Supporting Information

Learning Atomic Interactions through Solvation Free Energy Prediction Using Graph Neural Network

Yashaswi Pathak, Sarvesh Mehta, and U. Deva Priyakumar*

Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500 032, India.

E-mail: deva@iiit.ac.in

Phone: 0091 40 6653 1161. Fax: 0091 40 6653 1413

Accuracy on Minnesota Solvation Database (MNSOL)

There are classical and quantum-mechanical simulation studies that use the MNSOL as the reference database. We choose solvation model based on density(SMD) for comparison with our CIGIN Model. We also compared CIGIN Model with semi-empirical methods, pure COSMO, and COSMO-RS.

Table S1: Comparison of the performances of the CIGIN (current study), DELFOS[1] on MNSOL dataset[2] using root mean square error (RMSE). The RMSE was calculated as an average on 9 independednt 10 fold cross validatin runs for both the methods

Method	N_{data}	RMSE
COSMO/BP86/TZVP[3]	2436	2.57
COSMO-RS/BP86/TZVP[3]	2346	0.75
SMD/PM3[4]	2500	4.8
SMD/PM6[4]	2500	3.6
DELFOS	2495	0.57
CIGIN	2525	**0.55**

Model Architecture

Grid search strategy was employed to tune the hyper parameters of the model. Table 2 and 3 shows the architecture of the model in the prediction phase and the value of various parameters obtained after hyper parameter tuning respectively.

Table S2: Architecture of model in prediction phase obtained after grid search

Layer	No. of Units	Activation Function	Layer Position
Set2Set[5] for Solute	90	-	1st
Set2Set[5] for Solvent	90	-	1st
Linear	360	ReLU	2nd
Linear	256	ReLU	3rd
Linear	128	None	4th

Table S3: Value of different hyperparameters obtained after grid search

Parameter	Value	Phase
T	6	Message Passing Phase
Interaction function f	$tanh(A.B)$	Interaction Phase

Transferability of CIGIN Model

To show that our model shows high accuracy on pairs of solute and solvent that are not in the Solv@TUM and FreeSolv datasets, we selected data points from the MNSOL dataset that were not present in the Solv@TUM dataset. The mean absolute error obtained on this

set containing 1860 distinct pairs of solute and solvent was 0.40 kcal/mol. This shows that the CIGIN model doesn't over fir and accurately predicts unseen pair of solute and solvent.

Solvent Holdout Test

Table S4: Solvent Holdout Test

Solvent	MAE (kcal/mol)
ISOPROPYLACETATE	0.001
4-HEPTANOL	0.0013
ETHYLBUTANOATE	0.0017
PENTYLACETATE	0.0022
ETHYLHEXANOATE	0.0026
4-OCTANOL	0.0051
3-HEXANOL	0.0054
3-PENTANONE	0.012
UNDECANOL	0.0145
2-HEXANOL	0.0145
4-METHYLPENTAN-2-OL	0.019
PENTAN-2-ONE	0.0258
PROPYLACETATE	0.0261
HEPTYLACETAT	0.0267
METHOXYPROPANE	0.0289
ISOBUTYLACETATE	0.0323
NONANOL	0.033
2-METHYLBUTAN-1-OL	0.0338
HEPTAN-1-OL	0.0348
2-METHOXYPROPANE	0.0461

PROPIONITRILE	0.0469
5,8,11,14-TETRAOXAOCTADECANE	0.0474
ETBE	0.0513
M-XYLENE	0.0522
3-HEXANONE	0.0528
HEXYLACETATE	0.0531
BUTYROLACTONE	0.0538
2-PENTANOL	0.0539
NONANE	0.0547
2-PROPOXYETHANOL	0.0557
1-METHOXYBUTANE	0.0585
HEXANOL	0.0592
1-ETHOXYBUTANE	0.0594
PENTANE	0.0598
1-ETHYL-2-PYRROLIDINONE	0.0608
TETRADECANE	0.0621
PENTAN-3-OL	0.0627
3-METHYLBUTAN-1-OL	0.0633
BENZONITRILE	0.0645
TETRAHYDROPYRAN	0.0654
4-METHYLPENTAN-2-ONE	0.0658
1,5-DIMETHYL-2-PYRROLIDINONE	0.0667
TERT-AMYLMETHYLETHER	0.0718
TERT-AMYLALCOHOL	0.0728
1-METHYL-2-PIPERIDINONE	0.0782
1-OCTANOL	0.0783
N-BUTANOL	0.0793

O-XYLENE	0.0805
PENTANOL	0.0843
OCTANE	0.0859
1-ETHOXYPROPANE	0.087
DIETHYLENEGLYCOLDIBUTYLETHER	0.0877
BUTYRONITRILE	0.0886
N,N-DIETHYLACETAMIDE	0.0899
PROPYLETHER	0.0938
2-ISOPROPOXYETHANOL	0.0942
ANISOLE	0.0942
N-METHYLFORMAMIDE	0.0952
ACETONITRILE	0.0968
P-XYLENE	0.0973
N-ETHYLFORMAMIDE	0.0977
CYCLOOCTANE	0.0977
HEXANE	0.0982
CYCLOHEXANONE	0.0986
ISOBUTANOL	0.0994
2-HEXANONE	0.0999
TOLUENE	0.1007
2-HEPTANOL	0.1017
HEPTANE	0.1031
2-ETHOXYETHANOL	0.1032
DECANE	0.1032
ETHYLBENZENE	0.1089
DIBUTYLETHER	0.1092
N,N-DIBUTYLFORMAMID	0.1135

METHYLACETATE	0.1136
1-HEXADECENE	0.1148
ETHYLBENZOATE	0.1149
2-HEPTANONE	0.1153
BROMOETHANE	0.1164
EPSILON-CAPROLACTONE	0.119
ETHYLACETATE	0.1229
BUTANONE	0.1255
NITROETHANE	0.1263
N,N-DIMETHYLACETAMIDE	0.1276
DIETHYLETHER	0.1334
2-PICOLINE	0.1351
1-PROPANOL	0.1356
ALLYLALCOHOL	0.1362
2-BUTOXYETHANOL	0.1364
DECAN-1-OL	0.1407
METHOXYETHANOL	0.1431
METHYLCYCLOHEXANE	0.1459
FLUOROBENZENE	0.1477
DIISOPROPYLETHER	0.1477
N-ETHYLACETAMIDE	0.1479
N-METHYLACETAMIDE	0.1483
ACETONE	0.1488
DODECANE	0.1529
BROMOBENZENE	0.1562
1-CHLOROBUTANE	0.1589
TRIBUTYLPHOSPHATE	0.161

ISOPROPANOL	0.1618
4-FORMYLMORPHOLINE	0.1718
PROPYLENECARBONATE	0.173
BUTYLACETATE	0.1737
PERFLUOROBENZENE	0.1761
2,2,4-TRIMETHYLPENTANE	0.1768
NITROBENZENE	0.1776
N-METHYLPYRROLIDONE	0.1783
METHYLTERT-BUTYLETHER	0.1795
TERT-BUTANOL	0.1834
TETRAETHYLENEGLYCOLDIMETHYLETHER	0.1842
BENZYLALCOHOL	0.1844
CHLOROBENZENE	0.1869
TRIETHYLAMINE	0.1878
HEXADECANE	0.1889
ETHANOL	0.1908
IODOBENZENE	0.192
1,4-DIOXANE	0.2094
DICHLOROMETHANE	0.2142
PYRIDINE	0.2307
ACETICACID	0.2354
DIBENZYLETHER	0.2393
SULFOLANE	0.2393
TETRAHYDROFURAN	0.2415
PHENYLAMINE	0.243
DICHLOROETHANE	0.2486
NITROMETHANE	0.2631

DIMETHYLFORMAMIDE	0.2655
UNDECANE	0.2687
ACETOPHENONE	0.3214
METHANOL	0.336
FORMAMIDE	0.4087
ETHYLPHENYLETHER	0.4175
DIETHYLDIGLYCOL	0.4546
DIIODOMETHANE	0.4969
CARBONTETRACHLORIDE	0.5469
CHLOROFORM	0.5881
ETHYLENEGLYCOL	0.7186
M-CRESOL	0.7716
CARBONDISULFIDE	0.8344
DIMETHYLSULFOXIDE	0.9416
WATER	3.0439

References

(1) Lim, H.; Jung, Y. Delfos: Deep Learning Model for Prediction of Solvation Free Energies in Generic Organic Solvents. *Chem. Sci.* **2019**, *10*, 8306–8315.

(2) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. Minnesota Solvation Database (MNSOL) Version 2012. Retrieved from the Data Repository for the University of Minnesota, `https://doi.org/10.13020/3eks-j059`, (accessed on June 7, 2019).

(3) Klamt, A.; Diedenhofen, M. Calculation of Solvation Free Energies with DCOSMO-RS. *J. Phys. Chem. A* **2015**, *119*, 5439–5445.

(4) Kromann, J. C.; Steinmann, C.; Jensen, J. H. Improving Solvation Energy Predictions using the SMD Solvation Method and Semiempirical Electronic Structure Methods. *J. Chem. Phys.* **2018**, *149*, 104102.

(5) Vinyals, O.; Bengio, S.; Kudlur, M. Order matters: Sequence to Sequence for Sets. *arXiv preprint arXiv:1511.06391* **2015**,