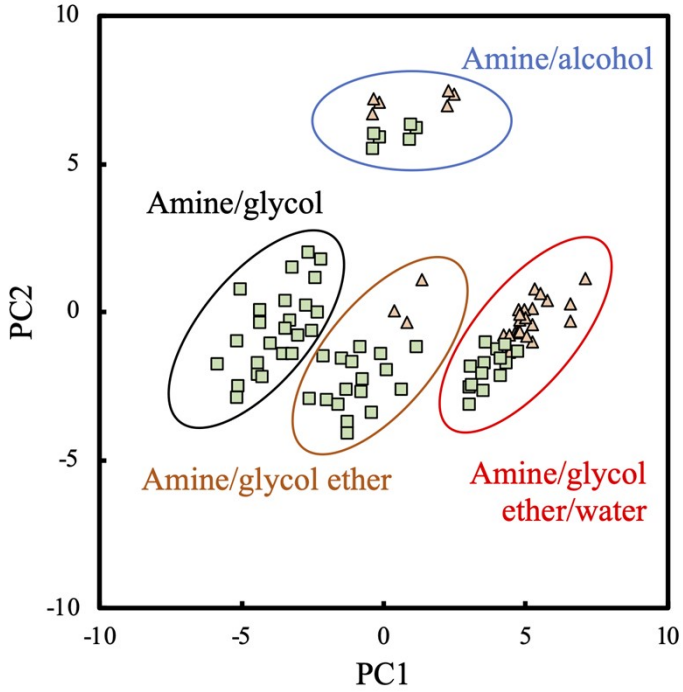

Fig. S1 Scatter plots of the first and second principal components. (■): Homogeneous phase, (△): Phase separated.

Table S1. All pairs of solvents in the dataset of mixed-solvent absorbents.

Amine	Organic solvent	Water	Type	Dataset[a]	Data source
IPMEA	DEGDEE	-	Miscible	Training	S1
BAE	DEGDEE	-	Miscible	Training	S1
BZMEA	DEGDEE	-	Miscible	Test	S1
EAE	DEGDME	-	Miscible	Test	S1
IPMEA	DEGDME	-	Miscible	Training	S1
BAE	DEGDME	-	Miscible	Training	S1
BZMEA	DEGDME	-	Miscible	Training	S1
MAE	DEGMEE	-	Miscible	Test	S1
EAE	DEGMEE	-	Miscible	Test	S1
IPMEA	DEGMEE	-	Miscible	Training	S1
BAE	DEGMEE	-	Miscible	Test	S1
BZMEA	DEGMEE	-	Miscible	Training	S1
MPA	DEGMEE	-	Miscible	Training	S1
AM2B	DEGMEE	-	Miscible	Training	S1
AM2P	DEGMEE	-	Miscible	Test	S1
MAE	DEGMME	-	Miscible	Test	S1
EAE	DEGMME	-	Miscible	Test	S1
IPMEA	DEGMME	-	Miscible	Training	S1
BAE	DEGMME	-	Miscible	Training	S1
BZMEA	DEGMME	-	Miscible	Test	S1
MPA	DEGMME	-	Miscible	Test	S1
AM2B	DEGMME	-	Miscible	Test	S1
AM2P	DEGMME	-	Miscible	Test	S1
MAE	EGBE	-	Miscible	Test	S1
EAE	EGBE	-	Miscible	Test	S1
IPMEA	EGBE	-	Miscible	Training	S1
BAE	EGBE	-	Miscible	Training	S1
BZMEA	EGBE	-	Miscible	Training	S1
MPA	EGBE	-	Miscible	Training	S1
AM2B	EGBE	-	Miscible	Test	S1
AM2P	EGBE	-	Miscible	Training	S1
MEA	Octanol	-	Biphasic	Test	S2

MEA	Heptanol	-	Biphasic	Training	S2
MEA	Isoocatanol	-	Biphasic	Test	S2
MAE	DEGDEE	-	Biphasic	Test	S1
MAE	DEGDME	-	Biphasic	Training	S1
EAE	DEGDEE	-	Biphasic	Training	S1
DEA	Octanol	-	Biphasic	Training	S2
DEA	Heptanol	-	Biphasic	Test	S2
DEA	Isoocatanol	-	Biphasic	Test	S2
BAE	DEGDEE	Contain	Miscible	Training	S3
BAE	DEGEME	Contain	Miscible	Training	S4
BAE	DEGDME	Contain	Miscible	Training	S4
AMB	DEGDEE	Contain	Miscible	Test	S3
AMB	DEGEME	Contain	Miscible	Training	S4
AMB	DEGDME	Contain	Miscible	Training	S4
AM2P	DEGDEE	Contain	Immiscible	Training	S3
MAE	DEGDEE	Contain	Immiscible	Test	S4
AEE	DEGDEE	Contain	Immiscible	Test	S3
AEE	DEGEME	Contain	Immiscible	Test	S4
MEA	DEGDEE	Contain	Immiscible	Training	S3
MEA	DEGEME	Contain	Immiscible	Test	S4
MEA	DEGDME	Contain	Immiscible	Test	S4
EAE	DEGDEE	Contain	Biphasic	Test	S3
EAE	DEGEME	Contain	Biphasic	Test	S4
EAE	DEGDME	Contain	Biphasic	Training	S4
AM2P	DEGEME	Contain	Biphasic	Training	S4
AM2P	DEGDME	Contain	Biphasic	Training	S4
MAE	DEGEME	Contain	Biphasic	Training	S4
MAE	DEGDME	Contain	Biphasic	Test	S4
AEE	DEGDME	Contain	Biphasic	Test	S4

[a] Classified into training or test sets by DS-known

Table S2. Ranges of hyperparameters in each model

Model	Parameters[a]	Range
RF	n_estimaters	100
	max_depth	1–7
	min_samples_split	3–20
LGR	*C*	0.125–1024
	penalty	L2 norm
SVM	*C*	0.125–1024
	penalty	L2 norm

[a] Name of parameters in scikit-learn library. n_estimaters is number of trees in forest, max_depth is maximum depth of trees, min_samples_split is minimum number of samples required to split an internal node, *C* is regularization parameter, and penalty is regularization term.

Table S3. Confusion matrix that summarizes prediction results for phase states by (a) RF, (b) LGR, and (c) SVM models using DS-unknown.

(a) RF

		Predicted class	
		Miscible	Immiscible
Actual class	Miscible	89	2
	Immiscible	8	23

(b) LGR

		Predicted class	
		Miscible	Immiscible
Actual class	Miscible	87	4
	Immiscible	3	28

(c) SVM

		Predicted class	
		Miscible	Immiscible
Actual class	Miscible	85	6
	Immiscible	3	28

Table S4. Confusion matrix that summarizes prediction results for phase states in test sets by (a) RF, (b) LGR, and (c) SVM models using DS-known.

(a) RF

		Predicted class	
		Miscible	Immiscible
Actual class	Miscible	40	1
	Immiscible	1	18

(b) LGR

		Predicted class	
		Miscible	Immiscible
Actual class	Miscible	40	1
	Immiscible	1	18

(c) SVM

		Predicted class	
		Miscible	Immiscible
Actual class	Miscible	36	5
	Immiscible	0	19

Table S5. Experimental and predicted phase behaviors[a].

Amine	Organic solvent	Water	Prediction for DS-known[b]	Prediction for DS-unknown[b]
IPMEA	DEGDEE	-	Miscible	Miscible
BAE	DEGDEE	-	Miscible	Miscible
BZMEA	DEGDEE	-	Miscible	Miscible
EAE	DEGDME	-	Miscible	Miscible
IPMEA	DEGDME	-	Miscible	Miscible
BAE	DEGDME	-	Miscible	Miscible
BZMEA	DEGDME	-	Miscible	Miscible
MAE	DEGMEE	-	Miscible	**Biphasic** *(Miscible)*
EAE	DEGMEE	-	Miscible	Miscible
IPMEA	DEGMEE	-	Miscible	Miscible
BAE	DEGMEE	-	Miscible	Miscible
BZMEA	DEGMEE	-	Miscible	Miscible
MPA	DEGMEE	-	Miscible	Miscible
AM2B	DEGMEE	-	Miscible	Miscible
AM2P	DEGMEE	-	Miscible	Miscible
MAE	DEGMME	-	Miscible	Miscible
EAE	DEGMME	-	Miscible	Miscible
IPMEA	DEGMME	-	Miscible	Miscible
BAE	DEGMME	-	Miscible	Miscible
BZMEA	DEGMME	-	Miscible	Miscible
MPA	DEGMME	-	Miscible	Miscible
AM2B	DEGMME	-	Miscible	Miscible
AM2P	DEGMME	-	Miscible	Miscible
MAE	EGBE	-	Miscible	**Biphasic** *(Miscible)*
EAE	EGBE	-	Miscible	Miscible
IPMEA	EGBE	-	Miscible	Miscible
BAE	EGBE	-	Miscible	Miscible
BZMEA	EGBE	-	Miscible	Miscible
MPA	EGBE	-	Miscible	Miscible
AM2B	EGBE	-	Miscible	Miscible
AM2P	EGBE	-	Miscible	Miscible
MEA	Octanol	-	Biphasic	Biphasic

MEA	Heptanol	-	Biphasic	Biphasic
MEA	Isoocatanol	-	Biphasic	Biphasic
MAE	DEGDEE	-	Biphasic	Biphasic
MAE	DEGDME	-	Biphasic	Biphasic
EAE	DEGDEE	-	Biphasic	Biphasic
DEA	Octanol	-	Biphasic	Biphasic
DEA	Heptanol	-	Biphasic	Biphasic
DEA	Isoocatanol	-	Biphasic	Biphasic
BAE	DEGDEE	Contain	Miscible	Miscible
BAE	DEGEME	Contain	Miscible	Miscible
BAE	DEGDME	Contain	Miscible	Miscible
AMB	DEGDEE	Contain	Miscible	Miscible
AMB	DEGEME	Contain	Miscible	Miscible
AMB	DEGDME	Contain	Miscible	Miscible
AM2P	DEGDEE	Contain	Immiscible	**Biphasic** *(Immiscible)*
MAE	DEGDEE	Contain	**Biphasic** *(Immiscible)*	**Biphasic** *(Immiscible)*
AEE	DEGDEE	Contain	Immiscible	Immiscible
AEE	DEGEME	Contain	Immiscible	Immiscible
MEA	DEGDEE	Contain	Immiscible	Immiscible
MEA	DEGEME	Contain	Immiscible	Immiscible
MEA	DEGDME	Contain	Immiscible	Immiscible
EAE	DEGDEE	Contain	Biphasic	Biphasic
EAE	DEGEME	Contain	Biphasic	**Miscible** *(Biphasic)*
EAE	DEGDME	Contain	Immiscible	Immiscible
AM2P	DEGEME	Contain	Immiscible	Immiscible
AM2P	DEGDME	Contain	Immiscible	Immiscible
MAE	DEGEME	Contain	Immiscible	Immiscible
MAE	DEGDME	Contain	Immiscible	Immiscible
AEE	DEGDME	Contain	**Immiscible** *(Biphasic)*	**Immiscible** *(Biphasic)*

Experimental data source as well as those in Table S1

[a] Predicted phase behaviors by LGR model.

[b] Bold letters: Prediction result was wrong, brackets: actual phase behavior (correct answer)

REFERENCES

(S1) Barzagli, F.; Mani, F.; Peruzzini, M. Novel Water-Free Biphasic Absorbents for Efficient CO_2 Capture. *Int. J. Greenh. Gas Control.* **2017**, 60, 100–109. https://doi.org/10.1016/j.ijggc.2017.03.010.

(S2) Kim, Y. E.; Park, J. H.; Yun, S. H.; Nam, S. C.; Jeong, S. K.; Yoon, Y. Il. Carbon Dioxide Absorption Using a Phase Transitional Alkanolamine-Alcohol Mixture. *J. Ind. Eng. Chem.* **2014**, 20 (4), 1486–1492. https://doi.org/10.1016/j.jiec.2013.07.036.

(S3) Machida, H.; Oba, K.; Tomikawa, T.; Esaki, T.; Yamaguchi, T.; Horizoe, H. Development of Phase Separation Solvent for CO_2 Capture by Aqueous (Amine + Ether) Solution. *J. Chem. Thermodyn.* **2017**, 113, 64–70. https://doi.org/10.1016/j.jct.2017.05.043.

(S4) Nakaoka, M.; Tran, K. V. B.; Yanase, K.; MacHida, H.; Norinaga, K. Prediction of Phase Behavior of CO_2 Absorbents Using Conductor-like Screening Model for Real Solvents (COSMO-RS): An Approach to Identify Phase Separation Solvents of Amine/Ether/Water Systems upon CO_2 Absorption. *Ind. Eng. Chem. Res.* **2020**, 59 (42), 19020–19029. https://doi.org/10.1021/acs.iecr.0c03233.