Electronic Supplementary Material

Design of bio-oil additives via molecular signature descriptors using a multi-stage computer-aided molecular design framework

Jia Wen Chong[1], Suchithra Thangalazhy Gopakumar[1], Kasturi Muthoosamy[2], Nishanth G Chemmangattuvalappil (✉)[1]

1 Department of Chemical and Environmental Engineering, University of Nottingham Malaysia, Selangor 43500, Malaysia

2 Nanotechnology Research Group, Centre of Nanotechnology and Advanced Materials, University of Nottingham Malaysia, Selangor 43500, Malaysia

E-mail: Nishanth.C@nottingham.edu.my

Appendix

Table 1 Property prediction models used in case study

Property	Property Model	SD	AAE
Normal Melting Point (T_m) [1]	$\exp\left(\frac{T_m}{T_{m0}}\right) = \sum_i N_i T_{mi} + \sum_i M_j T_{mj} + \sum_i O_k T_{mk}$ $T_{m0} = 147.450K$	25.34	18.76
Normal Boiling Point (T_b) [1]	$\exp\left(\frac{T_b}{T_{b0}}\right) = \sum_i N_i T_{bi} + \sum_i M_j T_{bj} + \sum_i O_k T_{bk}$ $T_{b0} = 222.543K$	8.01	5.89
Flash Point (F_p) [2]	$F_p - F_{p0} = \sum_i N_i T_{pi} + \sum_i M_j T_{pj} + \sum_i E_k T_{pk}$ $F_{p0} = 150.0218K$	12.10	8.97
Molar Volume (V_m) [3]	$V_{m(25^oC)} - d = \sum_i N_i v_{m1i} + \sum_j M_j v_{m2j}$ $d = 0.01211$	0.002	0.001
Vapour Pressure (P_{vp}) [4]	$P_{sat} = 5.58 - 2.7\left(\frac{T_b}{298.15}\right)^{1.7}$ $T_b = Normal\ Boiling\ Point$	N/A	N/A
Higher Heating Value (HHV) [5]	$HHV = \frac{\sum_i N_i H_i}{\sum_i N_i M_i} \left({}^{kJ}/_{g}\right)$ $M = molecular\ weight\ (g/mol)$	N/A	N/A
Density (ρ)	$\rho = \frac{M_i}{V_m}$	N/A	N/A

Dynamic Viscosity (η_L) [6]	$\ln \eta_L = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k$	0.89	0.37
Octanol/water partition coefficient (K_{ow}) [7]	$\log K_{ow} = 1.267\,({}^v\chi^1) + 0.612({}^v\chi^3) - 0.976({}^v\chi^3) - 2.130$	N/A	N/A
Acute Toxicity (Aquatic, LC_{50}) [8]	$\log LC_{50}^{-1} = 0.81 \log K_{ow} + 1.744$	N/A	N/A
Acute Toxicity (Aquatic, EC_{50}) [9]	$\log EC_{50} = -0.95 \log K_{ow} - 1.32$	N/A	N/A
Acute Toxicity (Oral, LD_{50}) [10]	$\log LD_{50}^{-1} = 0.805 \log K_{ow} - 0.971 \log(0.0807 K_{ow} + 1) + 0.984$	N/A	N/A
Relative Toxicity (IGC_{50}) [11]	$\log IGC_{50}^{-1} = 0.723(0.14) \log K_{ow} - 1.79(0.031)$	N/A	N/A
Bioconcentration Factor [12]	$\log BCF = 0.032 + 0.636 \log K_{ow}$	N/A	N/A
Soil-water partition coefficient (K_{oc}) [13]	$\log K_{oc} = 0.59\,({}^v\chi^1) - 0.97$	N/A	N/A
Global Warming Potential [14]	$\log GWP = \sum_i a_i A_i + b({}^v\chi^0) + 2c({}^v\chi^1) + d$ $b = -0.01877; c = -1.52848; d = -0.52073$	0.48	0.36
Photochemical Oxidation Potential [14]	$-\log PCO = \sum_i a_i A_i + b({}^v\chi^0) + 2c({}^v\chi^1) + d$ $b = -0.10486; c = 0.005087; d = -0.25708$	0.33	0.27

Group Contribution Functional Groups

First-order GC groups considered:

1. CH_3
2. CH_2
3. CH
4. C
5. CH_2=CH
6. CH=CH
7. CH_2=C
8. CH=C
9. C=C
10. OH
11. COOH
12. CH_3CO
13. CH_2CO
14. CHCO
15. CCO
16. CHO
17. CH_3COO
18. CH_2COO
19. CHCOO
20. CCOO
21. HCOO
22. CH_3O
23. CH_2O
24. CH-O
25. C-O
26. CH_2CN
27. CHCN
28. CCN

Second-order GC groups considered:

1. $(CH_3)_2CH$
2. $(CH_3)_3C$
3. $CH(CH_3)CH(CH_3)$
4. $CH(CH_3)C(CH_3)_2$
5. $C(CH_3)_2C(CH_3)_2$
6. $CH_n=CH_m-CH_p=CH_k$ (k,m,n,p in 0..2)
7. $CH_3-CH_m=CH_n$ (m,n in 0..2)
8. $CH_2-CH_m=CH_n$ (m,n in 0..2)
9. $CH_p-CH_m=CH_n$ (m,n in 0..2; p in 0..1)
10. CHCHO or CCHO
11. CH_3COCH_2
12. CH_3COCH or CH_3COC
13. CHCOOH or CCOOH
14. CH_3COOCH or CH_3COOC
15. CHOH
16. COH
17. CH_3COCH_nOH (n in 0..2)
18. NCCHOH or NCCOH
19. $OH-CH_n-COO$ (n in 0.2)
20. $CH_m(OH)CH_n(OH)$ (m,n in 0..2)
21. $CH_m(OH)CH_n(-)$ (m,n in 0..2)
22. $HOOC-CH_n-COOH$ (n on 1..2)
23. $HOOC-CH_n-CH_m-COOH$ (n,m on 1..2)
24. $HO-CH_n-COOH$ (n on 1..2)
25. CH_3-O-CH_n-COOH (n in 1..2)
26. $NC-CH_n-CH_m-CN$ (n,m in 1..2)
27. $OH-CH_n-CH_m-CN$ (n,m in 1..2)
28. $COO-CH_n-CH_m-OOC$ (n,m in 1..2)
29. $OOC-CH_n-CH_m-COO$ (n,m in 1..2)
30. $NC-CH_n-COO$ (n in 1..2)
31. $COCH_nCOO$ (n in 1..2)
32. $CH_m-O-CH_n=CH_p$ (m,n,p in 0..3)
33. $CH_m=CH_n-CN$ (m,n in 0..2)
34. $CH_m=CH_n-COO-CH_p$ (m,n,p in 0..3)
35. $CH_m=CH_n-CHO$ (m,n in 0..2)
36. $CH_m=CH_n-COOH$ (m,n in 0..2)

Third-order GC groups considered:

1. $HOOC-(CH_n)_m-COOH$ (m>2, n in 0..2)
2. $OH-(CH_n)_m-OH$ (m>2, n in 0..2)
3. $OH-(CH_p)_k-O-(CH_n)_m-H$ (m,k>0; p,n in 0..2)
4. $CH_p-O-(CH_n)_m-OH$ (m>2; n,p in 0..2)
5. $NC-(CH_n)_m-CN$ (m>2; n in 0..2)
6. $COO-(CH_n)_m-OOC$ (m>2; n in 0..2)

Table 2 Chemical classes considered and their respective chemical groups

Chemical Class	Chemical Group
Alkanes	CH_3, CH_2, CH, C
Alkenes	CH_2=CH, CH=CH, CH_2=C, CH=C, C=C
Alcohol	OH
Carboxylic Acid	COOH
Ketones	CH_3CO, CH_2CO, CHCO, CCO
Aldehyde	CHO
Esters	CH_3COO, CH_2COO, CHCOO, CCOO, HCOO
Ethers	CH_3O, CH_2O, CH-O, C-O
Nitriles	CH_2CN, CHCN, CCN

Table 3 Height 1 signature and their corresponding GC group

No	Signature	Corresponding Group	No	Signature	Corresponding Group
S1	C1(C)	CH_3	**S13**	C4(=C=O)	C = C
S2	C1(O)	CH_3O	**S14**	C4(=CCC)	C = C
S3	C2(=C)	CH_2 = C	**S15**	C4(=CCO)	C = C
S4	C2(CC)	CH_2	**S16**	C4(=OCC)	CHCO
S5	C2(CO)	CH_2	**S17**	C4(=OCO)	CH_2COO
S6	C3(=CC)	CH = C	**S18**	C4(≡NC)	CCN
S7	C3(=CO)	CH = C	**S19**	C4(CCCC)	C
S8	C3(=OC)	CHO	**S20**	C4(CCCO)	C
S9	C3(=OO)	HCOO	**S21**	N3(≡C)	CCN
S10	C3(CCC)	CH	**S22**	O1(C)	OH
S11	C3(CCO)	CH	**S23**	O2(=C)	COOH
S12	C4(=C=C)	C = C	**S24**	O2(CC)	CH-O

Table 4 Height 2 signature and their corresponding GC group

No	Signature	Corresponding Group
D1	C1(C3(CCO))	CH_3
D2	C1(C2(CC))	CH_3
D3	C1(C2(CO))	CH_3
D4	C2(C1(C)C2(CC))	CH_2
D5	C2(C1(C)C2(CO))	CH_2
D6	C2(C1(C)C3(CCO))	CH_2
D7	C2(C2(CC)C2(CC))	CH_2
D8	C2(C2(CC)C2(CO))	CH_2
D9	C2(C2(CC)C3(CCO))	CH_2
D10	C2(C2(CO)C3(CCO))	CH_2
D11	C2(C1(CO)O1(C))	CH_2
D12	C2(C2(CC)O1(C))	CH_2
D13	C3(C1(C)C1(C)O1(C))	CH

D14	C3(C1(C)C2(CC)O1(C))	CH
D15	C3(C2(CC)C2(CC)O1(C))	CH
D16	O1(C2(CO))	OH
D17	O1(C3(CCO))	OH

Table 5 Height 3 signature and their corresponding GC group

No	Signature	Corresponding Group
T1	C1(C3(C1(C)C2(CC)O1(C)))	CH_3
T2	C1(C2(C1(C)C2(CC)))	CH_3
T3	C2(C1(C2(CC))C2(C1(C)C2(CC)))	CH_2
T4	C2(C1(C2(CC))C2(C2(CC)C2(CC)))	CH_2
T5	C2(C1(C2(CC))C2(C2(CC)C3(CCO)))	CH_2
T6	C2(C2(C1(C)C2(CC))C2(C1(C)C2(CC)))	CH_2
T7	C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC)))	CH_2
T8	C2(C2(C1(C)C2(CC))C2(C2(CC)C3(CCO)))	CH_2
T9	C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC)))	CH_2
T10	C2(C2(C2(CC)C2(CC))C2(C2(CC)C3(CCO)))	CH_2
T11	C2(C2(C1(C)C2(CC))C3(C1(C)C2(CC)O1(C)))	CH_2
T12	C2(C2(C2(CC)C2(CC))C3(C1(C)C2(CC)O1(C)))	CH_2
T13	C3(C1(C3(CCO))C2(C2(CC)C3(CCO))O1(C3(CCO)))	CH
T14	O1(C3(C1(C)C2(CC)O1(C)))	OH

Table 66 Height 4 signature and their corresponding GC group

No	Signature
Q1	C1(C3(C1(C3(CCO))C2(C2(CC)C3(CCO))O1(C3(CCO))))
Q2	C1(C2(C1(C2(CC))C2(C2(CC)C2(CC))))
Q3	C2(C1(C2(C1(C)C2(CC))C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC))))
Q4	C2(C1(C2(C1(C)C2(CC))C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC))))
Q5	C2(C1(C2(C1(C)C2(CC))C2(C2(C2(CC)C2(CC))C2(C2(CC)C3(CCO))))
Q6	C2(C2(C1(C2(CC))C2(C2(CC)C2(CC)))C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC))))
Q7	C2(C2(C1(C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC))))
Q8	C2(C2(C1(C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C2(C2(CC)C3(CCO))))
Q9	C2(C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC)))C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC))))
Q10	C2(C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC))))
Q11	C2(C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC))))

Q12	C2(C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C2(C2(CC)C3(CCO))))
Q13	C2(C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C2(C2(CC)C3(CCO))))
Q14	C2(C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C3(C1(C)C2(CC)O1(C))))
Q15	C2(C2(C2(C2(CC)C2(CC))C2(C2(CC)C2(CC)))C2(C2(C2(CC)C2(CC))C3(C1(C)C2(CC)O1(C))))
Q16	C2(C2(C2(C2(CC)C2(CC))C2(C2(CC)C3(CCO)))C2(C2(C2(CC)C2(CC))C3(C1(C)C2(CC)O1(C))))
Q17	C2(C2(C2(C1(C)C2(CC))C2(C2(CC)C2(CC)))C3(C1(C3(CCO))C2(C2(CC)C3(CCO))O1(C3(CCO))))
Q18	C2(C2(C2(C2(CC)C2(CC))C2(C2(CC)C3(CCO)))C3(C1(C3(CCO))C2(C2(CC)C3(CCO))O1(C3(CCO))))
Q19	C2(C2(C2(C2(CC)C2(CC))C3(C1(C)C2(CC)O1(C)))C3(C1(C3(CCO))C2(C2(CC)C3(CCO))O1(C3(CCO))))
Q20	C3(C1(C3(C1(C)C2(CC)O1(C)))C2(C2(C2(CC)C2(CC))C3(C1(C)C2(CC)O1(C)))O1(C3(C1(C)C2(CC)O1(C))))
Q21	O1(C3(C1(C3(CCO))C2(C2(CC)C3(CCO))O1(C3(CCO))))

Phase analysis

To estimate the activity coefficients in non-ideal liquid mixture, group contribution estimation approach developed by Fredenslund *et al.* (1975)[12] was applied. In this work, the GC prediction model combines the solution-of-functional-groups concept with a model for activity coefficient based on UNIQUAC. In a multi-component mixture, the UNIQUAC equation for the activity coefficient of component i is given by:

$$\ln \gamma_i = \ln \gamma_i^C + \ln \gamma_i^R \tag{1}$$

In Equation 1, C represent the combinatorial part while the residual part is denoted as R. Here, Equation 2 and Equation 3 calculates the value $\ln \gamma_i^C$ and $\ln \gamma_i^R$:

$$\ln \gamma_i^C = \ln \frac{\phi_i}{x_i} + 5 q_i \ln \frac{\theta_i}{\phi_i} + l_i - \frac{\phi_i}{x_i} \sum_j x_j l_j \tag{2}$$

$$\ln \gamma_i^R = \sum_k v_k^{(i)} (\ln \Gamma_k - \ln \Gamma_\mathrm{k}^{(i)}) \tag{3}$$

Equation 4 to Equation 11 represents the calculation for terms in Equation 2 and Equation 3:

$$l_j = 5(r_i - q_i) - (r_i - 1) \tag{4}$$

$$\phi_i = \frac{r_i x_i}{\sum_j r_j x_j} \tag{5}$$

$$\theta_i = \frac{q_i x_i}{\sum_j q_j x_j} \tag{6}$$

$$r_i = \sum_k v_k^{(i)} R_k \tag{7}$$

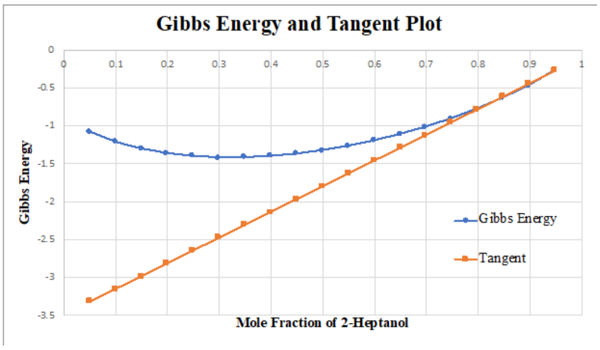
$$q_i = \sum_k v_k^{(i)} Q_k \tag{8}$$

$$\ln \Gamma_k = Q_k \left[1 - \ln \sum_m \vartheta_m \psi_{m,k} - \sum_m \frac{\vartheta_m \psi_{m,k}}{\sum_n \vartheta_n \psi_{n,m}} \right] \tag{9}$$

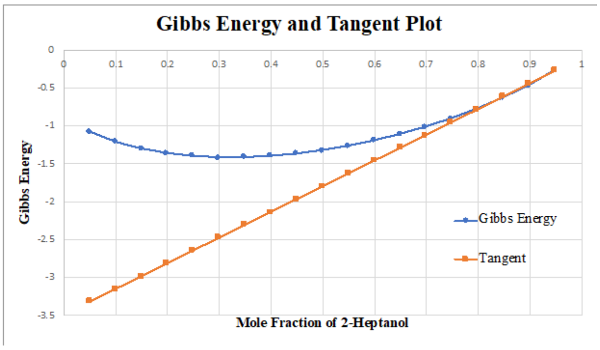
$$\vartheta_m = \frac{Q_m X_m}{\sum_n Q_n X_n} \tag{10}$$

$$\psi_{m,n} = -\exp(\frac{a_{mn}}{T}) \tag{11}$$

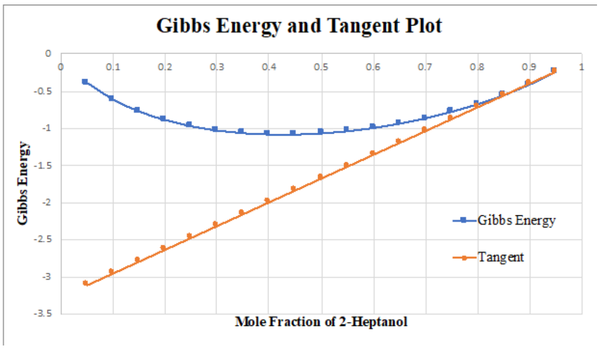
Where γ_i = activity coefficient of component i
ϕ_i = segment fraction (volume fraction) of component i
θ_i = area fraction of component i
x_i = mole fraction of component i
r_i = pure component molecular van der Waals volume parameter
q_i = pure component molecular surface areas parameter
$v_k^{(i)}$ = number of groups of type k in molecule i
R_k = group volume parameters
Q_k = group area parameters
Γ_k = group residual activity coefficient
$\Gamma_k^{(i)}$ = residual activity coefficient of group k in pure component i
ϑ_m = area fraction of group m
$\psi_{m,k}$ = group interaction parameter
X_m = mole fraction of group m in the mixture
$a_{m,n}$ = group interaction parameters obtained from experimental phase equilibrium data



(a)

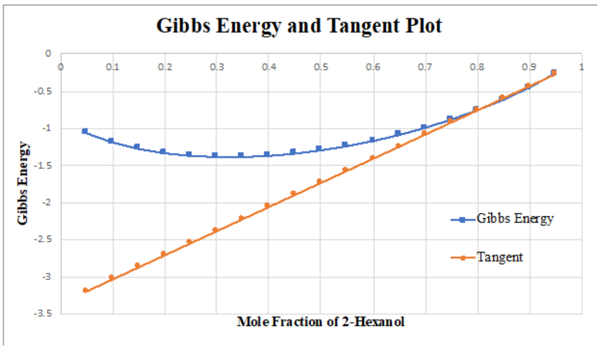


(b)

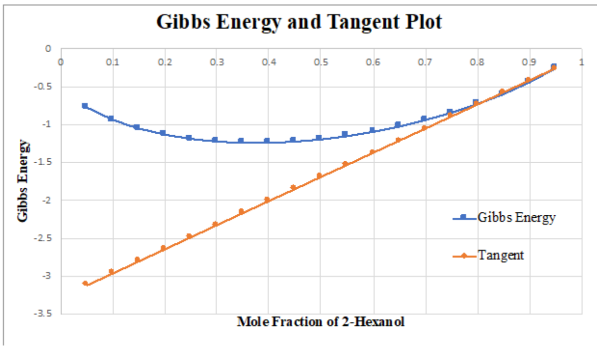


(c)

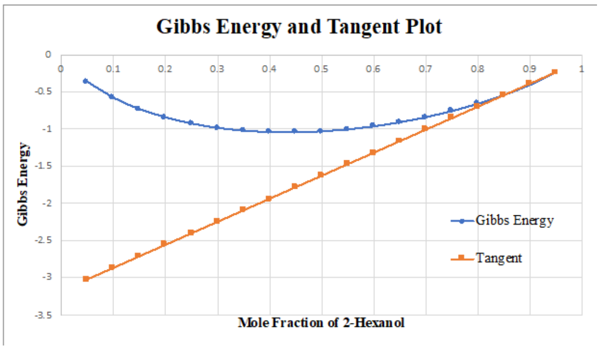
Figure S-1 Gibbs energy and tangent plot for 2-heptanol and bio-oil at (a) 16% water content (b) 25% water content and (c) 40% water content



(a)

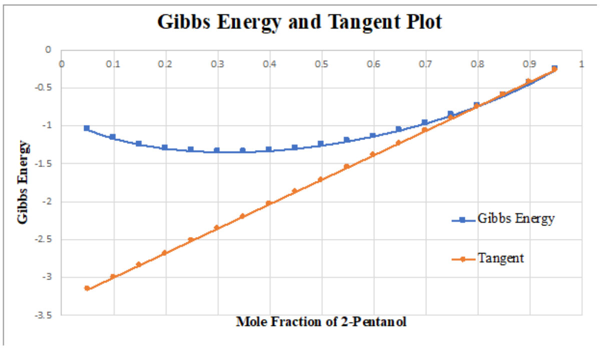


(b)

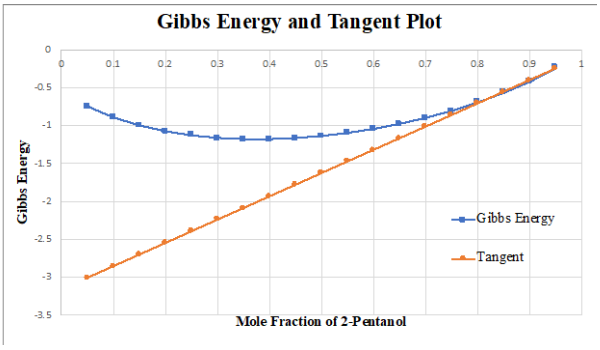


(c)

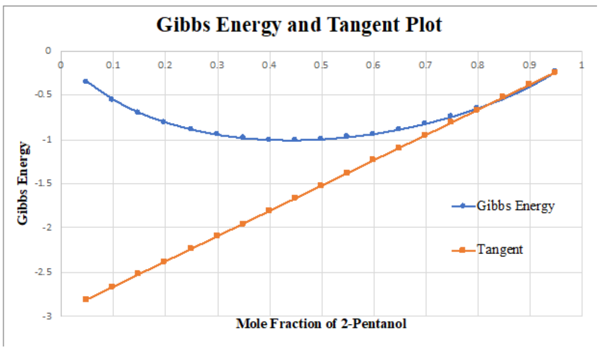
Figure S-2 Gibbs energy and tangent plot for 2-hexanol and bio-oil at (a) 16% water content (b) 25% water content and (c) 40% water content



(a)

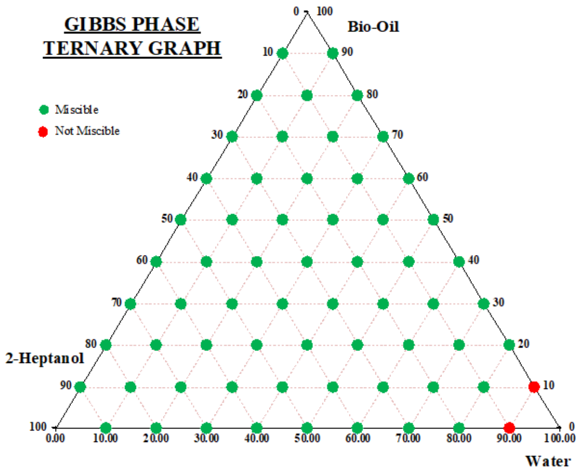


(b)

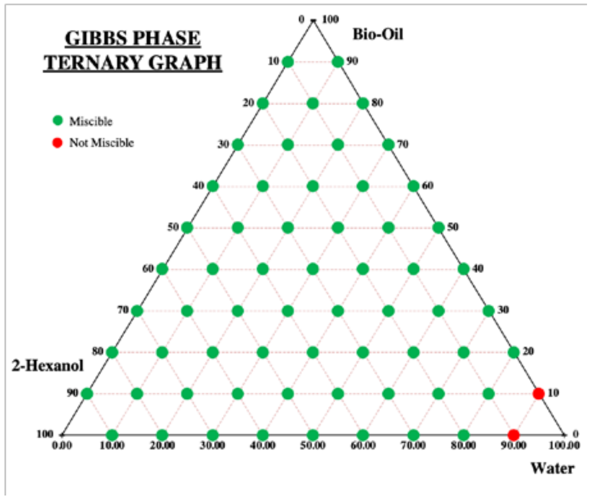


(c)

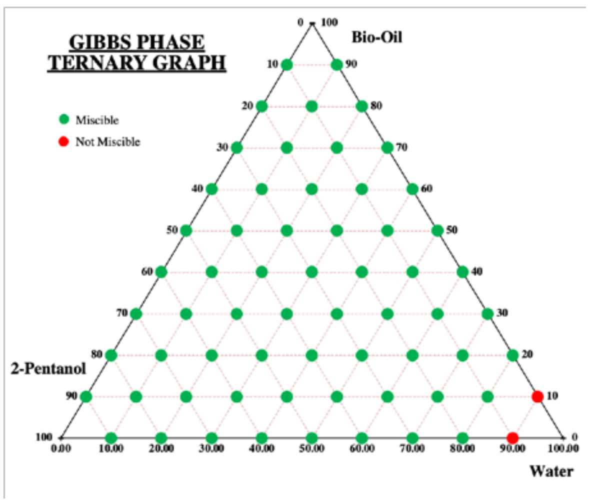
Figure S-3 Gibbs energy and tangent plot for 2-pentanol and bio-oil at (a) 16% water content (b) 25% water content and (c) 40% water content



(a)



(b)



(c)

Figure S-4 Gibbs phase ternary graph of bio-oil, water and (a) 2-heptanol, (b) 2-hexanol and (c) 2-heptanol

Reference

1. Marrero J, Gani R. Group-contribution based estimation of pure component properties. Fluid Phase Equilibria, 2001, 183–208.

2. Hukkerikar AS, Sarup B, Kate A Ten, Abildskov J, Sin G, Gani R. Group-contribution + (GC +) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilibria, 2012, 321: 25–43.

3. Constantinou L, Gani R, O'Connell JP. Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method. Fluid Phase Equilibria, 1995, 103(1): 11–22.

4. Sinha M, Achenie LEK. Systematic design of blanket wash solvents with recovery considerations. Advances in Environmental Research, 2001, 5(3): 239–249.

5. Walters RN. Molar group contributions to the heat of combustion. Fire Mater, 2002, 26(3): 131–145.

6. Conte E, Martinho A, Matos HA, Gani R. Combined group-contribution and atom connectivity index-based methods for estimation of surface tension and viscosity. Industrial & Engineering Chemistry Research, 2008, 47(20): 7940–7954.

7. Boyd JC, Millership JS, Woolfson AD. The Relationship between Molecular Connectivity and Partition Coefficients. Journal of Pharmacy and Pharmacology, 1982, 34(6): 364–366.

8. Pavan M, Worth A, Netzeva T. Preliminary Analysis of an Aquatic Toxicity Dataset and Assessment of QSAR Models for Narcosis. EUR - Scientific and Technical Research Reports. 2005.

9. Carpanini FM. QSARs in the Assessment of the Environmental Fate and Effects of Chemicals. Technical Report No. 74. 1998.

10. Lipnick RL. Structure-Activity relationships in environmental toxicology and chemistry : Narcosis, electrophile and proelectrophile toxicity mechanisms: Application of SAR and QSAR. Environmental Toxicology and Chemistry, 1989, 8(1): 1–12.

11. Schultz TW, Cronin MTD, Netzeva TI, Aptula AO. Structure-toxicity relationships for aliphatic chemicals evaluated with Tetrahymena pyriformis. Chemical Research in Toxicology, 2002, 15(12): 1602–1609.

12. Lu X, Tao S, Hu H, Dawson R. Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors. Chemosphere, 2000, 41(10): 1675–1688.

13. Liao YY, Wang ZT, Chen JW, Han SK, Wang LS, Lu GY, Zhao TN. The Prediction of Soil Sorption Coefficients of Heterocyclic Nitrogen Compounds by Octanol/Water Partition Coefficient, Water Solubility, and by Molecular Connectivity Indices. Bulletin of Environmental Contamination and Toxicology, 1996, 56(5): 711–716.

14. Hukkerikar AS, Kalakul S, Sarup B, Young DM, Sin G, Gani R. Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution+ (GC +) property models and uncertainty analysis. Journal of Chemical Information and Modeling, 2012, 52(11): 2823–2839.

15. Fredenslund A, Jones RL, Prausnitz JM. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. AIChE Journal, 1975, 21(6): 1086–1099.