Supporting Information for: Predicting Thermodynamic Properties of Alkanes by High-throughput Force Field Simulation and Machine Learning

*Zheng Gong, Yanze Wu, Liang Wu and Huai Sun**

School of Chemistry and Chemical Engineering, Materials Genome Initiative Center, and Key Laboratory of Scientific and Engineering Computing of Ministry of Education, Shanghai Jiao Tong University, Shanghai, China 200240

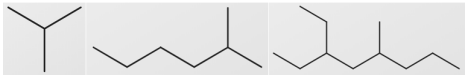
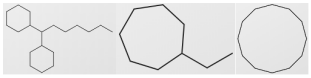
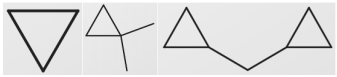
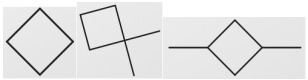
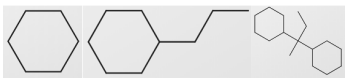
*To whom the correspondence should be addressed: huaisun@sjtu.edu.cn

1. Details of molecules

The details of 876 molecules are listed in the attached Excel (.xlsx) file. The formula, CAS number, name, melting point (T_{m}), normal boiling point (T_{vap}) and critical point (T_{c}) are taken from Yaws' Critical Property Data for Chemical Engineers and Chemists[1]. The SMILES string are obtained by using NCBI PubChem PUG API[2]. For molecule whose melting point is not available, the lowest temperature for simulation is set to $0.4 * T_{\mathrm{vap}} + 100\mathrm{K}$. For those moleucules whose critical point is not available, the highest temperature is set to $T_{\mathrm{vap}} + 75\mathrm{K}$ for liquid simulation or $T_{\mathrm{vap}} + 100\mathrm{K}$ for VLE simulation.

2. Categories of molecules

Table S1 Number of different type of alkanes and representative structures. Note that one molecule can belong to several categories

Category	Number	Representative structures
normal alkane	19	
iso-alkane	743	
neo-alkane	410	
cyclic alkane	186	
cyclopropane	14	
cyclobutane	7	
cyclopentane	62	
cyclohexane	93	
bicyclo-alkane	13	

3. Convergence of vacuum simulation for dodecane at 288 K

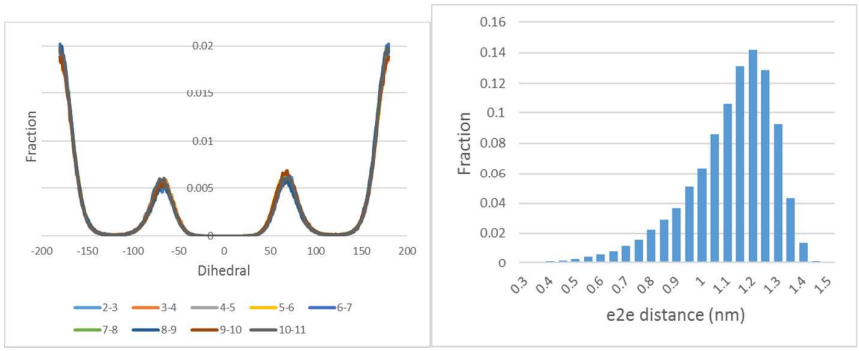


Figure S1 Distributions of nine C-C-C-C dihedral angles ('2-3' denotes the dihedral formed by C1-C2-C3-C4), and distribution of the end-to-end distances for dodecane at 288 K in vacuum simulation.

4. Distributions of kinetic energies of VLE simulation

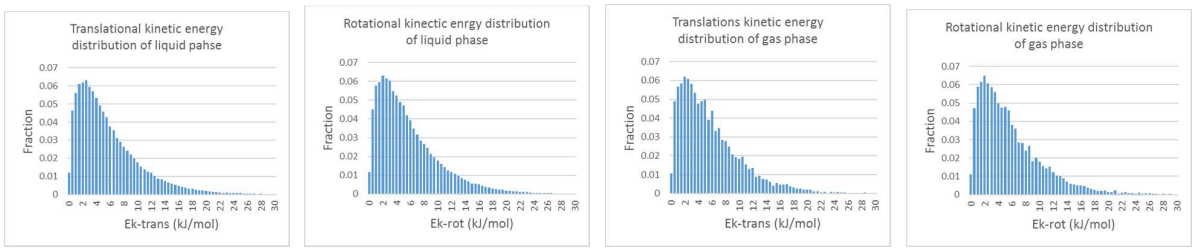
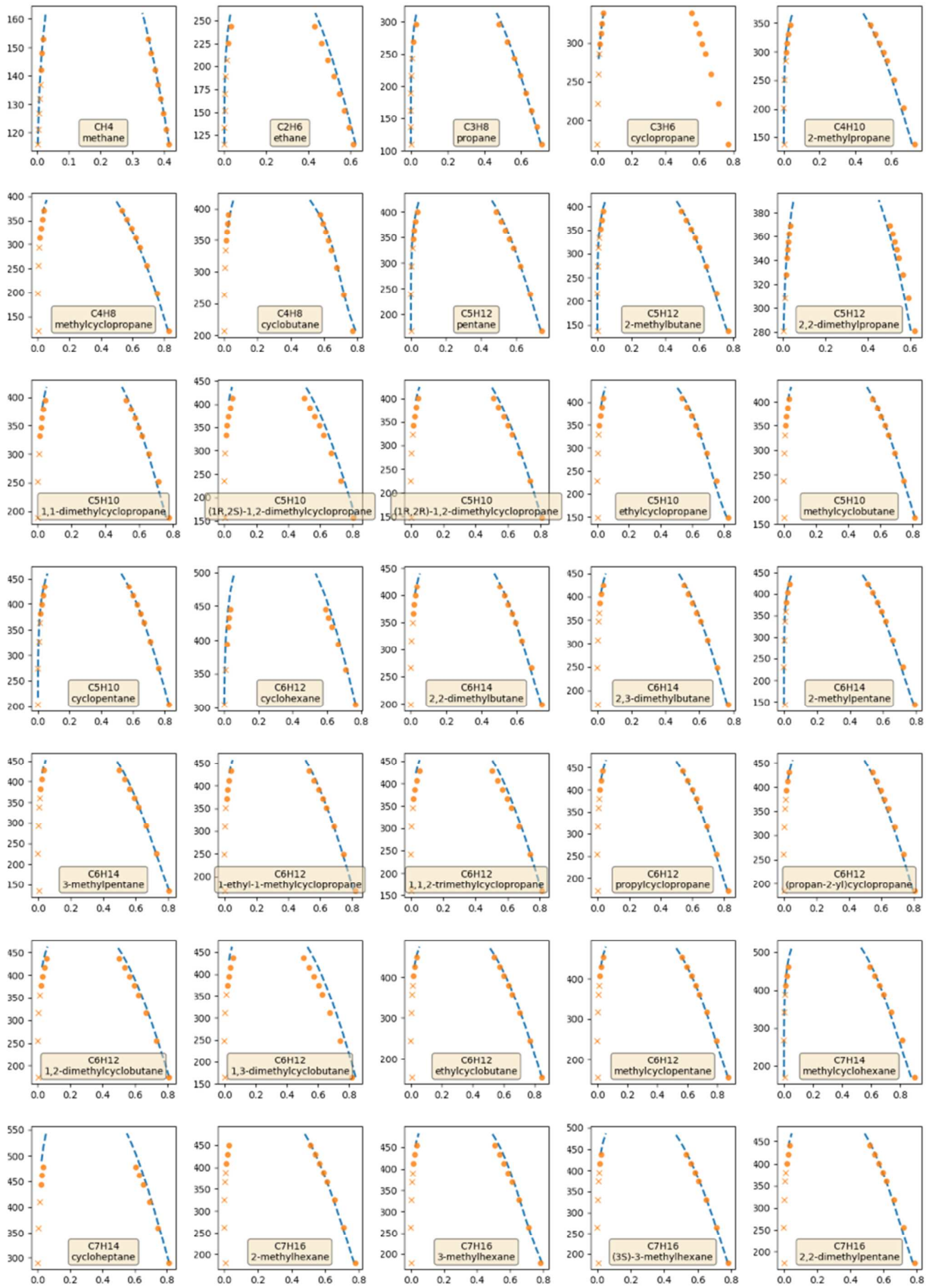
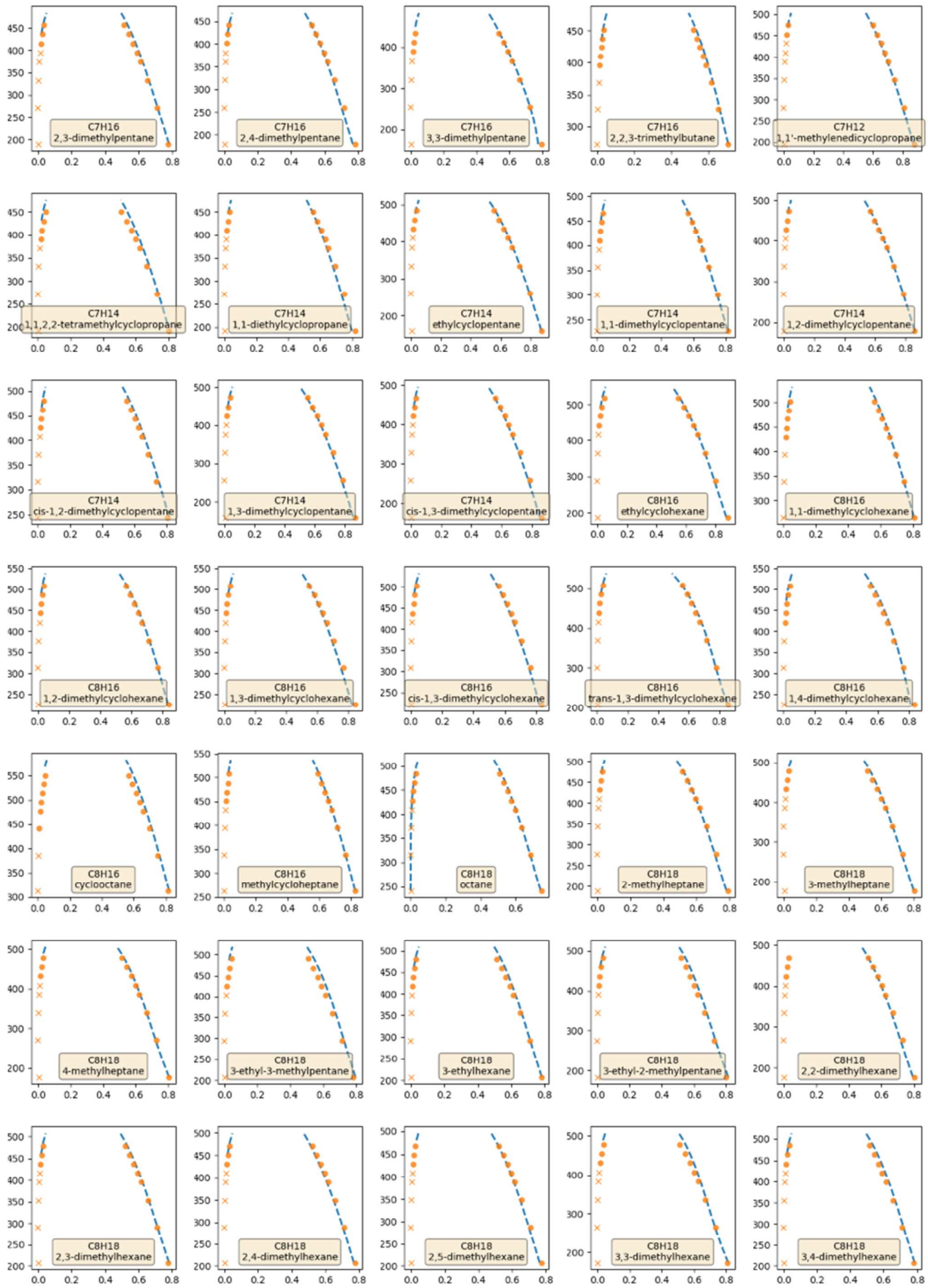


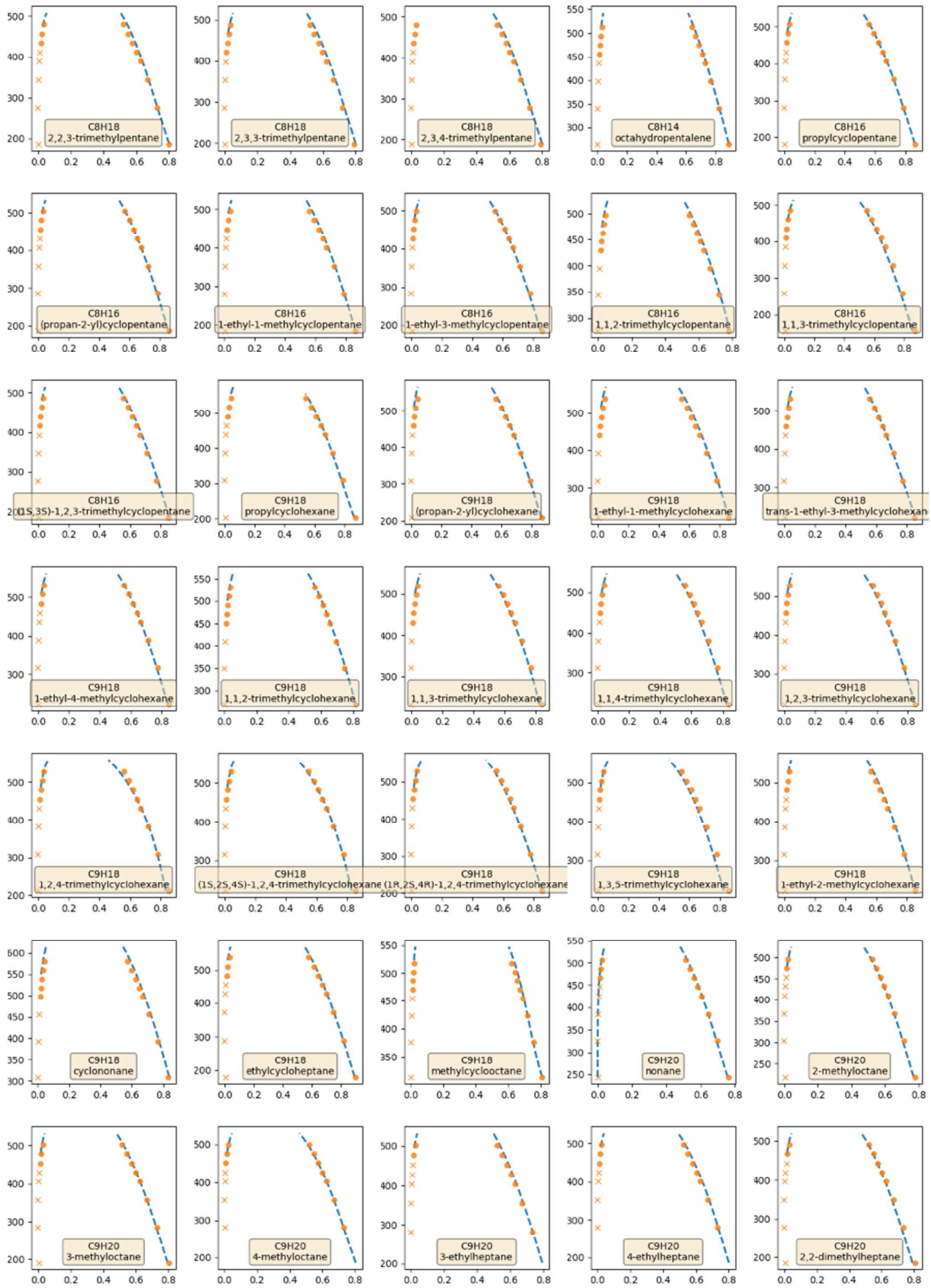
Figure S2 Distributions of translational and rotational kinetic energy of molecules in liquid phase and gas phase in VLE simulation for decane at 466 K.

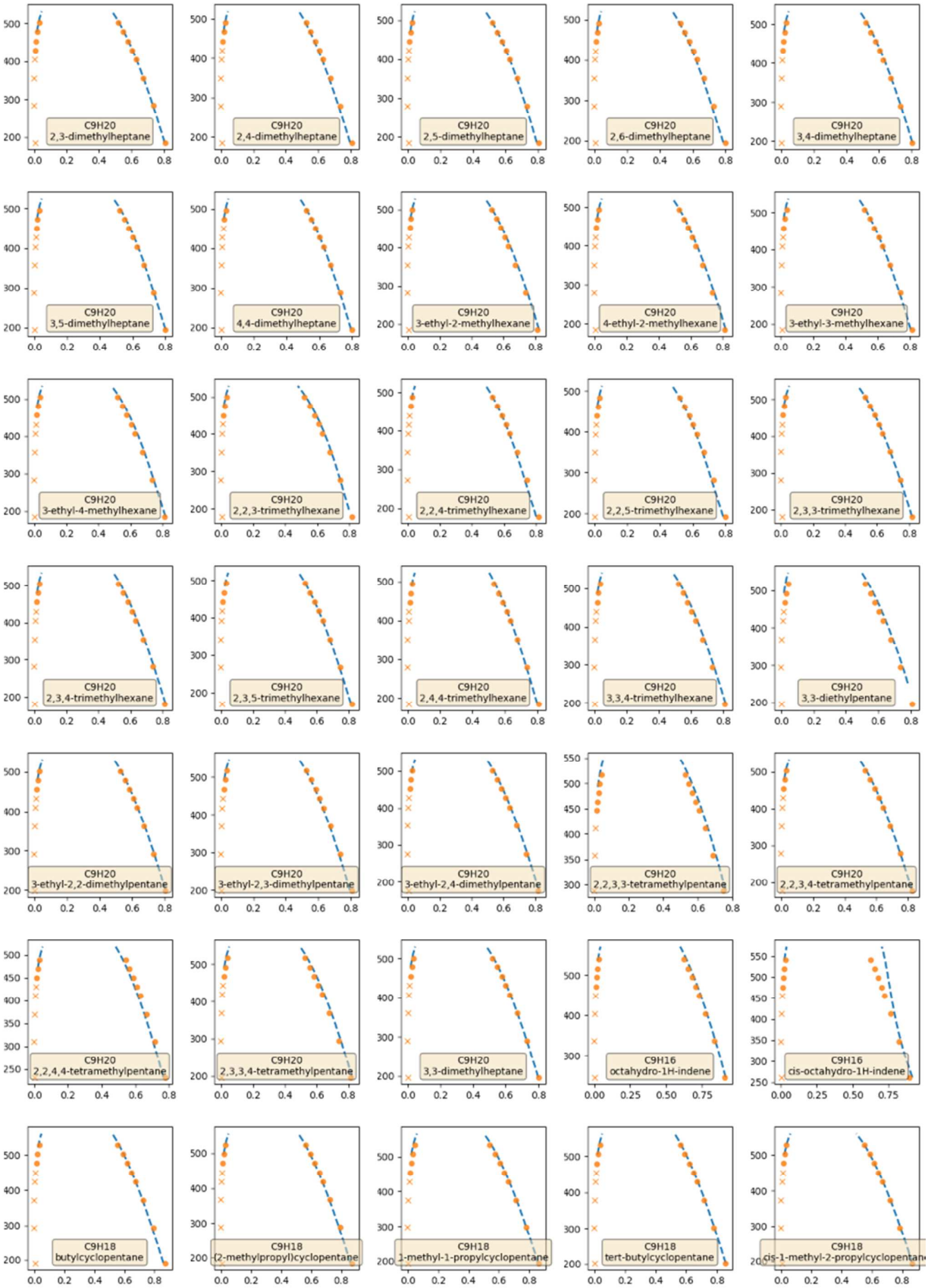
5. Vapor-Liquid-Equilibria Phase Diagrams and Clausius-Clapeyron plots

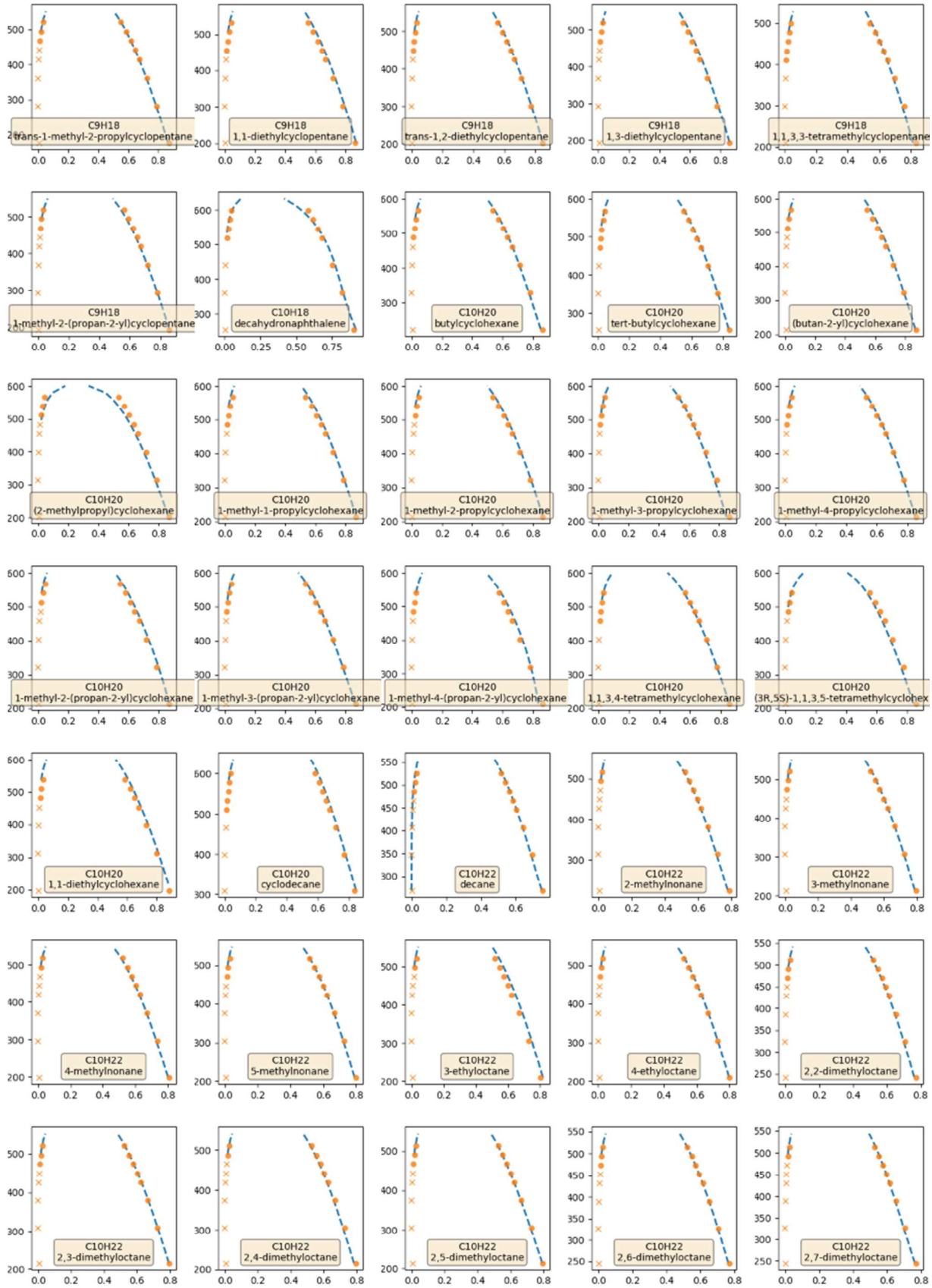
The VLE phase diagrams of 240 molecules are compared against experimental data taken from NIST Standard Reference Database 103b[3-5], As shown in Figure S3. The Clausius-Clapeyron plots are given in Figure S4.

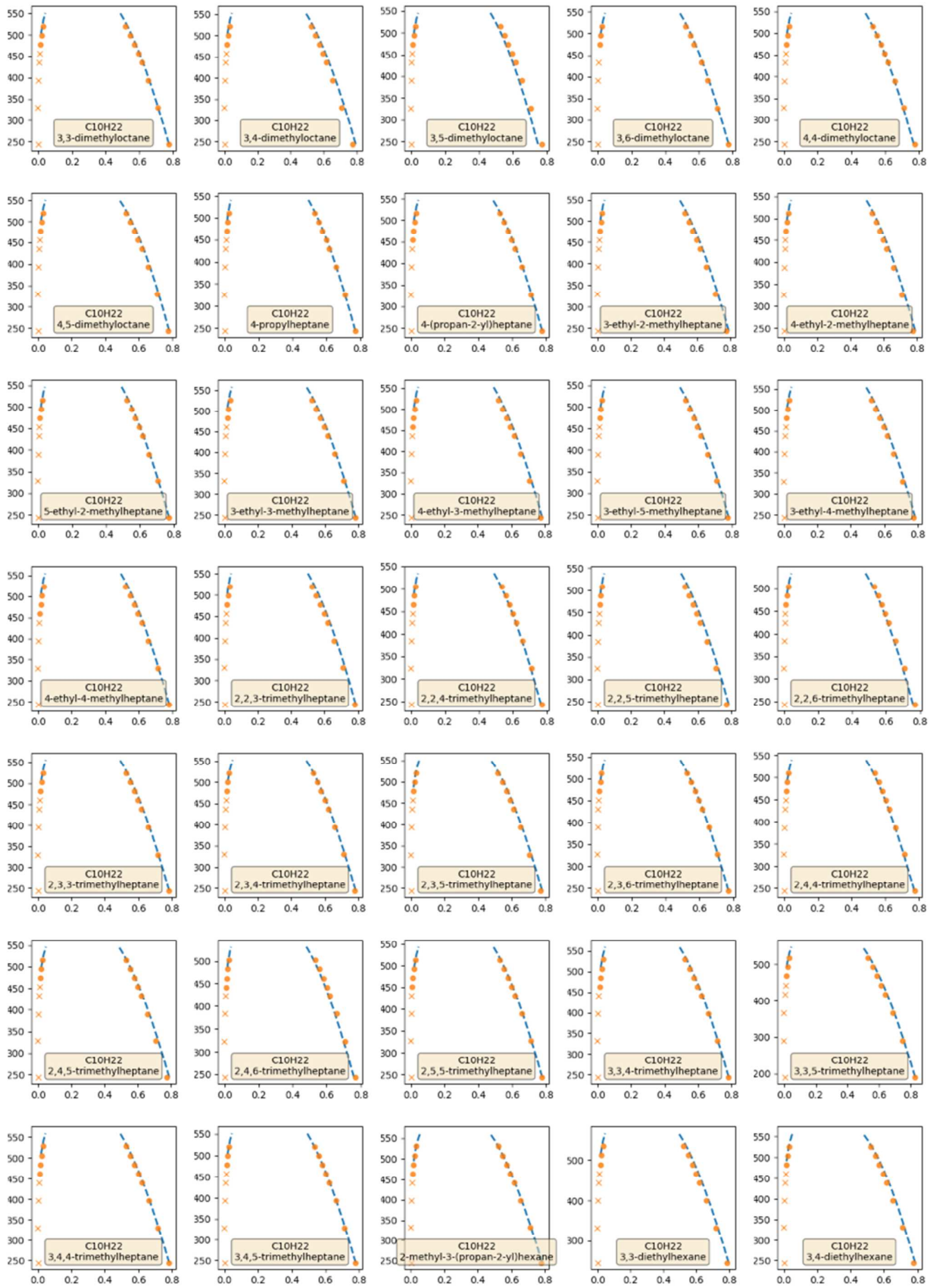

CH4
methane
C2H6
ethane
C3H8
propane
C3H6
cyclopropane
C4H10
2-methylpropane
C4H8
methylcyclopropane
C4H8
cyclobutane
C5H12
pentane
C5H12
2-methylbutane
C5H12
2,2-dimethylpropane
C5H10
1,1-dimethylcyclopropane
C5H10
(1R,2S)-1,2-dimethylcyclopropane
C5H10
(1R,2R)-1,2-dimethylcyclopropane
C5H10
ethylcyclopropane
C5H10
methylcyclobutane
C5H10
cyclopentane
C6H12
cyclohexane
C6H14
2,2-dimethylbutane
C6H14
2,3-dimethylbutane
C6H14
2-methylpentane
C6H14
3-methylpentane
C6H12
1-ethyl-1-methylcyclopropane
C6H12
1,1,2-trimethylcyclopropane
C6H12
propylcyclopropane
C6H12
(propan-2-yl)cyclopropane
C6H12
1,2-dimethylcyclobutane
C6H12
1,3-dimethylcyclobutane
C6H12
ethylcyclobutane
C6H12
methylcyclopentane
C7H14
methylcyclohexane
C7H14
cycloheptane
C7H16
2-methylhexane
C7H16
3-methylhexane
C7H16
(3S)-3-methylhexane
C7H16
2,2-dimethylpentane


C7H16
2,3-dimethylpentane
C7H16
2,4-dimethylpentane
C7H16
3,3-dimethylpentane
C7H16
2,2,3-trimethylbutane
C7H12
1,1'-methylenedicyclopropane
C7H14
1,1,2,2-tetramethylcyclopropane
C7H14
1,1-diethylcyclopropane
C7H14
ethylcyclopentane
C7H14
1,1-dimethylcyclopentane
C7H14
1,2-dimethylcyclopentane
C7H14
cis-1,2-dimethylcyclopentane
C7H14
1,3-dimethylcyclopentane
C7H14
cis-1,3-dimethylcyclopentane
C8H16
ethylcyclohexane
C8H16
1,1-dimethylcyclohexane
C8H16
1,2-dimethylcyclohexane
C8H16
1,3-dimethylcyclohexane
C8H16
cis-1,3-dimethylcyclohexane
C8H16
trans-1,3-dimethylcyclohexane
C8H16
1,4-dimethylcyclohexane
C8H16
cyclooctane
C8H16
methylcycloheptane
C8H18
octane
C8H18
2-methylheptane
C8H18
3-methylheptane
C8H18
4-methylheptane
C8H18
3-ethyl-3-methylpentane
C8H18
3-ethylhexane
C8H18
3-ethyl-2-methylpentane
C8H18
2,2-dimethylhexane
C8H18
2,3-dimethylhexane
C8H18
2,4-dimethylhexane
C8H18
2,5-dimethylhexane
C8H18
3,3-dimethylhexane
C8H18
3,4-dimethylhexane


C8H18
2,2,3-trimethylpentane
C8H18
2,3,3-trimethylpentane
C8H18
2,3,4-trimethylpentane
C8H14
octahydropentalene
C8H16
propylcyclopentane
C8H16
(propan-2-yl)cyclopentane
C8H16
1-ethyl-1-methylcyclopentane
C8H16
1-ethyl-3-methylcyclopentane
C8H16
1,1,2-trimethylcyclopentane
C8H16
1,1,3-trimethylcyclopentane
C8H16
(1S,3S)-1,2,3-trimethylcyclopentane
C9H18
propylcyclohexane
C9H18
(propan-2-yl)cyclohexane
C9H18
1-ethyl-1-methylcyclohexane
C9H18
trans-1-ethyl-3-methylcyclohexan
C9H18
1-ethyl-4-methylcyclohexane
C9H18
1,1,2-trimethylcyclohexane
C9H18
1,1,3-trimethylcyclohexane
C9H18
1,1,4-trimethylcyclohexane
C9H18
1,2,3-trimethylcyclohexane
C9H18
1,2,4-trimethylcyclohexane
C9H18
(1S,2S,4S)-1,2,4-trimethylcyclohexane (1R,2S,4R)-1,2,4-trimethylcyclohexane
C9H18
1,3,5-trimethylcyclohexane
C9H18
1-ethyl-2-methylcyclohexane
C9H18
cyclononane
C9H18
ethylcycloheptane
C9H18
methylcyclooctane
C9H20
nonane
C9H20
2-methyloctane
C9H20
3-methyloctane
C9H20
4-methyloctane
C9H20
3-ethylheptane
C9H20
4-ethylheptane
C9H20
2,2-dimethylheptane


C9H20
2,3-dimethylheptane
C9H20
2,4-dimethylheptane
C9H20
2,5-dimethylheptane
C9H20
2,6-dimethylheptane
C9H20
3,4-dimethylheptane
C9H20
3,5-dimethylheptane
C9H20
4,4-dimethylheptane
C9H20
3-ethyl-2-methylhexane
C9H20
4-ethyl-2-methylhexane
C9H20
3-ethyl-3-methylhexane
C9H20
3-ethyl-4-methylhexane
C9H20
2,2,3-trimethylhexane
C9H20
2,2,4-trimethylhexane
C9H20
2,2,5-trimethylhexane
C9H20
2,3,3-trimethylhexane
C9H20
2,3,4-trimethylhexane
C9H20
2,3,5-trimethylhexane
C9H20
2,4,4-trimethylhexane
C9H20
3,3,4-trimethylhexane
C9H20
3,3-diethylpentane
C9H20
3-ethyl-2,2-dimethylpentane
C9H20
3-ethyl-2,3-dimethylpentane
C9H20
3-ethyl-2,4-dimethylpentane
C9H20
2,2,3,3-tetramethylpentane
C9H20
2,2,3,4-tetramethylpentane
C9H20
2,2,4,4-tetramethylpentane
C9H20
2,3,3,4-tetramethylpentane
C9H20
3,3-dimethylheptane
C9H16
octahydro-1H-indene
C9H16
cis-octahydro-1H-indene
C9H18
butylcyclopentane
C9H18
(2-methylpropyl)cyclopentane
C9H18
1-methyl-1-propylcyclopentane
C9H18
tert-butylcyclopentane
C9H18
cis-1-methyl-2-propylcyclopentane


C9H18
trans-1-methyl-2-propylcyclopentane
C9H18
1,1-diethylcyclopentane
C9H18
trans-1,2-diethylcyclopentane
C9H18
1,3-diethylcyclopentane
C9H18
1,1,3,3-tetramethylcyclopentane
C9H18
1-methyl-2-(propan-2-yl)cyclopentane
C10H18
decahydronaphthalene
C10H20
butylcyclohexane
C10H20
tert-butylcyclohexane
C10H20
(butan-2-yl)cyclohexane
C10H20
(2-methylpropyl)cyclohexane
C10H20
1-methyl-1-propylcyclohexane
C10H20
1-methyl-2-propylcyclohexane
C10H20
1-methyl-3-propylcyclohexane
C10H20
1-methyl-4-propylcyclohexane
C10H20
1-methyl-2-(propan-2-yl)cyclohexane
C10H20
1-methyl-3-(propan-2-yl)cyclohexane
C10H20
1-methyl-4-(propan-2-yl)cyclohexane
C10H20
1,1,3,4-tetramethylcyclohexane
C10H20
(3R,5S)-1,1,3,5-tetramethylcyclohex
C10H20
1,1-diethylcyclohexane
C10H20
cyclodecane
C10H22
decane
C10H22
2-methylnonane
C10H22
3-methylnonane
C10H22
4-methylnonane
C10H22
5-methylnonane
C10H22
3-ethyloctane
C10H22
4-ethyloctane
C10H22
2,2-dimethyloctane
C10H22
2,3-dimethyloctane
C10H22
2,4-dimethyloctane
C10H22
2,5-dimethyloctane
C10H22
2,6-dimethyloctane
C10H22
2,7-dimethyloctane


C10H22
3,3-dimethyloctane
C10H22
3,4-dimethyloctane
C10H22
3,5-dimethyloctane
C10H22
3,6-dimethyloctane
C10H22
4,4-dimethyloctane
C10H22
4,5-dimethyloctane
C10H22
4-propylheptane
C10H22
4-(propan-2-yl)heptane
C10H22
3-ethyl-2-methylheptane
C10H22
4-ethyl-2-methylheptane
C10H22
5-ethyl-2-methylheptane
C10H22
3-ethyl-3-methylheptane
C10H22
4-ethyl-3-methylheptane
C10H22
3-ethyl-5-methylheptane
C10H22
3-ethyl-4-methylheptane
C10H22
4-ethyl-4-methylheptane
C10H22
2,2,3-trimethylheptane
C10H22
2,2,4-trimethylheptane
C10H22
2,2,5-trimethylheptane
C10H22
2,2,6-trimethylheptane
C10H22
2,3,3-trimethylheptane
C10H22
2,3,4-trimethylheptane
C10H22
2,3,5-trimethylheptane
C10H22
2,3,6-trimethylheptane
C10H22
2,4,4-trimethylheptane
C10H22
2,4,5-trimethylheptane
C10H22
2,4,6-trimethylheptane
C10H22
2,5,5-trimethylheptane
C10H22
3,3,4-trimethylheptane
C10H22
3,3,5-trimethylheptane
C10H22
3,4,4-trimethylheptane
C10H22
3,4,5-trimethylheptane
C10H22
2-methyl-3-(propan-2-yl)hexane
C10H22
3,3-diethylhexane
C10H22
3,4-diethylhexane

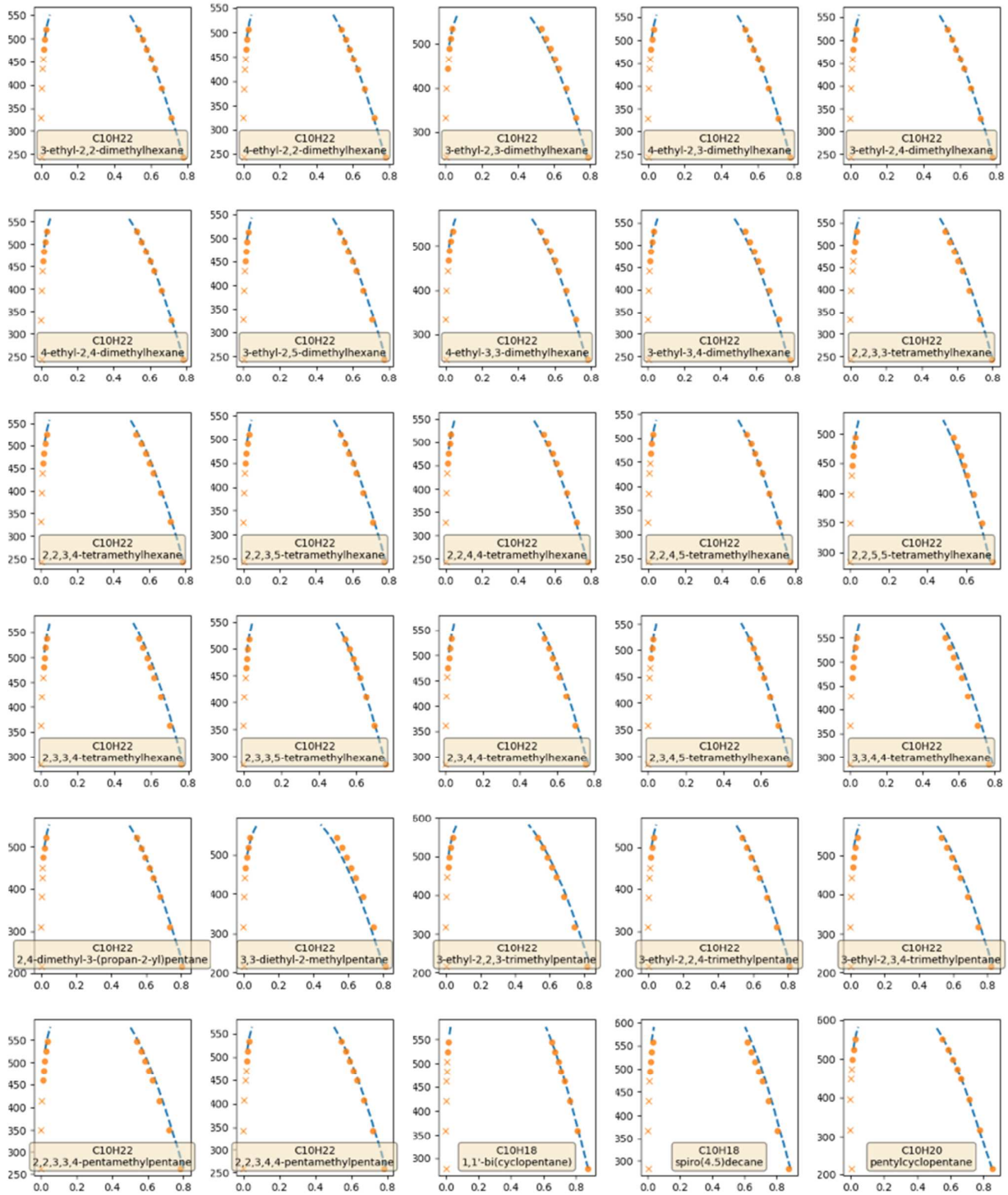
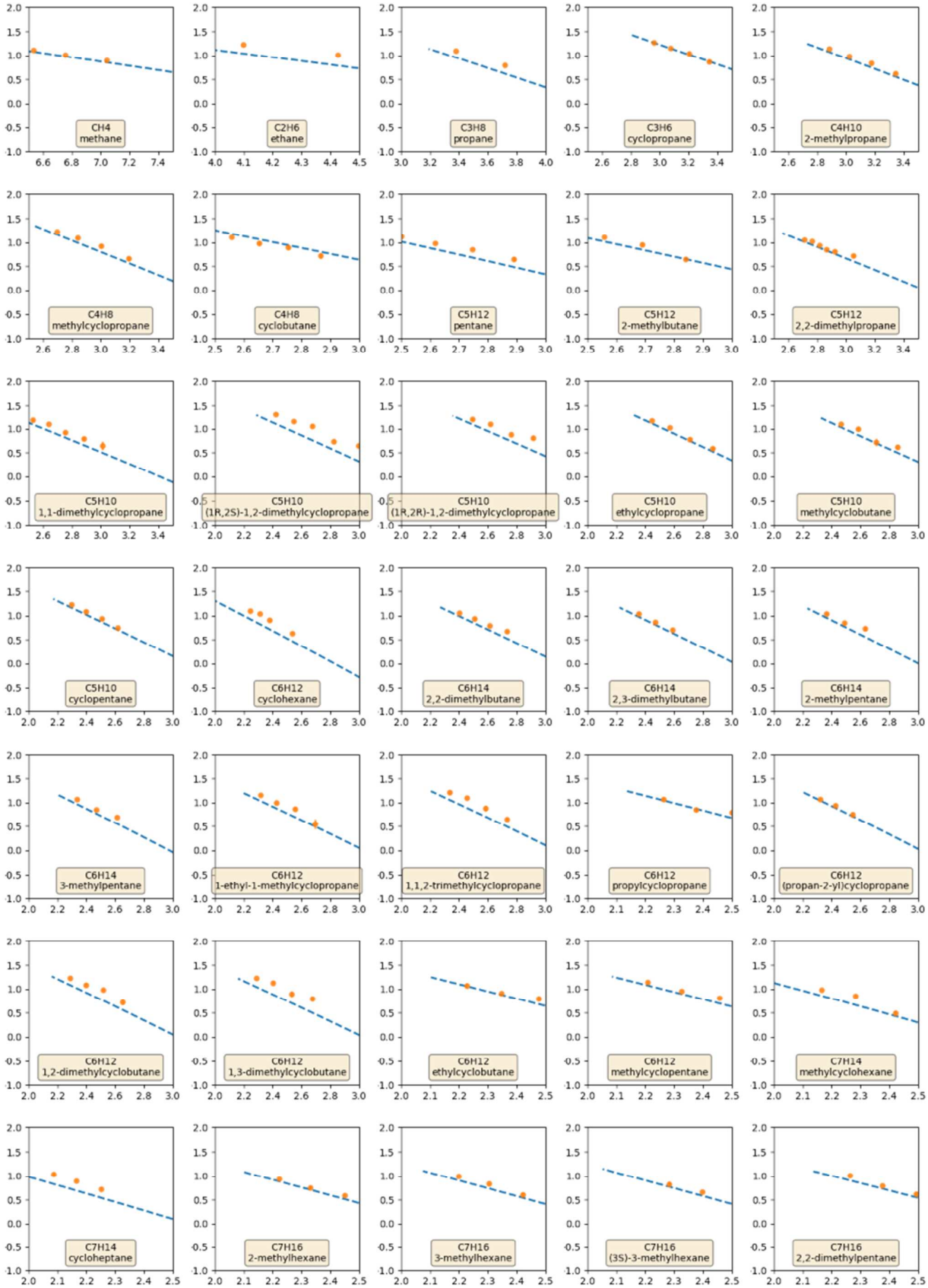
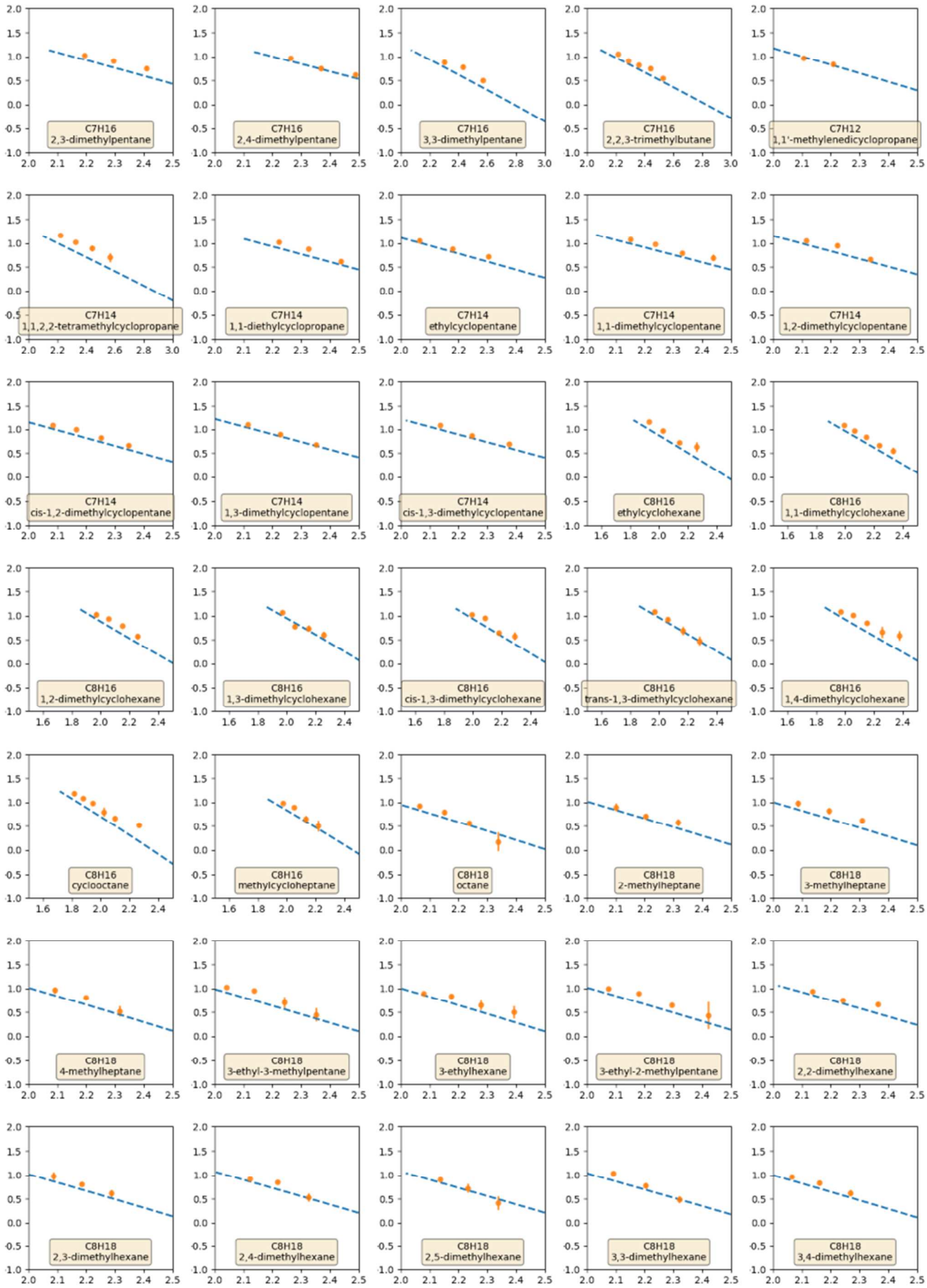
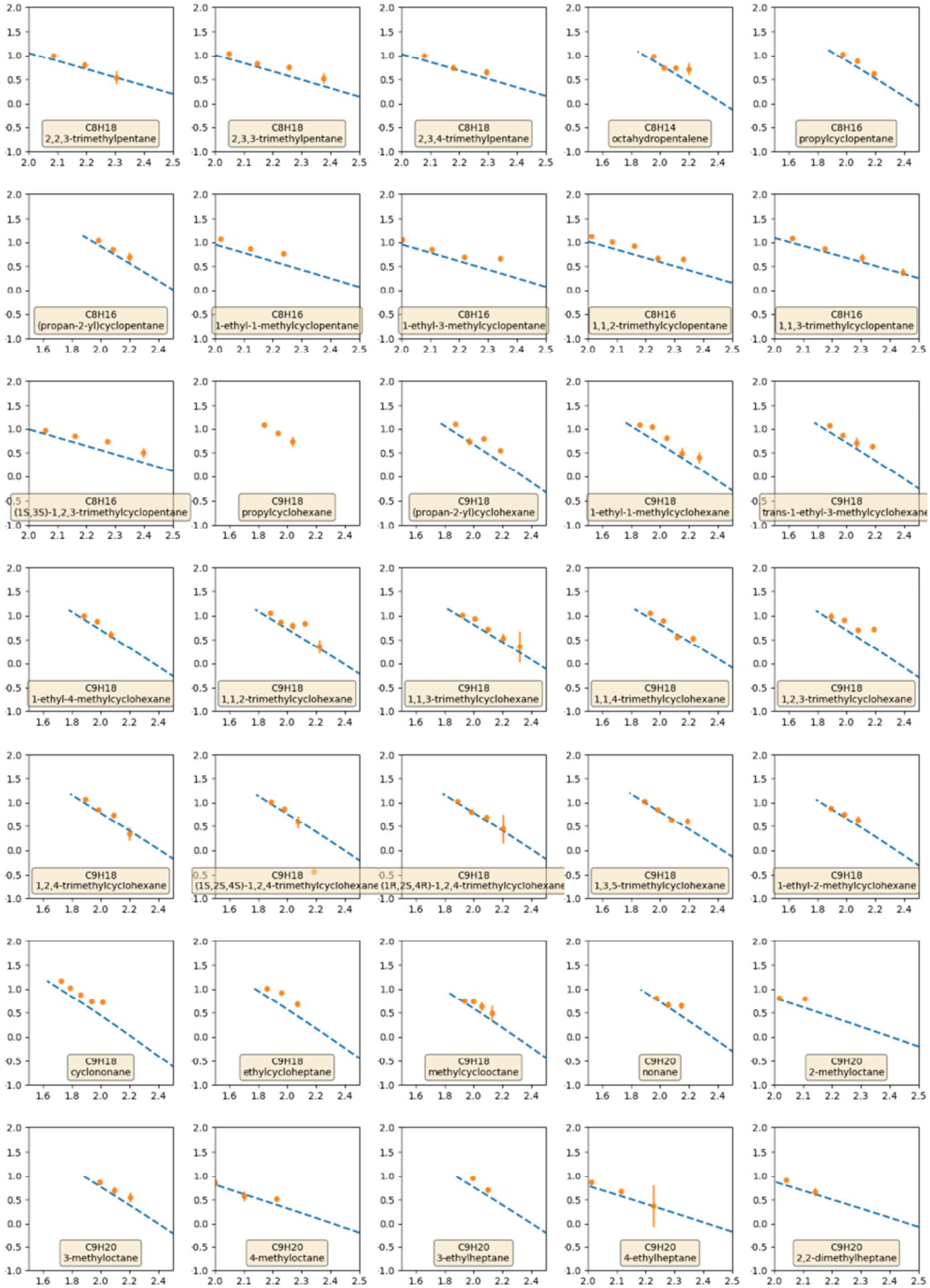
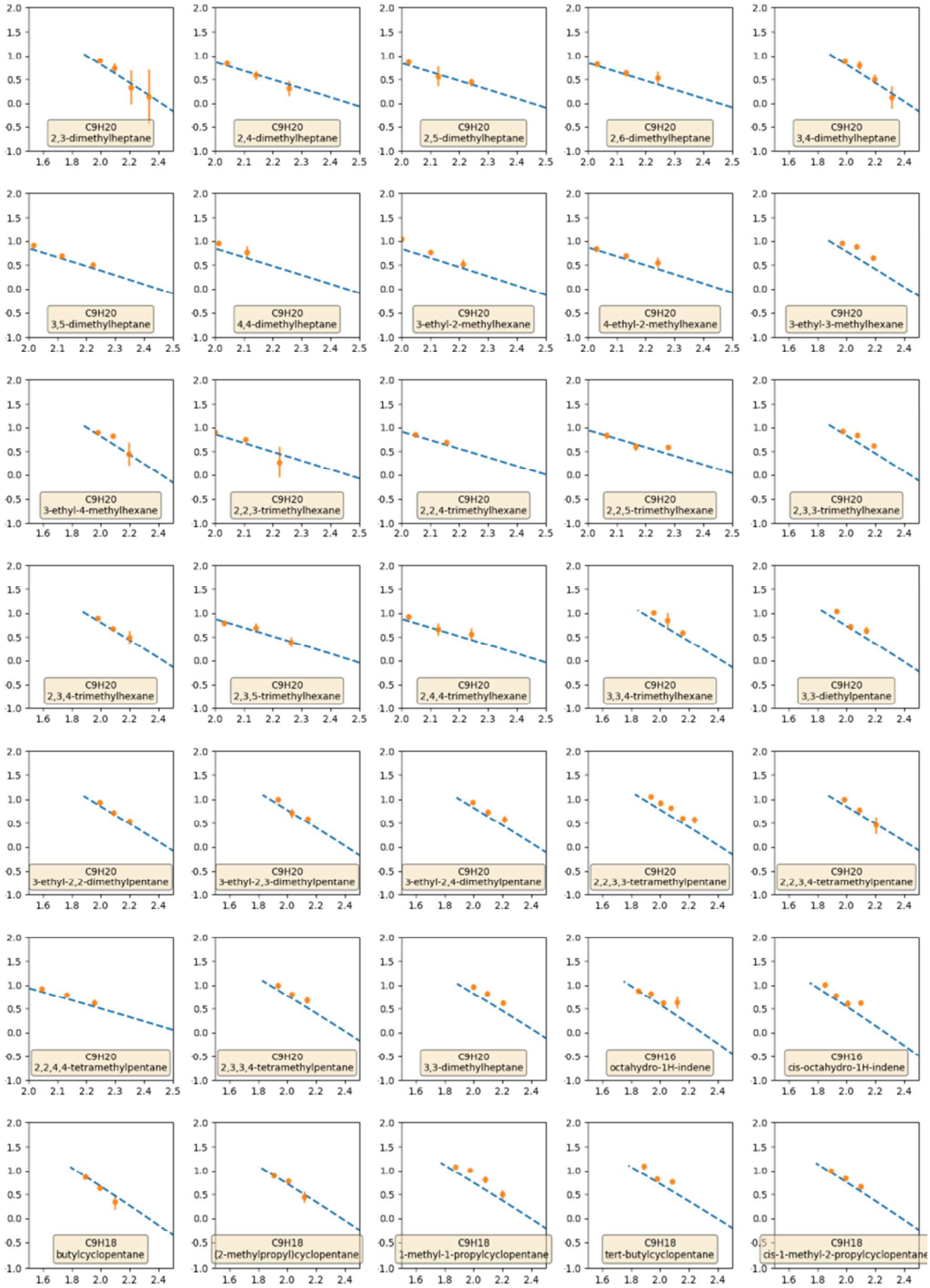


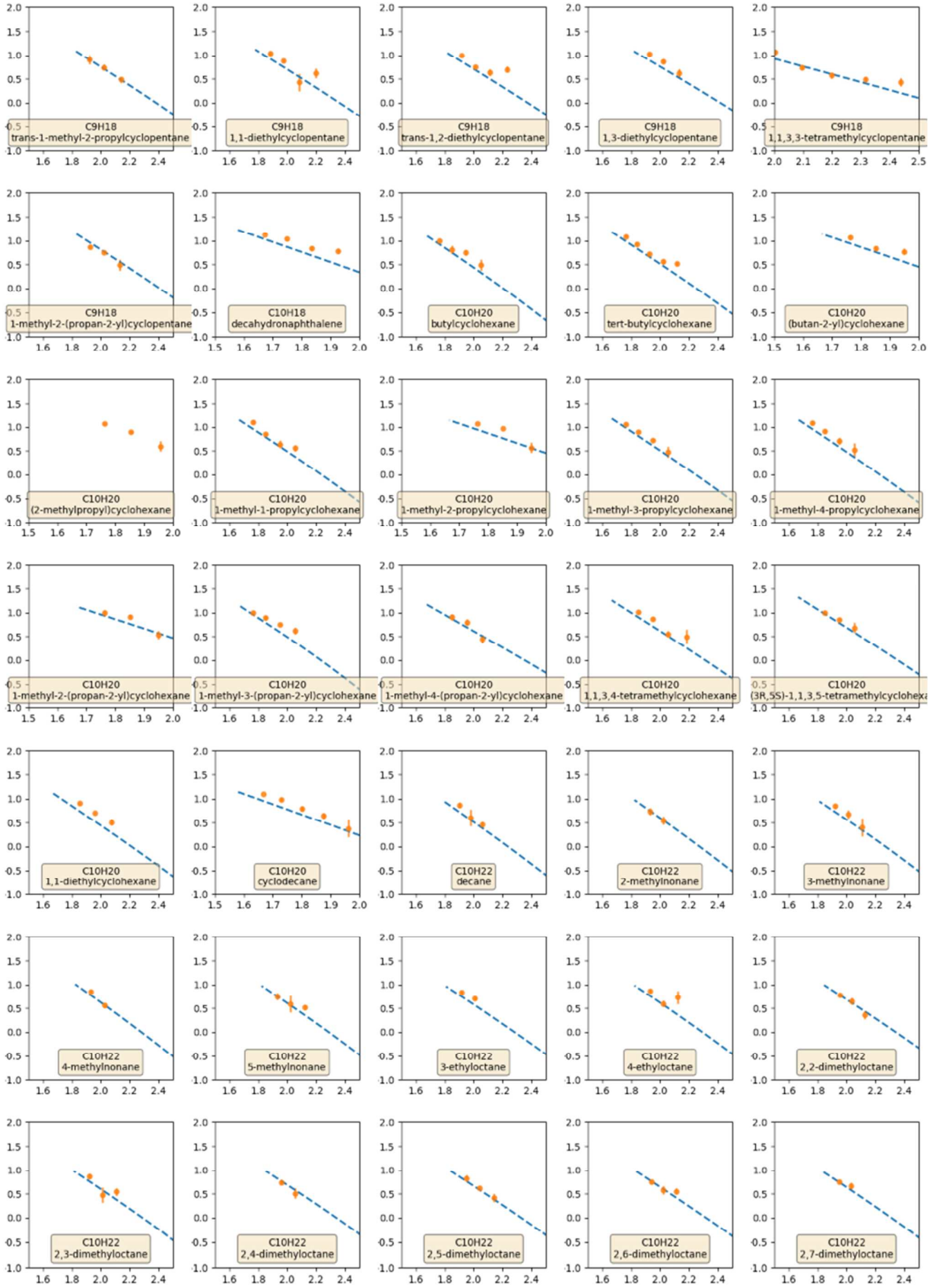
Figure S3 Comparison of predicted (dots and crosses) and experimental (lines) VLE phase diagram of 240 molecules. The cross symbol denotes less confident data (densities lower than 0.01 g/cm^3).

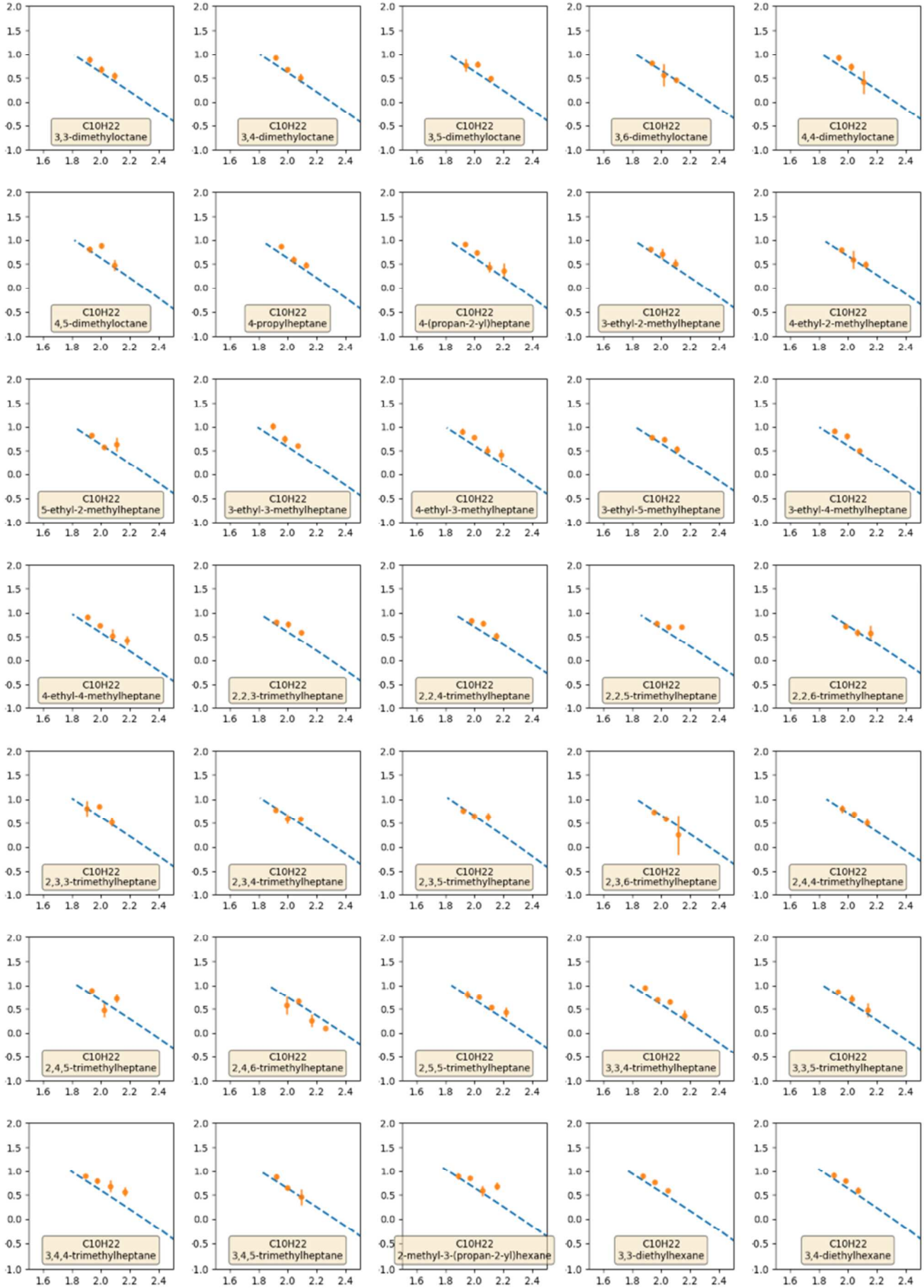

CH4
methane
C2H6
ethane
C3H8
propane
C3H6
cyclopropane
C4H10
2-methylpropane
C4H8
methylcyclopropane
C4H8
cyclobutane
C5H12
pentane
C5H12
2-methylbutane
C5H12
2,2-dimethylpropane
C5H10
1,1-dimethylcyclopropane
C5H10
(1R,2S)-1,2-dimethylcyclopropane
C5H10
(1R,2R)-1,2-dimethylcyclopropane
C5H10
ethylcyclopropane
C5H10
methylcyclobutane
C5H10
cyclopentane
C6H12
cyclohexane
C6H14
2,2-dimethylbutane
C6H14
2,3-dimethylbutane
C6H14
2-methylpentane
C6H14
3-methylpentane
C6H12
1-ethyl-1-methylcyclopropane
C6H12
1,1,2-trimethylcyclopropane
C6H12
propylcyclopropane
C6H12
(propan-2-yl)cyclopropane
C6H12
1,2-dimethylcyclobutane
C6H12
1,3-dimethylcyclobutane
C6H12
ethylcyclobutane
C6H12
methylcyclopentane
C7H14
methylcyclohexane
C7H14
cycloheptane
C7H16
2-methylhexane
C7H16
3-methylhexane
C7H16
(3S)-3-methylhexane
C7H16
2,2-dimethylpentane


C7H16
2,3-dimethylpentane
C7H16
2,4-dimethylpentane
C7H16
3,3-dimethylpentane
C7H16
2,2,3-trimethylbutane
C7H12
1,1'-methylenedicyclopropane
C7H14
1,1,2,2-tetramethylcyclopropane
C7H14
1,1-diethylcyclopropane
C7H14
ethylcyclopentane
C7H14
1,1-dimethylcyclopentane
C7H14
1,2-dimethylcyclopentane
C7H14
cis-1,2-dimethylcyclopentane
C7H14
1,3-dimethylcyclopentane
C7H14
cis-1,3-dimethylcyclopentane
C8H16
ethylcyclohexane
C8H16
1,1-dimethylcyclohexane
C8H16
1,2-dimethylcyclohexane
C8H16
1,3-dimethylcyclohexane
C8H16
cis-1,3-dimethylcyclohexane
C8H16
trans-1,3-dimethylcyclohexane
C8H16
1,4-dimethylcyclohexane
C8H16
cyclooctane
C8H16
methylcycloheptane
C8H18
octane
C8H18
2-methylheptane
C8H18
3-methylheptane
C8H18
4-methylheptane
C8H18
3-ethyl-3-methylpentane
C8H18
3-ethylhexane
C8H18
3-ethyl-2-methylpentane
C8H18
2,2-dimethylhexane
C8H18
2,3-dimethylhexane
C8H18
2,4-dimethylhexane
C8H18
2,5-dimethylhexane
C8H18
3,3-dimethylhexane
C8H18
3,4-dimethylhexane


C8H18
2,2,3-trimethylpentane
C8H18
2,3,3-trimethylpentane
C8H18
2,3,4-trimethylpentane
C8H14
octahydropentalene
C8H16
propylcyclopentane
C8H16
(propan-2-yl)cyclopentane
C8H16
1-ethyl-1-methylcyclopentane
C8H16
1-ethyl-3-methylcyclopentane
C8H16
1,1,2-trimethylcyclopentane
C8H16
1,1,3-trimethylcyclopentane
C8H16
(1S,3S)-1,2,3-trimethylcyclopentane
C9H18
propylcyclohexane
C9H18
(propan-2-yl)cyclohexane
C9H18
1-ethyl-1-methylcyclohexane
C9H18
trans-1-ethyl-3-methylcyclohexane
C9H18
1-ethyl-4-methylcyclohexane
C9H18
1,1,2-trimethylcyclohexane
C9H18
1,1,3-trimethylcyclohexane
C9H18
1,1,4-trimethylcyclohexane
C9H18
1,2,3-trimethylcyclohexane
C9H18
1,2,4-trimethylcyclohexane
C9H18
(1S,2S,4S)-1,2,4-trimethylcyclohexane
C9H18
(1R,2S,4R)-1,2,4-trimethylcyclohexane
C9H18
1,3,5-trimethylcyclohexane
C9H18
1-ethyl-2-methylcyclohexane
C9H18
cyclononane
C9H18
ethylcycloheptane
C9H18
methylcyclooctane
C9H20
nonane
C9H20
2-methyloctane
C9H20
3-methyloctane
C9H20
4-methyloctane
C9H20
3-ethylheptane
C9H20
4-ethylheptane
C9H20
2,2-dimethylheptane


C9H20
2,3-dimethylheptane
C9H20
2,4-dimethylheptane
C9H20
2,5-dimethylheptane
C9H20
2,6-dimethylheptane
C9H20
3,4-dimethylheptane
C9H20
3,5-dimethylheptane
C9H20
4,4-dimethylheptane
C9H20
3-ethyl-2-methylhexane
C9H20
4-ethyl-2-methylhexane
C9H20
3-ethyl-3-methylhexane
C9H20
3-ethyl-4-methylhexane
C9H20
2,2,3-trimethylhexane
C9H20
2,2,4-trimethylhexane
C9H20
2,2,5-trimethylhexane
C9H20
2,3,3-trimethylhexane
C9H20
2,3,4-trimethylhexane
C9H20
2,3,5-trimethylhexane
C9H20
2,4,4-trimethylhexane
C9H20
3,3,4-trimethylhexane
C9H20
3,3-diethylpentane
C9H20
3-ethyl-2,2-dimethylpentane
C9H20
3-ethyl-2,3-dimethylpentane
C9H20
3-ethyl-2,4-dimethylpentane
C9H20
2,2,3,3-tetramethylpentane
C9H20
2,2,3,4-tetramethylpentane
C9H20
2,2,4,4-tetramethylpentane
C9H20
2,3,3,4-tetramethylpentane
C9H20
3,3-dimethylheptane
C9H16
octahydro-1H-indene
C9H16
cis-octahydro-1H-indene
C9H18
butylcyclopentane
C9H18
(2-methylpropyl)cyclopentane
C9H18
1-methyl-1-propylcyclopentane
C9H18
tert-butylcyclopentane
C9H18
cis-1-methyl-2-propylcyclopentane


C9H18
trans-1-methyl-2-propylcyclopentane
C9H18
1,1-diethylcyclopentane
C9H18
trans-1,2-diethylcyclopentane
C9H18
1,3-diethylcyclopentane
C9H18
1,1,3,3-tetramethylcyclopentane
C9H18
1-methyl-2-(propan-2-yl)cyclopentane
C10H18
decahydronaphthalene
C10H20
butylcyclohexane
C10H20
tert-butylcyclohexane
C10H20
(butan-2-yl)cyclohexane
C10H20
(2-methylpropyl)cyclohexane
C10H20
1-methyl-1-propylcyclohexane
C10H20
1-methyl-2-propylcyclohexane
C10H20
1-methyl-3-propylcyclohexane
C10H20
1-methyl-4-propylcyclohexane
C10H20
1-methyl-2-(propan-2-yl)cyclohexane
C10H20
1-methyl-3-(propan-2-yl)cyclohexane
C10H20
1-methyl-4-(propan-2-yl)cyclohexane
C10H20
1,1,3,4-tetramethylcyclohexane
C10H20
(3R,5S)-1,1,3,5-tetramethylcyclohexa
C10H20
1,1-diethylcyclohexane
C10H20
cyclodecane
C10H22
decane
C10H22
2-methylnonane
C10H22
3-methylnonane
C10H22
4-methylnonane
C10H22
5-methylnonane
C10H22
3-ethyloctane
C10H22
4-ethyloctane
C10H22
2,2-dimethyloctane
C10H22
2,3-dimethyloctane
C10H22
2,4-dimethyloctane
C10H22
2,5-dimethyloctane
C10H22
2,6-dimethyloctane
C10H22
2,7-dimethyloctane


C10H22
3,3-dimethyloctane
C10H22
3,4-dimethyloctane
C10H22
3,5-dimethyloctane
C10H22
3,6-dimethyloctane
C10H22
4,4-dimethyloctane
C10H22
4,5-dimethyloctane
C10H22
4-propylheptane
C10H22
4-(propan-2-yl)heptane
C10H22
3-ethyl-2-methylheptane
C10H22
4-ethyl-2-methylheptane
C10H22
5-ethyl-2-methylheptane
C10H22
3-ethyl-3-methylheptane
C10H22
4-ethyl-3-methylheptane
C10H22
3-ethyl-5-methylheptane
C10H22
3-ethyl-4-methylheptane
C10H22
4-ethyl-4-methylheptane
C10H22
2,2,3-trimethylheptane
C10H22
2,2,4-trimethylheptane
C10H22
2,2,5-trimethylheptane
C10H22
2,2,6-trimethylheptane
C10H22
2,3,3-trimethylheptane
C10H22
2,3,4-trimethylheptane
C10H22
2,3,5-trimethylheptane
C10H22
2,3,6-trimethylheptane
C10H22
2,4,4-trimethylheptane
C10H22
2,4,5-trimethylheptane
C10H22
2,4,6-trimethylheptane
C10H22
2,5,5-trimethylheptane
C10H22
3,3,4-trimethylheptane
C10H22
3,3,5-trimethylheptane
C10H22
3,4,4-trimethylheptane
C10H22
3,4,5-trimethylheptane
C10H22
2-methyl-3-(propan-2-yl)hexane
C10H22
3,3-diethylhexane
C10H22
3,4-diethylhexane

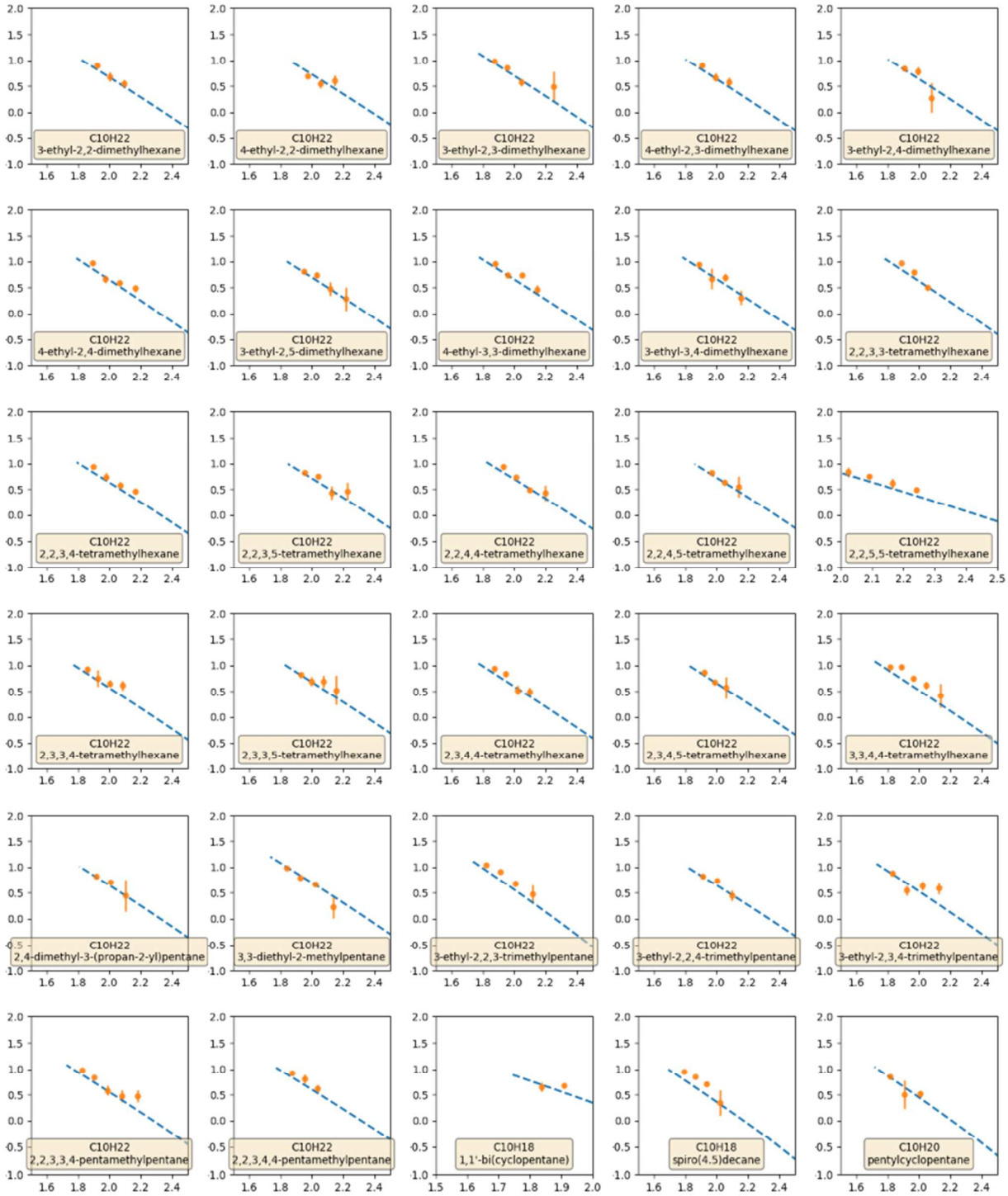


Figure S4 Comparison of predicted (dots) and experimental (lines) Clausius-Clapeyron plots of 240 molecules. The x-axis is $\frac{1000\ \mathrm{K}}{T}$, y axis is $\log_{10}\left(\frac{P}{1\mathrm{bar}}\right)$. The vapor pressure of systems with gas density smaller than 0.01 g/cm^3 are not shown.

References

(1) Yaws, C. L. *Yaws' Critical Property Data for Chemical Engineers and Chemists*; Knovel, 2012.

(2) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202-D1213.

(3) Knovel Data Analytics. https://app.knovel.com/web/poc/ms/discovery.html (accessed Mar. 2018).

(4) Diky, V.; Muzny, C. D.; Smolyanitsky, A. Y.; Bazyleva, A.; Chirico, R. D.; Magee, J. W.; Paulechka, Y.; Kazakov, A. F.; Townsend, S. A.; Lemmon, E. W. ThermoData Engine (TDE) Version 10 (Pure Compounds, Binary Mixtures, Ternary Mixtures, and Chemical Reactions): NIST Standard Reference Database 103b. **2015**.

(5) Frenkel, M.; Chirico, R. D.; Diky, V.; Yan, X.; Dong, Q.; Muzny, C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. *J. Chem. Inf. Model.* **2005**, *45*, 816-838.