Supporting Information

Fast Prediction of Equivalent Alkane Carbon Number Using Graph Machines and Neural Networks

Lucie Delforce¹, François Duprat*, Jean-Luc Ploix², Jesus Fermín Ontiveros¹,
Valentin Goussard¹, Véronique Nardello-Rataj*, Jean-Marie Aubry*

¹Université de Lille, CNRS, ENSCL, UMR 8181, UCCS - Unité de Catalyse et de Chimie
du Solide, 59655 Villeneuve d’Ascq, France

²Chimie Moléculaire, Macromoléculaire, Matériaux, ESPCI Paris, CNRS, PSL University,
10 rue Vauquelin, 75005 Paris, France

*To whom correspondence should be addressed

Emails: jean-marie-aubry@univ-lille.fr, veronique.rataj-nardello@univ-lille.fr and arthur.duprat@espci.psl.eu
| Entry | Compound       | SMILES code  | $M_0^X$ [a] | $M_2^X$ [b] | $M_3^X$ [c] | n  | EACN$_{exp}$ | Ref.   | EACN$_{exm}$ [d] | EACN$_{exm}$ [e] | MF | CAS RN |
|-------|----------------|--------------|-------------|-------------|-------------|----|-------------|--------|----------------|----------------|----|--------|
| 1     | Hexane         | CCCCCCCC     | 159.25      | 8.17        | -0.08       | 6  | 6$^{[l]}$   | -      | 5.6            | 6.2            | CsH$_{14}$ | 110-54-3 |
| 2     | Heptane        | CCCCCCCC     | 179.32      | 9.14        | -0.03       | 7  | 7$^{[l]}$   | -      | 6.6            | 7.2            | CsH$_{16}$ | 142-82-5 |
| 3     | Octane         | CCCCCCCCC    | 198.88      | 10.04       | 0.00        | 8  | 8$^{[l]}$   | -      | 7.8            | 8.2            | CsH$_{18}$ | 111-65-9 |
| 4     | Nonane         | CCCCCCCCCC   | 218.16      | 10.71       | 0.00        | 9  | 9$^{[l]}$   | -      | 8.8            | 9.1            | CsH$_{20}$ | 111-84-2 |
| 5     | Decane         | CCCCCCCCCCC  | 238.31      | 11.78       | 0.06        | 10 | 10$^{[l]}$  | -      | 9.9            | 10.0           | CsH$_{22}$ | 124-18-5 |
| 6     | Undecane       | CCCCCCCCCCCC | 258.33      | 12.51       | 0.11        | 11 | 11$^{[l]}$  | -      | 11.0           | 11.0           | CsH$_{24}$ | 1120-21-4 |
| 7     | Dodecane       | CCCCCCCCCCCC | 278.75      | 13.46       | 0.12        | 12 | 12$^{[l]}$  | -      | 12.1           | 11.9           | CsH$_{26}$ | 112-40-3 |
| 8     | Tridecane      | CCCCCCCCCCCC | 298.9       | 14.31       | 0.22        | 13 | 13$^{[l]}$  | -      | 13.1           | 12.9           | CsH$_{28}$ | 629-50-5 |
| 9     | Tetradecane    | CCCCCCCCCCCCCC | 319.04    | 15.18       | 0.26        | 14 | 14$^{[l]}$  | -      | 14.1           | 13.9           | CsH$_{30}$ | 629-59-4 |
| 10    | Pentadecane    | CCCCCCCCCCCCCC | 339.46    | 16.11       | 0.28        | 15 | 15$^{[l]}$  | -      | 15.2           | 15.0           | CsH$_{32}$ | 629-62-9 |
| 11    | Hexadecane     | CCCCCCCCCCCCCC | 359.04    | 17.29       | 0.38        | 16 | 16$^{[l]}$  | -      | 16.2           | 15.9           | CsH$_{34}$ | 544-76-3 |
| 12    | Heptadecane    | CCCCCCCCCCCCCCCC | 379.07    | 18.2        | 0.43        | 17 | 17$^{[l]}$  | -      | 17.1           | 17.0           | CsH$_{36}$ | 629-78-7 |
| 13    | Octadecane     | CCCCCCCCCCCCCCCCCC | 399.34    | 18.9        | 0.42        | 18 | 18$^{[l]}$  | -      | 18.1           | 18.1           | CsH$_{38}$ | 593-45-3 |
| 14    | Nonadecane     | CCCCCCCCCCCCCCCCCC | 419.21    | 20.01       | 0.52        | 19 | 19$^{[l]}$  | -      | 18.9           | 18.8           | CsH$_{40}$ | 629-92-5 |
| 15    | Eicosane       | CCCCCCCCCCCCCCCCCC | 440.35    | 20.71       | 0.52        | 20 | 20$^{[l]}$  | -      | 19.4           | 19.5           | C$_{20}$H$_{42}$ | 112-95-8 |
| 16    | Cyclohexane    | C1CCCCC1      | 135.77      | 5.86        | -0.12       | 6  | 2.1         | 1,4   | 1.9            | 2.9            | CsH$_{12}$ | 110-82-7 |
| 17    | Methylcyclohexane | C1CCCCC1    | 153.55      | 7.16        | -0.04       | 7  | 3.2         | 1     | 2.8            | 3.4            | CsH$_{14}$ | 108-87-2 |
| 18    | Ethylcyclohexane | C1CC1CCC1    | 171.52      | 7.89        | -0.13       | 8  | 4.2         | 1,5   | 4.6            | 4.7            | CsH$_{16}$ | 1678-91-7 |
| 19    | Propylcyclohexane | C1CC1CCC1    | 191.55      | 8.77        | -0.09       | 9  | 5.9         | 1,5   | 5.6            | 6.2            | CsH$_{18}$ | 1678-92-8 |
| 20    | Butylcyclohexane | C1CC1CCC1    | 210.84      | 9.62        | -0.05       | 10 | 7.3         | 1,5   | 6.8            | 7.6            | CsH$_{20}$ | 1678-93-9 |
| 21    | Decylcyclohexane | C1CCCCCCCCC1CCC1 | 332.16    | 14.95       | 0.24        | 16 | 14.4        | 5     | 14.3           | 15.2           | CsH$_{32}$ | 1795-16-0 |
| 22    | Dodecylcyclohexane | C1CCCCCCCCC1CCC1 | 371.81    | 16.81       | 0.29        | 18 | 17.5        | 5     | 15.6           | 16.7           | CsH$_{36}$ | 1795-17-1 |
| 23    | 1-Octene       | C1CCCCC=C     | 192.38      | 18.77       | 3.59        | 8  | 3.9         | 6     | 3.4            | 3.4            | CsH$_{16}$ | 111-66-0 |
| 24    | 1-Decene       | C1CCCCC=C     | 234.26      | 20.46       | 3.17        | 10 | 5.5         | 6     | 5.7            | 6.2            | CsH$_{20}$ | 872-05-9 |
| 25    | 1-Dodecane     | C1CCCCC=C     | 274.46      | 22.29       | 3.31        | 12 | 8.1         | 6     | 8.1            | 8.8            | CsH$_{24}$ | 112-41-4 |
| 26    | 1-Octadecane   | C1CCCCCCCCC=C | 394.66      | 27.89       | 3.6         | 18 | 14.2        | 6     | 15.1           | 15.1           | CsH$_{36}$ | 112-88-9 |
| 27    | Dipropylether  | CCCCCC       | 171.36      | 26.39       | 20.59       | 6  | 0.4         | [g]   | 1.3            | 1.1            | CsH$_{2}O$ | 111-43-3 |
| 28    | Dibutylether   | CCCCCC       | 211.38      | 28.47       | 21.23       | 8  | 3.0         | 5,7,8 | 2.6            | 2.1            | CsH$_{4}O$ | 142-96-1 |
| No. | Compound               | Molecular Formula | Mass, m/z (g/mol) | pK<sub>a</sub>  |
|-----|------------------------|-------------------|-------------------|---------------|
| 29  | Dipentylether          | CCCCCCCCCCCC      | 251.19 29.57 20.44 10 4.2 | 7 4.5 4.8 C<sub>10</sub>H<sub>20</sub>O 693-65-2 |
| 30  | Dihexylether           | CCCCCCCCCCC       | 291.31 31.96 21.34 12 6.2 | 7 6.2 6.0 C<sub>12</sub>H<sub>26</sub>O 112-58-3 |
| 31  | Diheptylether          | CCCCCCCCCCCCC     | 331.49 32.86 20.36 14 8  | 7 7.7 8.7 C<sub>14</sub>H<sub>30</sub>O 629-64-1 |
| 32  | Diododecylen         | CCCCCCCCCCCCCC     | 371.28 35.31 21.42 16 10.3 | 7 9.8 10.1 C<sub>16</sub>H<sub>34</sub>O 629-82-3 |
| 33  | 1-Chlorodecane        | CCCCCCCCCCCl      | 257.74 29.94 3.76 10 3.5 | 6,9 3.7 4.2 C<sub>10</sub>H<sub>20</sub>Cl 1002-69-3 |
| 34  | 1-Chlorododecane      | CCCCCCCCCCCCl     | 297.96 31.52 3.82 12 5.6 | 6,9 5.8 6 C<sub>12</sub>H<sub>26</sub>Cl 112-52-7 |
| 35  | 1-Chlorotetradecane   | CCCCCCCCCCCCCCl   | 338.17 33.58 3.92 14 8  | 6,9 7.6 7.2 C<sub>14</sub>H<sub>30</sub>Cl 2425-54-9 |
| 36  | 1-Chlorohexadecane    | CCCCCCCCCCCCCCcl  | 378.22 35.15 3.98 16 9.8 | 6,9 9.9 9.7 C<sub>16</sub>H<sub>34</sub>Cl 4860-03-1 |
| 37  | Butylbenzene          | CCCCCC1ccccc1     | 197.06 25.15 0.54 10 0.4 | 4,6 -0.7 -0.3 C<sub>6</sub>H<sub>14</sub> 104-51-8 |
| 38  | Octylbenzene          | CCCCCCCCCC1ccccc1 | 280.39 29.6 0.94 14 4  | 1,5 3.5 5.2 C<sub>8</sub>H<sub>18</sub> 2189-60-8 |
| 39  | Decylbenzene          | CCCCCCCCCCCCCccc1 | 320.42 31.55 1.1 16 6  | 1,5 5.9 6.6 C<sub>10</sub>H<sub>22</sub> 104-72-3 |
| 40  | Dodecylbenzene        | CCCCCCCCCCCCCC1ccc1 | 359.87 33.08 1.12 18 7.8 | 2,4,5 8.3 8.6 C<sub>12</sub>H<sub>26</sub> 123-01-3 |
| 41  | 1-Octyne              | CCCCCC#C          | 188.77 33.22 1.53 8 8  | 2 0.6 1.5 C<sub>6</sub>H<sub>14</sub> 629-05-0 |
| 42  | 1-Decyne              | CCCCCCCCC#C       | 230.44 35.32 1.68 10 0.1 | 6 0.4 0.2 C<sub>10</sub>H<sub>18</sub> 764-93-2 |
| 43  | 1-Dodecyne            | CCCCCCCCCC#C      | 265.87 36.64 1.52 12 2  | 6 2 1.4 C<sub>12</sub>H<sub>22</sub> 765-03-7 |
| 44  | 1-Tetradecyne         | CCCCCCCCCCCC#C    | 310.75 38.72 1.96 14 3.9 | 6 3.5 3.5 C<sub>14</sub>H<sub>26</sub> 765-10-6 |
| 45  | Ethyl dcanoato        | CCCCCCCCCC(=O)OCC | 294.96 53.81 33.56 12 2.1 | 10,11 1.9 1.5 C<sub>12</sub>H<sub>22</sub>O<sub>2</sub> 110-38-3 |
| 46  | Ethyl dodecanoato     | CCCCCCCCCC(=O)OCC | 333.62 54.56 32.64 14 3.8 | 11 3.7 4.7 C<sub>14</sub>H<sub>26</sub>O<sub>2</sub> 106-33-2 |
| 47  | Ethyl myristate       | CCCCCCCCCC(=O)OCC | 375.02 56.48 32.88 16 5.2 | 11 5.5 6.5 C<sub>16</sub>H<sub>32</sub>O<sub>2</sub> 124-06-1 |
| 48  | Ethyl palmitate       | CCCCCCCCCCCCC(=O)OCC | 415.24 59.07 33.25 18 6.8 | 11 7.2 8.2 C<sub>18</sub>H<sub>36</sub>O<sub>2</sub> 628-97-7 |
| 49  | Octanenitrile         | CCCCCCCCC#N       | 204.22 49.16 22.76 8 1.7 | 7 -1.8 -1.7 C<sub>10</sub>H<sub>20</sub>N 127-12-9 |
| 50  | Decanenitrile         | CCCCCCCCC#N       | 243.52 50.82 22.88 10 0.5 | 7 -0.4 -0.5 C<sub>10</sub>H<sub>20</sub>N 1975-78-6 |
| 51  | Dodecanenitrile       | CCCCCCCCC#N       | 283.99 52.55 23.07 12 0.4 | 7 1.4 1.4 C<sub>12</sub>H<sub>24</sub>N 2437-25-4 |
| 52  | 2-Octanone            | CCCCCCCCC(=O)     | 201.89 48.96 35.68 8 3.4 | 7 -3.2 -2.6 C<sub>10</sub>H<sub>20</sub>O 111-13-7 |
| 53  | 2-Decanone            | CCCCCCCCCC(=O)    | 242.35 50.66 35.46 10 2.1 | 7 -2.1 -2 C<sub>10</sub>H<sub>20</sub>O 693-54-9 |
| 54  | 2-Undecanone          | CCCCCCCCCC(=O)    | 261.87 51.41 35.83 11 -1.3 | 7 -0.7 -1 C<sub>10</sub>H<sub>20</sub>O 112-12-9 |
| 55  | 2-Dodecanone          | CCCCCCCCCC(=O)    | 282.29 52.17 35.35 12 0.6 | 7 -0.8 0.8 C<sub>12</sub>H<sub>24</sub>O 6175-49-1 |
| 56  | 2-methylpentane       | CCCCCC(C)C        | 154.78 8.47 -0.17 6 6.4 | 9 5.3 5.5 C<sub>6</sub>H<sub>14</sub> 107-83-5 |
| 57  | 2,2,4-trimethylpentane| CC(C)CC(C)(C)C    | 181.75 10.89 -0.13 8 8.3 | 9 6.4 5.6 C<sub>8</sub>H<sub>18</sub> 540-84-1 |
| 58  | 2,3-dimethylbutane    | CC(C)(C)C        | 149.31 8.56 -0.25 6 4.8 | 9 4.7 5 C<sub>6</sub>H<sub>14</sub> 79-29-8 |
| 59  | 3-methylpentane       | CCCCCC(C)C        | 152.61 8.3 -0.28 6 5.2 | 9 5.3 5.5 C<sub>6</sub>H<sub>14</sub> 96-14-0 |
| 60  | Isododecane           | CC(C)(C)CC(C)CC(C)C | 244.43 14.90 -0.11 12 11.7 | 9 11.5 9.8 C<sub>12</sub>H<sub>26</sub> 13475-82-6 |
| 61-Test | Hemisqualane | CCC(C)CCCC(C)CCCC(C)C | 320.82 | 16.86 | -0.11 | 15 | 14.8 | [g] | 14.9 | 14.6 | C_{58}H_{32} | 3891-98-3 |
|---------|-------------|------------------------|---------|-------|--------|----|------|-----|------|------|-------------|-------------|
| 62      | Isohexadecane| CC(CC(C)(C)CC(C)CC(C)C) | 303.13 | 18.48 | -0.11 | 16 | 13.9 | [g] | 15.6 | 12.1 | C_{58}H_{34} | 4390-04-9   |
| 63      | Pristane    | CC(C)CCCC(C)CCCC(C)CCCC(C)C | 396.57 | 20.64 | 0.06 | 19 | 17.6 | [g] | 17.9 | 18.3 | C_{58}H_{40} | 1921-70-6   |
| 64      | 1,2-Dimethylcyclohexane | CC1CCCCCC1C | 168.83 | 8.2 | -0.05 | 8 | 3.3 | 1 | 4.6 | 4.3 | C_{8}H_{16} | 583-57-3   |
| 65      | 1,4-Dimethylcyclohexane | CC1CCC(C)CC1 | 171.39 | 8.45 | 0.03 | 8 | 4.6 | 1 | 4.6 | 4.4 | C_{16}H_{30} | 589-90-2   |
| 66      | Isopropylcyclohexane | CC(C)C1CCCCC1C | 185.48 | 9.03 | -0.13 | 9 | 5.3 | 1 | 5.4 | 5.1 | C_{18}H_{30} | 696-29-7   |
| 67      | Cyclooctane | C1CCCCCCCCC1 | 163.91 | 7.28 | -0.04 | 8 | 4.1 | 6 | 3.8 | 3.4 | C_{16}H_{26} | 292-64-8   |
| 68      | Cyclodecane  | C1CCCCCCCCC1 | 192.46 | 9.22 | -0.38 | 10 | 5.6 | 6 | 6.6 | 5.2 | C_{20}H_{40} | 293-96-9   |
| 69      | cis-Decalin  | C1CC[C@@H]2CCCCC[C@@H]2C1 | 188.05 | 7.8 | 0.05 | 10 | 5.3 | 6 | 3.3 | 4.2 | C_{16}H_{34} | 91-17-8    |
| 70      | Pinane      | CC1CCC2CC1C2(C)C | 186.28 | 10.46 | -0.38 | 10 | 4.1 | 12 | 4.3 | 5 | C_{16}H_{34} | 473-55-2   |
| 71      | p-Menthanne | CC1CCC(C)C(C)C | 200.29 | 10.29 | -0.15 | 10 | 5.8 | 12 | 6.8 | 6 | C_{20}H_{42} | 99-82-1    |
| 72      | Cyclohexene | C1CCC=C=CC1 | 132.04 | 14.28 | 3.82 | 6 | -1.2 | 1 | -0.8 | -1.1 | C_{9}H_{10} | 110-83-8   |
| 73      | 1-Methyl-1-cyclohexene | CC1=CCCC1C | 150.83 | 14.02 | 3.11 | 7 | 0 | 13 | 1.1 | 0.7 | C_{9}H_{12} | 591-49-1   |
| 74      | 4-Methyl-1-cyclohexene | CC1C=CCCC1C | 149.89 | 15.41 | 3.72 | 7 | 0.1 | 13 | -0.2 | -0.2 | C_{9}H_{12} | 591-47-9   |
| 75      | 3-Methyl-1-cyclohexene | CC1CCCC=C1 | 150.06 | 15.25 | 3.73 | 7 | -0.5 | 13 | 0.5 | 0 | C_{9}H_{12} | 591-48-0   |
| 76      | cis-Cyclooctene | C1CCGCC=C=CC1 | 160.62 | 14.36 | 3.25 | 8 | 1.5 | 6 | 1.2 | 0.4 | C_{10}H_{14} | 931-87-3   |
| 77      | 1,3-Cyclohexadiene | C1C=CC=C1 | 127.77 | 20.68 | 3.54 | 6 | -3.1 | 13 | -3.6 | -3.5 | C_{8}H_{8} | 592-57-4   |
| 78      | 1,4-Cyclohexadiene | C1C=C=CC1 | 128.11 | 21.47 | 4.65 | 6 | -4 | 13 | -3.5 | -3.4 | C_{8}H_{8} | 628-41-1   |
| 79      | 2,5-Norbornadiene | C1CC2=C=CC=C2 | 133.62 | 22.95 | 3.78 | 7 | -3.2 | 13 | -2.5 | -4.2 | C_{8}H_{8} | 121-46-0   |
| 80      | 2,6,10-trimethylundeca-2,6-diene | CC(C)CCC=C(C)CCCC=C=CC1C | 291.67 | 28.17 | 4.2 | 14 | 10.3 | 12 | 8.4 | 6.1 | C_{44}H_{26} | 53837-34-6 |
| 81      | p-Xylene    | C1ccc(C)c1 | 161.17 | 23.96 | 1.22 | 8 | -2.4 | 6 | -3.1 | -2.3 | C_{10}H_{10} | 106-42-3   |
| 82      | Phenyl-1-butylene | CCC#Cc1ccccc1 | 195.47 | 35.72 | -0.41 | 10 | -3.3 | 6 | -3.7 | -2.7 | C_{10}H_{10} | 622-76-4   |
| 83      | p-Cymene    | CC(C)c1ccc(C)c1 | 195.88 | 25.88 | 1.52 | 10 | -0.8 | 12 | -1.6 | -0.3 | C_{10}H_{14} | 99-87-6    |
| 84      | α-Pinenene  | CC1=CCC2CC1C2(C)C | 186.35 | 14.89 | 1.06 | 10 | 3.5 | 12 | 2 | 3.6 | C_{10}H_{16} | 80-56-8    |
| 85      | p-Menth-2-ene | CC(C)C1CCCCC(C)=C1 | 197.75 | 17.00 | 3.01 | 10 | 3.4 | 12 | 3.8 | 3.7 | C_{10}H_{18} | 138-86-3   |
| 86      | Δ-3-Carene  | CC1=CCC2(C)C2(C)C | 194.74 | 20.82 | 3.22 | 10 | 2.5 | 12 | 2 | 1.9 | C_{10}H_{16} | 13466-78-9 |
| 87      | β-Pinenene  | C1(C)C2CC(C)(C)C1C2 | 184.7 | 18.79 | 3.18 | 10 | 2.2 | 12 | 3.4 | 1.8 | C_{10}H_{16} | 127-91-3   |
| 88-Test | Limonene    | C1(C)=C1CCCC(C)=CC1 | 197.8 | 24.85 | 5.78 | 10 | 1.8 | 12 | 1.1 | 1.1 | C_{10}H_{16} | 5256-65-5   |
| 89      | γ-Terpine   | CC(C)C1=CCCC(C)=CC1C | 199.76 | 22.82 | 4.25 | 10 | 1.7 | 12 | 0.8 | 1.5 | C_{10}H_{16} | 99-85-4    |
| 90      | α-Terpine   | CC(C)C1=CC=C(C)C1 | 199.72 | 23.33 | 4.58 | 10 | 1.2 | 12 | 1.3 | 1.4 | C_{10}H_{16} | 99-86-5    |
| 91      | Terpinolene | CC(C)=C1CCCC(C)=CC1C | 198.78 | 22.97 | 3.89 | 10 | 0.7 | 12 | 2 | 1.6 | C_{10}H_{16} | 13877-93-5 |
|   | Longifolene | Caryophyllene | Diisopropylether | 1,4-diproxybutane | 1,2-diproxyethane | Decyl butyrate | Hexyl octanoate | Butyl dodecanoate | Myristyl propionate |
|---|-------------|---------------|-----------------|-------------------|-------------------|---------------|----------------|------------------|-------------------|
| 92 | C(1)CCCC2(C)C3CC(C13)C2=C | C(1)C=CCCC=CC(=C)C2CC(C)C2CCC1 | CC(C)OCC(C)C | CCCOCOCOCOCOC | CCCOCOCOCOC | CCCOCOCOCOC | CCCOCOCOC | CCCOCOCOCOC | CCCOCOCOCOC |
| 93-Test | | | | | | | | | |
| 94 | Diisopropylether | CC(C)OCC(C)C | | CCCOCOCOCOCOC | CCCOCOCOCOCOC | | | | |
| 95 | 1,4-diproxybutane | CCCOCOCOCOCOC | | | | | | | |
| 96 | 1,2-diproxyethane | CCCOCOCOCOCOC | | | | | | | |
| 97 | 1,2-dibutoxyethane | CCCOCOCOCOCOC | | | | | | | |
| 98 | Decyl butyrate | CCCOCOCOCOC(C)=OCC | | | | | | | |
| 99 | Hexyl octanoate | CCC[C:1]CCCC(C)=OCC | | | | | | | |
| 100 | Butyl dodecanoate | CCCOCOCOCOCOC(C)=OCC | | | | | | | |
| 101 | Myristyl propionate | CCCOCOCOCOCOC(C)=OCC | | | | | | | |
| 102-Test | Octyl octanoate | CCCOCOCOCOC(C)=OCC | | | | | | | |
| 103-Test | Isopropyl myristate | CCCOCOCOCOC(C)=OCC(C)C | | | | | | | |
| 104 | Isoamyl laureate | CCCOCOCOCOC(C)=OCC(C)C | | | | | | | |
| 105 | Hexyl dodecanoate | CCCOCOCOCOCOC | | | | | | | |
| 106 | Ethyl oleate | CCCOCOCOCOC(C)=OCC | | | | | | | |
| 107 | Menthone | CC(C)C1CC(CC)CC1=O | | | | | | | |
| 108 | Eucalyptol | CC12CC(CC1)C(C)O2 | | | | | | | |
| 109-Test | Rose oxide | CC1CCOC(C)C=CC(C)C=CC | | | | | | | |
| 110 | D-Carvone | CC(C)=C[C@H]1CC=C(C)C(C)=O | | | | | | | |
| 111 | Hexyl methacrylate | CCCOCOCOC(C)=OCC(C)=C | | | | | | | |
| 112 | Menthol acetate | CC(C)C1CC(CC)CC1OCC(C)=O | | | | | | | |
| 113 | Citronellyl acetate | CC(C)COC(C)=OCC=C(C)C | | | | | | | |
| 114 | Geranyl acetate | CC(C)=OCC(C)=OCC(C)=C(C)C | | | | | | | |
| 115 | α-Damascone | CC(C)=C(C)=C(C)=C(C)C(C)C | | | | | | | |
| 116-Test | Linalyl acetate | CC(C)=CCCC(C)(OC)=C(C)=C | | | | | | | |
| 117-Test | B-lonone | CC(C)=OCC(C)=C(C)C1CC(C)C | | | | | | | |
| 118 | Methyl dihydrojasmonate | CCC1C(CC1)=OC(C)=C(C)C | | | | | | | |
| 119 | Ethylene brassylate | O=C1CCCC1CCCC1=OCCOCC1 | | | | | | | |
| 120 | Methyl cedryl ether | O=C[C@H][CC23][C@H][C(C)(C)C=C[C@H][CC23][C@H][C(C)]C | | | | | | | |
| 121 | Ambretolid | O=C1CCCC1CCCC1COC | | | | | | | |

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
|   | $M_1^X$, expressed in Å$^2$, is equal to the whole surface area of molecule $X$; $M_2^X$, expressed in e°Å$^{-2}$, reflects the polarity of molecule $X$; $M_3^X$, expressed in e°Å$^{-4}$, reflects |   |   |   |   |   |   |   |   |
|   |   |   |   |   |   |   |   |   |   |

[a] $M_1^X$, expressed in Å$^2$, is equal to the whole surface area of molecule $X$; [b] $M_2^X$, expressed in e°Å$^{-2}$, reflects the polarity of molecule $X$; [c] $M_3^X$, expressed in e°Å$^{-4}$, reflects
the electrostatic asymmetry of molecule X; [g] LOO estimated EACN values with the GM-5N model having the best LOO score (out of five) for the 111 molecules of the training set and EACN estimated values with the GM-5N model having the best VLOO score average (out of ten) for the 10 molecules of the test set; [h] LOO estimated EACN values with the NN-6N model having the best LOO score (out of five) for the 111 molecules of the training set and EACN predicted values with the NN-6N model having the best VLOO score average (out of ten) for the 10 molecules of the test set; [i] EACN value equal to number of carbon atoms by definition; [j] This work. For the test molecules the given data can be obtained in running the Docker image "espclign/eacn:demo", see section "graph machine and neural network results with Docker".

Table S2. Names, SMILES notations, three first α-moments (different from zero) calculated with COSMO-RS, number of carbon atoms, experimental EACN and average EACN values determined from the Fish-tail-temperature T° reported in the literature for ternary systems C6E6Oil/Water, estimated EACN values with the GM-5N and NN-6N models, molecular formulas and CAS RN for the 111 molecules of the homologous series.

| Entry | Compound   | SMILES code          | $M_0^{x}$ [a] | $M_2^{x}$ [b] | $M_4^{x}$ [c] | n | EACN$_{exp}$ | Ref. | EACN$_{exp}$ [d] | EACN$_{nn}$ [e] | MF   | CAS RN  |
|-------|------------|----------------------|---------------|---------------|---------------|---|-------------|------|----------------|----------------|------|---------|
| 1     | hex-1-ene  | CCCCC=C              | 154.47        | 17.08         | 3.03          | 6  | -           | -    | 1.7           | 1.0            | C6H12 | 592-41-6 |
| 2     | hept-1-ene | CCCCCCC=C            | 174.54        | 17.93         | 3.06          | 7  | -           | -    | 2.8           | 2.6            | C7H14 | 592-76-7 |
| 3     | oct-1-ene  | CCCCCCCC=C           | 192.38        | 18.77         | 3.59          | 8  | 3.96        | 6    | 3.6           | 3.7            | C8H16 | 111-66-0 |
| 4     | non-1-ene  | CCCCCCCCC=C          | 213.71        | 19.66         | 3.2           | 9  | -           | -    | 4.8           | 5.0            | C9H18 | 124-11-8 |
| 5     | dec-1-ene  | CCCCCCCCCCC=C        | 234.26        | 20.46         | 3.17          | 10 | 5.57        | 6    | 5.6           | 6.2            | C10H20 | 872-05-9 |
| 6     | undec-1-ene| CCCCCCCCCCCC=C       | 253.11        | 21.06         | 2.99          | 11 | -           | -    | 6.8           | 7.3            | C11H22 | 821-95-4 |
| 7     | dodec-1-ene| CCCCCCCCCCCCC=C      | 274.46        | 22.29         | 3.31          | 12 | 8.1         | 6    | 7.9           | 8.3            | C12H24 | 112-41-4 |
| 8     | tridec-1-ene| CCCCCCCCCCCCCC=C    | 293.79        | 23.18         | 3.23          | 13 | -           | -    | 9.0           | 9.5            | C13H26 | 2437-56-1 |
| 9     | tetradec-1-ene| CCCCCCCCCCCCCCC=C | 314.61        | 24.11         | 3.38          | 14 | -           | -    | 10.1          | 10.8           | C14H28 | 1120-36-1 |
| 10    | pentadec-1-ene| CCCCCCCCCCCCCCCC=C | 334.71        | 25.24         | 3.47          | 15 | -           | -    | 11.2          | 12.0           | C15H30 | 13360-61-7 |
| 11    | hexadec-1-ene| CCCCCCCCCCCCCCCCC=C | 355.28        | 25.93         | 3.5           | 16 | -           | -    | 12.3          | 13.0           | C16H32 | 629-73-2 |
| 12    | heptadec-1-ene| CCCCCCCCCCCCCCCCC=C | 374.83        | 27.05         | 3.56          | 17 | -           | -    | 13.3          | 13.7           | C17H34 | 6765-39-5 |
| 13    | octadec-1-ene| CCCCCCCCCCCCCCCCCC=C | 394.66       | 27.89         | 3.6           | 18 | 14          | 6    | 14.4          | 14.4           | C18H36 | 112-88-9 |
| 14    | 1-chlorohexane | CC[C:1]CCCCl | 178.83       | 26.47         | 4.0           | 6  | -           | -    | -0.6          | 1.2            | C6H12Cl | 544-10-5 |
| 15    | 1-chloroheptane| CCCCCCCCl          | 198.14       | 27.2          | 3.78          | 7  | -           | -    | 0.6           | 1.8            | C7H15Cl | 629-06-1 |
| 16    | 1-chlorooctane| CCCCCCCCCCl        | 219.07       | 28.31         | 4.13          | 8  | -           | -    | 1.8           | 2.4            | C8H16Cl | 111-85-3 |
| 17    | 1-chlorononane| CCCCCCCCCCCl       | 238.09       | 29.01         | 3.89          | 9  | -           | -    | 2.5           | 3.0            | C9H18Cl | 2473-01-0 |
| 18    | 1-chlorodecane| CCCCCCCCCCCCl      | 257.74       | 29.94         | 3.76          | 10 | 3.57        | 6.9  | 3.6           | 3.7            | C10H20Cl | 1002-69-3 |
| 19    | 1-chloroundecane| CCCCCCCCCCCCl     | 278.35       | 30.89         | 3.8          | 11 | -           | -    | 4.5           | 4.6            | C11H22Cl | 2473-03-2 |
| 20    | 1-chlorododécane| CCCCCCCCCCCCCCl    | 297.96       | 31.52         | 3.82          | 12 | 5.6         | 6.9  | 5.7           | 5.7            | C12H24Cl | 112-52-7 |
| 21    | 1-chlorotridecane| CCCCCCCCCCCCCCl   | 318.51       | 32.66         | 3.9          | 13 | -           | -    | 6.7           | 6.8            | C13H26Cl | 822-13-9 |
|   | Compound                          | Molecular Formula | Mass | Mol. % | Value   | Compound                          | Molecular Formula | Mass | Mol. % | Value   |
|---|-----------------------------------|-------------------|------|--------|--------|-----------------------------------|-------------------|------|--------|--------|--------|
| 22| 1-chlorotetradecane               | CCCCCCCCCCCCCCIC | 338.17 | 33.58 | 3.92  | 14 | 8 | 6.9 | 7.8 | 7.9 | C_{14}H_{29}Cl | 2425-54-9 |
| 23| 1-chloropentadecane               | CCCCCCCCCCCCCCIC | 358.14 | 34.49 | 3.93  | 15 | - | - | 8.9 | 8.7 | C_{15}H_{31}Cl | 4862-03-7 |
| 24| 1-chlorohexadecane                | CCCCCCCCCCCCCCCCI | 378.22 | 35.15 | 3.98  | 16 | 9.8 | 6.9 | 9.8 | 9.6 | C_{16}H_{33}Cl | 4860-03-1 |
| 25| 1-chloroheptadecane               | CCCCCCCCCCCCCCCCI | 398.86 | 36.28 | 4.05  | 17 | - | - | 10.9 | 10.1 | C_{17}H_{35}Cl | 62016-75-5 |
| 26| 1-chlorooctadecane                | CCCCCCCCCCCCCCCCI | 418.41 | 37.03 | 4.09  | 18 | - | - | 11.7 | 11.1 | C_{18}H_{37}Cl | 3386-33-2 |
| 27| hex-1-ynel                        | C(C[1]CCCC#C     | 150.44 | 31.68 | 1.42  | 6 | - | - | -3.9 | -2.3 | C_{8}H_{10} | 693-02-7 |
| 28| hept-1-ynel                       | C(C[1]CCCC#C     | 170.69 | 32.6 | 1.53  | 7 | - | - | -2.5 | -1.9 | C_{9}H_{12} | 628-71-7 |
| 29| oct-1-ynel                        | CCCCCCCCC#C       | 188.77 | 33.22 | 1.53  | 8 | -1.8 | 6 | -1.3 | -1.8 | C_{10}H_{14} | 629-05-0 |
| 30| non-1-ynel                        | C(C[1]CCCC#C     | 209.6 | 33.93 | 1.37  | 9 | - | - | -1.0 | -1.0 | C_{10}H_{16} | 3452-09-3 |
| 31| dec-1-ynel                        | CCCCCCCCC#C       | 230.44 | 35.32 | 1.68  | 10 | 0.1 | 6 | 0.2 | 0.1 | C_{11}H_{18} | 764-93-2 |
| 32| undec-1-ynel                      | CCCCCCCCC#C       | 249.55 | 35.61 | 1.43  | 11 | - | - | 1.5 | 1.4 | C_{11}H_{20} | 2243-98-3 |
| 33| dodec-1-ynel                      | CCCCCCCCC#C       | 265.87 | 36.64 | 1.52  | 12 | 2.0 | 6 | 2.0 | 1.8 | C_{12}H_{22} | 765-03-7 |
| 34| tridec-1-ynel                     | CCCCCCCCC#C       | 290.57 | 38.01 | 1.86  | 13 | - | - | 3.2 | 3.2 | C_{13}H_{24} | 26186-02-7 |
| 35| tetradec-1-ynel                   | CCCCCCCCC#C       | 310.75 | 38.72 | 1.96  | 14 | 3.9 | 6 | 3.8 | 3.8 | C_{14}H_{26} | 765-10-6 |
| 36| pentadec-1-ynel                   | CCCCCCCCC#C       | 330.13 | 39.49 | 1.74  | 15 | - | - | 4.9 | 4.4 | C_{15}H_{28} | 765-13-9 |
| 37| hexadec-1-ynel                    | CCCCCCCCC#C       | 350.9 | 40.53 | 2.02  | 16 | - | - | 5.6 | 4.8 | C_{16}H_{30} | 629-74-3 |
| 38| heptadec-1-ynel                   | CCCCCCCCC#C       | 371.5 | 41.48 | 2.09  | 17 | - | - | 6.6 | 5.1 | C_{17}H_{32} | 26186-00-5 |
| 39| octa dec-1-ynel                   | CCCCCCCCC#C       | 391.04 | 42.25 | 2.09  | 18 | - | - | 7.4 | 5.4 | C_{18}H_{34} | 629-89-0 |
| 40| hexan-2-one                       | C(C[1]CC(O)=O     | 162.69 | 47.19 | 35.22 | 6 | - | - | -4.9 | -4.5 | C_{8}H_{12}O | 591-78-6 |
| 41| heptan-2-one                      | CCCCCC(O)=O       | 182.02 | 47.29 | 34.83 | 7 | - | - | -3.8 | -3.7 | C_{9}H_{14}O | 110-43-0 |
| 42| octan-2-one                       | CCCCCC(O)=O       | 201.89 | 48.96 | 35.68 | 8 | -3.4 | 7 | -3.4 | -3.2 | C_{10}H_{16}O | 111-13-7 |
| 43| nonan-2-one                       | CCCCCC(O)=O       | 221.83 | 48.89 | 34.74 | 9 | - | - | -2.2 | -2.4 | C_{11}H_{18}O | 821-55-6 |
| 44| decan-2-one                       | CCCCCC(O)=O       | 242.35 | 50.66 | 35.46 | 10 | -2.1 | 7 | -2.2 | -1.8 | C_{12}H_{20}O | 693-54-9 |
| 45| undecan-2-one                     | CCCCCC(O)=O       | 262.46 | 51.58 | 35.88 | 11 | -1.3 | 7 | -1.0 | -1.2 | C_{13}H_{22}O | 112-12-9 |
| 46| dodecan-2-one                     | CCCCCC(O)=O       | 282.29 | 52.17 | 35.35 | 12 | -0.6 | 7 | -0.8 | -0.1 | C_{14}H_{24}O | 6175-49-1 |
| 47| tridecan-2-one                    | CCCCCC(O)=O       | 302.53 | 52.44 | 34.62 | 13 | - | - | 0.4 | 1.9 | C_{15}H_{26}O | 593-08-8 |
| 48| tetradecan-2-one                  | CCCCCC(O)=O       | 322.06 | 54.01 | 35.42 | 14 | - | - | 0.5 | 3.3 | C_{16}H_{28}O | 2345-27-9 |
| 49| pentadecan-2-one                  | CCCCCC(O)=O       | 342.6 | 54.37 | 34.66 | 15 | - | - | 1.6 | 5.1 | C_{17}H_{30}O | 2345-28-0 |
| 50| hexadecan-2-one                   | CCCCCC(O)=O       | 362.4 | 55.79 | 35.63 | 16 | - | - | 1.9 | 6.3 | C_{18}H_{32}O | 18787-63-8 |
| 51| heptadecan-2-one                  | CCCCCC(O)=O       | 383.45 | 57.07 | 35.93 | 17 | - | - | 3.0 | 7.5 | C_{19}H_{34}O | 2922-51-2 |
| 52| octadecan-2-one                   | CCCCCC(O)=O       | 402.65 | 57.78 | 35.77 | 18 | - | - | 3.2 | 8.5 | C_{20}H_{36}O | 7373-13-9 |
| 53| cyclohexane                       | C1CCCCC1          | 135.76 | 5.88 | -0.12 | 6 | 2.1 | 1-4 | 2.0 | 2.4 | C_{6}H_{12} | 110-82-7 |
| 54 | methylcyclohexane | CC1CCCCC1 | 153.55 | 7.16 | -0.05 | 7 | 3.2 | 1 | 2.9 | 3.3 | C_{12}H_{14} | 108-87-2 |
| 55 | ethylcyclohexane | CCC1CCCCC1 | 171.53 | 7.91 | -0.12 | 8 | 4.2 | 1,5 | 4.4 | 4.6 | C_{12}H_{16} | 1678-91-7 |
| 56 | Propylcyclohexane | CCC1CC1CCC1 | 191.60 | 8.79 | -0.08 | 9 | 5.9 | 1,5 | 5.7 | 6.1 | C_{14}H_{18} | 1678-92-8 |
| 57 | butylcyclohexane | CCCCC1CCCCC1 | 210.84 | 9.64 | -0.04 | 10 | 7.3 | 1,5 | 7.1 | 7.4 | C_{16}H_{20} | 1678-93-9 |
| 58 | pentylcyclohexane | CCCCC1CCCCC1 | 230.87 | 10.52 | 0.01 | 11 | - | - | 8.2 | 8.7 | C_{15}H_{22} | 4292-92-6 |
| 59 | hexylcyclohexane | CCCCCC1CCCCC1 | 251.83 | 11.39 | 0.03 | 12 | - | - | 9.3 | 9.9 | C_{16}H_{24} | 4292-75-5 |
| 60 | heptylcyclohexane | CCCCCC1CCCCC1 | 267.31 | 12.16 | -0.03 | 13 | - | - | 10.4 | 10.8 | C_{17}H_{26} | 5617-41-4 |
| 61 | octylcyclohexane | CCCCCCCC1CCCCC1 | 291.22 | 13.25 | 0.13 | 14 | - | - | 12.1 | 12.0 | C_{18}H_{28} | 1795-15-9 |
| 62 | nonylcyclohexane | CCCCCCCC1CCCCC1 | 311.7 | 14.12 | 0.18 | 15 | - | - | 13.2 | 13.3 | C_{19}H_{30} | 2883-02-5 |
| 63 | decylcyclohexane | CCCCCCC1CCCCC1 | 332.16 | 14.95 | 0.24 | 16 | 14.4 | 5 | 14.5 | 14.6 | C_{20}H_{32} | 1795-16-0 |
| 64 | undecylcyclohexane | CCCCCC1CCCC1CCC1 | 352.12 | 15.76 | 0.3 | 17 | - | - | 15.6 | 15.9 | C_{21}H_{34} | 54105-66-7 |
| 65 | dodecylcyclohexane | CCCCCC1CCCC1CCC1 | 371.81 | 16.81 | 0.29 | 18 | 17.5 | 5 | 17.1 | 17.1 | C_{22}H_{36} | 1795-17-1 |
| 66 | 1-propanol | CCCOCC | 171.36 | 26.39 | 20.59 | 6 | 0.4 | || | 0.7 | 0.7 | C_{6}H_{14}O | 114-43-3 |
| 67 | 1-propanol | CCCOCC | 191.74 | 28.16 | 21.83 | 7 | - | - | 1.7 | 1.5 | C_{7}H_{16}O | 3073-92-5 |
| 68 | 1-butanol | CCCOCOC | 211.38 | 28.47 | 21.23 | 8 | 3.0 | 5,7,8 | 2.6 | 2.7 | C_{8}H_{18}O | 142-96-1 |
| 69 | 1-butanol | CCCOCOC | 231.24 | 29.01 | 20.98 | 9 | - | - | 3.5 | 3.7 | C_{9}H_{20}O | 18636-66-3 |
| 70 | 1-pentanol | CCCOCCOC | 251.19 | 29.57 | 20.44 | 10 | 4.2 | 7 | 4.4 | 4.5 | C_{10}H_{22}O | 693-65-2 |
| 71 | 1-pentanol | CCCOCCOC | 271.23 | 30.7 | 21.06 | 11 | - | - | 5.3 | 5.2 | C_{11}H_{24}O | 32357-83-8 |
| 72 | 1-hexanol | CCCOCCOC | 291.31 | 31.96 | 21.34 | 12 | 6.2 | 7 | 6.2 | 6.0 | C_{12}H_{26}O | 112-58-3 |
| 73 | 1-hexanol | CCCOCCOC | 311.43 | 32.38 | 20.97 | 13 | - | - | 7.1 | 7.0 | C_{13}H_{28}O | 7289-40-9 |
| 74 | 1-heptanol | CCCOCCOC | 331.49 | 32.86 | 20.36 | 14 | 8.0 | 7 | 8.0 | 8.1 | C_{14}H_{30}O | 629-64-1 |
| 75 | 1-heptanol | CCCOCCOC | 351.57 | 33.84 | 20.58 | 15 | - | - | 8.9 | 9.1 | C_{15}H_{32}O | 32357-84-9 |
| 76 | 1-octanol | CCCOCCOC | 371.28 | 35.31 | 21.42 | 16 | 10.3 | 7 | 9.8 | 10.1 | C_{16}H_{34}O | 629-82-3 |
| 77 | 1-octanol | CCCOCCOC | 391.55 | 35.95 | 21.17 | 17 | - | - | 10.6 | 11.3 | C_{17}H_{36}O | 858238-61-6 |
| 78 | 1-nonylalcohol | CCCOCCOC | 411.72 | 36.47 | 20.58 | 18 | - | - | 11.4 | 12.7 | C_{18}H_{38}O | 2456-27-1 |
| 79 | ethyl butyrate | CC[C:1]C(OCC)=O | 173.82 | 47.52 | 31.71 | 6 | - | - | -3.0 | -4.7 | C_{7}H_{14}O_{2} | 105-54-4 |
| 80 | ethyl butyrate | CC[C:1]C(OCC)=O | 194.22 | 48.26 | 32.23 | 7 | - | - | -1.9 | -3.9 | C_{8}H_{16}O_{2} | 539-82-2 |
| 81 | ethyl caproate | CCC[C:1]C(OCC)=O | 213.46 | 49.02 | 31.67 | 8 | - | - | -0.7 | -3.0 | C_{9}H_{18}O_{2} | 123-66-0 |
| 82 | ethyl caproate | CCC[C:1]C(OCC)=O | 234.15 | 50.07 | 32.26 | 9 | - | - | 0.4 | -1.9 | C_{10}H_{20}O_{2} | 123-66-0 |
| 83 | ethyl caproate | CCC[C:1]C(OCC)=O | 253.0 | 49.88 | 30.92 | 10 | - | - | 0.5 | -0.4 | C_{11}H_{22}O_{2} | 106-32-1 |
| 84 | ethyl caproate | CCC[C:1]C(OCC)=O | 274.18 | 51.89 | 32.1 | 11 | - | - | 1.0 | 0.8 | C_{11}H_{22}O_{2} | 123-29-5 |
| 85 | ethyl caproate | CCC[C:1]C(OCC)=O | 294.96 | 53.81 | 33.56 | 12 | 2.1 | 10,11 | 2.1 | 1.9 | C_{12}H_{24}O_{2} | 110-38-3 |
| No. | Name                        | Structural Formula                  | MW | MP  | LogP | S | Value  | Ref. |
|-----|-----------------------------|------------------------------------|-----|-----|------|---|--------|------|
| 86  | ethyl undecylate            | CCCCCCCCCCC(OCC)=O                | 314.51 | 54.14 | 33.04 | 13 | -      |      |
| 87  | ethyl laurate               | CCCCCCCCCCC(OCC)=O                | 333.62 | 54.56 | 32.64 | 14 | 3.8    | 11   |
| 88  | ethyl tridecanoate          | CCCCCCCCCCCC(OCC)=O               | 354.72 | 55.83 | 33.02 | 15 | -      |      |
| 89  | ethyl myristate             | CCCCCCCCCCCCCCC(OCC)=O            | 375.02 | 56.46 | 32.88 | 16 | 5.2    | 11   |
| 90  | ethyl pentadecanoate        | CCCCCCCCCCCCCCCC(OCC)=O           | 395.34 | 57.79 | 33.24 | 17 | -      |      |
| 91  | ethyl palmitate             | CCCCCCCCCCCCCCCCCC(OCC)=O         | 415.24 | 59.07 | 33.25 | 18 | 6.8    | 11   |
| 92  | capronitrile                | CCCCCC#N                          | 163.84 | 47.37 | 22.7  | 6  | -      |      |
| 93  | enanthonitrile              | CCCC[C:1]CC#N                    | 183.34 | 48.24 | 22.69 | 7  | -      |      |
| 94  | caprylonitrile              | CCCCCCCCC#N                      | 204.22 | 49.16 | 22.76 | 8  | -1.7   | 7    |
| 95  | pelargonoritrile            | CCCCCCC#N                        | 222.16 | 49.71 | 22.71 | 9  | -      |      |
| 96  | capronitrile                | CCCCCCCCCC#N                     | 243.52 | 58.02 | 22.88 | 10 | -0.5   | 7    |
| 97  | undecanenitrile             | CCCCCCCCCC#N                     | 263.88 | 51.76 | 23.03 | 11 | -      |      |
| 98  | lauronitrile                | CCCCCCCCCCCC#N                   | 283.99 | 52.55 | 23.07 | 12 | 0.4    | 7    |
| 99  | tridecanenitrile            | CCCCCCCCCCCCC#N                  | 304.6  | 53.71 | 23.08 | 13 | -      |      |
| 100 | myristonitrile              | CCCCCCCCCCC#N                    | 324.22 | 54.56 | 23.13 | 14 | -      |      |
| 101 | pentadecanenitrile          | CCCCCCCCCCCCC#N                  | 342.98 | 55.14 | 22.97 | 15 | -      |      |
| 102 | palmitonitrile              | CCCCCCCCCCCCC#N                  | 364.4  | 56.25 | 23.18 | 16 | -      |      |
| 103 | heptadecanenitrile          | CCCCCCCCCCCCCCCC#N               | 384.41 | 57.18 | 23.27 | 17 | -      |      |
| 104 | stearonitrile               | CCCCCCCCCCCCCCCC#N               | 404.51 | 57.93 | 23.31 | 18 | -      |      |
| 105 | benzene                     | C1=CC=CC=C1                      | 122.42 | 22.98 | -1.27 | 6  | -      |      |
| 106 | toluene                     | CC1=CC=CC=C1                     | 141.81 | 23.64 | 0.06  | 7  | -      |      |
| 107 | ethylbenzene                | CCC1=CC=CC=C1                    | 160.64 | 24.31 | 0.44  | 8  | -      |      |
| 108 | propylbenzene               | CCC1=CC=CC=C1                    | 180.59 | 25.28 | 0.67  | 9  | -      |      |
| 109 | butylbenzene                | CCC1=CC=CC=C1                    | 200.65 | 26.12 | 0.72  | 10 | 0.4    | 4.6  |
| 110 | pentybenzene                | CCCCCC1=CC=CC=C1                 | 220.24 | 27.00 | 0.82  | 11 | -      |      |
| 111 | hexylbenzene                | CCCCCCC1=CC=CC=C1                | 237.83 | 27.56 | 0.78  | 12 | -      |      |
| 112 | heptylbenzene               | CCCCCCC1=CC=CC=C1                | 258.5  | 28.41 | 0.80  | 13 | -      |      |
| 113 | octylbenzene                | CCCCCCCCC1=CC=CC=C1              | 278.92 | 29.37 | 0.88  | 14 | 4.0    | 1.5  |
| 114 | nonylbenzene                | CCCCCCCCCC1=CC=CC=C1             | 299.80 | 30.47 | 0.96  | 15 | -      |      |
| 115 | decylbenzene                | CCCCCCCCCCCCC1=CC=CC=C1          | 320.4  | 31.55 | 1.1   | 16 | 6.0    | 1.5  |
| 116 | undecylbenzene              | CCCCCCCCCC1=CC=CC=C1             | 339.84 | 32.10 | 1.04  | 17 | -      |      |
| 117 | dodecylbenzene              | CCCCCCCCCCCCCC1=CC=CC=C1         | 359.87 | 33.08 | 1.12  | 18 | 7.8    | 2.45 |
\[ M_0^X, \text{ expressed in } \text{Å}^2, \text{ is equal to the whole surface area of molecule } X; \]
\[ M_2^X, \text{ expressed in e}^2.\text{Å}^{-2}, \text{ reflects the polarity of molecule } X; \]
\[ M_3^X, \text{ expressed in e}^3.\text{Å}^{-4}, \text{ reflects the electrostatic asymmetry of molecule } X; \]
\[ \text{Estimated EACN values with the GM-5N model having the best VLOO score average (out of ten) for the 117 molecules of the homologous series; } \]
\[ \text{Estimated EACN values with the NN-6N model having the best VLOO score average (out of ten) for the 117 molecules of the homologous series; } \]
\[ \text{This work. For the 117 molecules the given data can be obtained in running the Docker image "espcigm/eacn:demo", see section "graph machine and neural networks results with Docker".} \]
Figure S1. Partial experimental fish plot of C₆E₄/dipropyl ether/Water-T (left) and C₆E₄/diisopropyl ether/Water-T (right) systems at Water/Oil ratio equal to 1 (w/w) used to determine EACN values.

The temperatures of the fish-tail points are 34.7 °C for dipropyl ether and 36.1 °C for diisopropyl ether. Those temperatures are reported to the calibration straight line obtained with C₆E₄/n-Alkanes/Water-T systems shown in Figure S2. Resulting EACN values are 0.4 for dipropyl ether and 0.6 for diisopropyl ether.

Figure S2. Calibration straight line obtained with C₆E₄/n-Alkanes/Water-T systems at Water/Oil ratio equal to 1 (w/w) used to determine EACN values.¹²
**Figure S3.** Evolution of experimental and estimated EACN with increasing number of carbon atoms for homologous series of molecules with various chemical functions: (left) alk-1-enes, 1-chloroalkanes, alk-1ynes and alk2-ones, (right) n-alkycyclohexanes, central ethers, ethyl alkanoates and n-alkane nitriles. The dotted and dashed lines indicate the experimental and neural network fits. Triangles (▲), diamonds (◇) and circles (○) are respectively markers for experimental, neural network predicted and graph machine predicted values.

**Figure S4.** Evolution of experimental and estimated EACN with increasing number of carbon atoms for the n-alkylbenzene series. The dotted line indicates the experimental fit. Triangles (▲), diamonds (◇) and circles (○) are respectively markers for experimental, neural network predicted and graph machine predicted values.
Installing Docker for Mac and downloading the demo image

In the following example the installation of Docker is performed with an intel version of Docker Desktop for Mac. The same operations can be done on a Mac with a M1 ARM-based system. The appropriate link to download the corresponding version is given at the end of this Section.

1) Download the Intel chip Docker application by copying and pasting the following line into a browser URL bar:

https://desktop.docker.com/mac/main/amd64/Docker.dmg?utm_source=docker&utm_medium=webreferral&utm_campaign=docs-driven-download-mac-amd64

2) After Docker installation (with administrator privileges), launch it. You can open the Docker preferences to increase the Docker allocated memory (e.g. 40 Go), and select a number of cores for the virtual machine (more or less according to the machine resources).

3) Open a terminal window and download the eacn:demo image by pasting the following line, and hitting return:

docker pull espcigm/eacn@sha256:202f92f4d55a68c96ec61fd4a73abc707103a0c6a79a670f4346a0378796821c

The image used to create containers is then downloaded. If the image is not genuine, the downloading operation will fail, and nothing will be copied on your machine.¹

The same image is used to launch either graph machine or neural network computations. The set up is now complete.

More information on Docker client installation can be obtained from the link below and from the Docker website (docker.com).

http://pubs.acs.org/doi/suppl/10.1021/acs.jcim.0c00083/suppl_file/ci0c00083_si_003.pdf

The link to install the M1 chip version of Docker is the following:

https://desktop.docker.com/mac/main/arm64/Docker.dmg?utm_source=docker&utm_medium=webreferral&utm_campaign=dd-smartbutton&utm_location=module

Installing Docker for Windows and downloading the demo image

The steps for the installation of the Docker Windows version and the demo image are given below.

1) Download the Windows version of Docker application by copying and pasting the following line into a browser URL bar:

https://desktop.docker.com/win/main/amd64/Docker%20Desktop%20Installer.exe

2) Install the Docker Desktop version 4.8.1 (or above) with administrator privileges

3) In the Dashboard opened at the beginning of the Docker installation do the following:
   a. UNCHECK the box "Use WSL 2 instead of Hyper-V (recommended)";
   b. verify that the box 'Add shortcut to Desktop' is checked.

¹ If the safe download failed or if the downloaded image has been recognized as not genuine, please send us a message, so that we can fix the problem.
4) When the installation has succeeded Docker Desktop is starting, you might then skip the tutorial. After a while, open a Powershell window (or a command prompt window) and to verify that Docker is running type the following command:

docker version

You can confirm in the lines returned that the Docker Desktop version is 4.8.1 (or above).

5) Open a terminal window, paste the following line, and hit return:

docker pull espcigm/eacn8sha256:202f92f4d55a68c96ec61fd4a73abc707103a0c6a79a670f4346a0378796821c

The image used to create containers is then downloaded.
If the image is not genuine, the downloading operation will fail, and nothing will be copied on your machine.2

The same image is used to launch either graph machine or neural network computations.
The set up is now complete.

Notes: 1) If at step 4 Docker does not start, it is because Hyper-V is probably not active on your system. You need to activate Hyper-V by typing in a powershell window the following command (as admin):

Enable-WindowsOptionalFeature -Online -FeatureName Microsoft-Hyper-V -All

then restart your machine.

2) Docker Desktop can be installed for a standard user. The user has to be a member of the docker-users group. This can be done with the Windows administration tools.

Loading and launching the Docker image

To open a container that will launch the default graph machine computations for the molecules of the test set, open a terminal window (or start a PowerShell session), and type the following line of text below (or copy and paste it), the argument demo1 being optional:

docker run -it --rm -v ~/docker:/host 2cd64516692f demo1

Note: with the Docker Desktop for Windows you will have to accept the files sharing message in the Docker popup for the docker folder located in your home folder. Simply click the 'Share it' button. This can also be customized in the File sharing window of the Docker Settings (Resources menu).

The first time this command is entered, you are asked to define a path for the shared folder as described below:

Please give the absolute path of the shared folder in the host
(for labelling output files only) [host]

Some suitable propositions for the absolute path that must be typed are given below for the Macintosh (Unix) and Windows OS, where ‘home’ is the home folder (or user folder), and ‘docker’ is the folder manually created in the home folder:

/Users/`home`/docker
C:Users\`home`\docker

If, for any reason, the folder path has been incorrectly entered, you can correct it by deleting the file “host.pth” located inside the docker folder, and give the proper path at the next launch.

At the end of the computations issued by this command, the results are written in an excel file located in a shared folder mounted inside the container. In the later command line, docker is the folder used; it is automatically

2 If the safe download failed or if the downloaded image has been recognized as not genuine, please send us a message, so that we can fix the problem.
created in the home folder at the Docker installation step (~/docker = /Users/home/docker on macOS, see below). If this folder does not exist, no results are saved.

The shared excel file is recorded in the result subdirectory of the docker directory on the host machine (~/docker/result, where ~ is a shortcut indicating the path to the connected user's home directory). The output filename is automatically incremented when the same command is issued.

Three other computations can be called with the same command line but with the arguments demo2, demo3 and demo4 instead of demo1. The explanations and the outputs of all four command lines are given in the Section 'Graph machine results with Docker'.

Notes: 1) The above run command is the minimal command; if for example, the "-v ~/docker:/host" is omitted in the command syntax, no excel file is created on the host machine, the computed results being lost when the container is deleted. In that case the path to the shared folder must also be entered for each computation.

2) The computed times reported during the demo depend on the machine used.

For more explanations on hyper-V you can use the following link: https://docs.microsoft.com/fr-fr/virtualization/hyper-v-on-windows/quick-start/enable-hyper-v

Predicting EACN for the test set

We explain hereafter the demo that describes the EACN computation for the ten test molecules with the selected complexity of the graph machine-based model; the command line for launching this demo is the following:

docker run -it --rm --v ~/docker:/host 2cd64516692f

The data file used is the file DEMOEACN.xlsx, that can be downloaded separately. This file is used to compute the EACN values of the molecules of the test set. It contains two sheets as described below:

- The first sheet named DATA, contains the training data for the 111 molecules of the training set in the DATA cell range. Every line of this sheet contains a compound name, its SMILES code, the three computed sigma-moments, the number of carbon atoms and the experimental value of the property of interest (here the dimensionless EACN). The other data present in the file, namely the entry, the molecular formula (MF), the molecular weight (MW) and the chemical abstract registry number (CAS RN) are ignored since they don't belong to the DATA range.
- The second sheet named TEST contains the data for the 10 molecules of the test set, in the TEST cell range. The same information as above is displayed in the eleven columns, but only the data present in the TEST range are taken into account.

During the training step achieved in an earlier stage, a graph machine model based on the SMILES code derived from the molecular structure was automatically generated for every molecule of the DATA sheet. All 111 models were then merged into a module that was trained with the desired property values. The parameters at the end of the training were stored to be used with a new model.

In the course of the present demo a graph machine model is generated for the 10 molecules (also inputed as their SMILES codes) of the TEST sheet. After the model constructions, the parameters saved during the training step are passed to the functions of the graph machines to predict the 10 EACN values of the test set molecules. At the end of the computation the file GM_test_5N.xlsx is written in the result folder (located also in the home folder). It can be read with Excel 2010-2021 (or more recent) or with LibreOffice 7.3.1 (or more recent).

The GM_test_5N.xlsx file has one sheet. For each molecule, the following quantities are displayed:

- the experimental value of the EACN;
- the mean estimated value of the EACN, computed in averaging the estimations produced by the 25 models that have the smallest VLOO training scores;
- the minimum estimation obtained for the 25 models that have the smallest VLOO training scores;
- the maximum estimation obtained for the 25 models that have the smallest VLOO training scores.
**Explanation of the demo command line**

The execution of the graph machine demonstration can be launched from the command line. The proposed default command line is (gm mode is invoked if demo1 is omitted):

```bash
docker run -it --rm -v ~/docker:/host 2cd64516692f demo1
```

It contains the following terms:

- "docker": calls the Docker daemon of the host machine;
- "run": launches a Docker container from the Docker image;
- "-it": opens and launches the interactive mode;
- "--rm": destroys the container at the end of the session;
- "-v ~/docker:/host": creates a volume in the container, and shares it with the ~/docker tree on the host machine;
- "2cd64516692f": identification of the Docker image launched with the run command;
- "demo1": this argument invokes the gm mode, as does demo3; if demo2 or demo4 are used, the nn mode is activated. In the former case a SMILES input is expected while a list of four descriptors is required for the later. The demo computations are made with the graph machine model and the neural network model that have the numbers of hidden neurons chosen when looking for the appropriate complexity (i.e. five and six hidden neurons respectively). After completion of the demo, the container is automatically deleted. A new demo session can be started with the same command, but within a new container.

**Other command line options**

Two subcommands get or draw can be appended to the command line instead of using demo1,...demo4.

The subcommand "get" can be passed to the demonstrator to compute the property value for a single compound using either a SMILES code input (with a GM argument) or some sigma-moment inputs (with a NN argument) as follows:

```bash
docker run -it --rm -v ~/docker:/host 2cd64516692f get (GM or NN)"id;inputs"
```

where "id" is the name of the compound and "inputs" corresponds either to a SMILES code or to a list of the four comma-separated descriptors $M_0$, $M_2$, $M_3$ and $N_C$. When a SMILES input is passed, the argument GM is used, while NN is used if the descriptors are the inputs. An excel file with the name "id".xlsx is also written in the result folder as described in the Section 'Predicting EACN for the test set'.

The subcommand "draw" can be passed to the demonstrator to generate a representation of a graph machine for a given compound. This representation is written in a svg graphic file that can be opened in any browser (or other svg compatible software). The command line used is as follows:

```bash
docker run -it --rm -v ~/docker:/host 2cd64516692f draw GM "id;SMILES"
```

where "id" and "SMILES" are the name and the SMILES code of the compound for which the representation of the graph machine is requested. The "id_5N".svg file produced is also written in the result folder. The computed representation of the graph machine can be simplified by adding the option "-H n" at the end of the above command line, where $n$ (<5) is the number of hidden neurons of the MLP. It is particularly useful to add the option -H 0 (or 1) to understand how a graph machine is built for a given compound. The file name produced is then "id_0N".svg (or "id_1N".svg).

The computation of the representation of the neural network with six hidden neurons is invoqued with the following command:

```bash
docker run -it --rm -v ~/docker:/host 2cd64516692f draw NN
```

It does not have any option. The name of the file written is model_6N.svg.
GRAPH MACHINE AND NEURAL NETWORK

RESULTS WITH DOCKER

All computations were made on a 3.2 GHz 8-Core Intel Xeon W iMac Pro with 64 GB of RAM running macOS Monterey 12.3 (Docker configuration: 16 CPUs, 40 GB allocated RAM). All graph machine tasks were launched with the maximum number of available CPUs minus 2, i.e. 14 CPUs on the iMac Pro. The shared folder /host is defined as /Users/home/docker.

Predictions of the EACN values for the 10-molecule test set

For the graph machine model with five hidden neurons, the predictions for the ten molecules are obtained with the following command (demo1 can be omitted):

docker run -it --rm -v ~/docker:/host 2cd64516692f demo1

The terminal output is then:

Run demo computing | computing time: 0.59 s
results :

| Compound                     | SMILES                        | EACN | Estimated_EACN | minmin_EACN | maxmax_EACN |
|------------------------------|-------------------------------|------|----------------|-------------|-------------|
| Diocylether                  | CCCCCCCCCCCCCCCCC           | 18.30| 9.76           | 8.12        | 12.84       |
| 2,2,4,6,6-pentamethylheptane | CC1(C)(C)(C)(C)(C)(C)(C)     | 11.70| 11.47          | 6.49        | 27.22       |
| Hemisqualane                 | CCCCCCCCCCCCCCCCCC(C)C       | 14.80| 14.07          | 11.21       | 17.05       |
| Limonene                     | CC=CC1(C)(C)(C)(C)(C)(C)(C)C | 1.80 | 1.10           | -1.22       | 5.47        |
| Caryophyllene                | C1=C(C)(C)(C)(C)(C)(C)(C)C   | 5.95 | 5.39           | 1.30        | 9.83        |
| Octyloctanoate               | CCCCCCCCCCCCCCCCCC(C)C       | 8.10 | 8.74           | 8.26        | 15.81       |
| Isopropyl myristate          | CCCCCCCCCCCCCC(C)C           | 7.25 | 6.94           | -3.00       | 29.67       |
| Rose oxide                   | CCCCCCCCOCCCCC             | 10.30| 9.76           | 8.12        | 12.04       |
| Linalyl acetate              | CCCCCCCOC(=O)CCC[=C:1]CCC   | 8.10 | 8.74           | 0.26        | 15.81       |
| beta-Ionone                  | CCCCCCCCOC(=O)CCC[=C:1]CCC   | 8.10 | 8.74           | 0.26        | 15.81       |

Writing results in file "~/Users/home/docker/result/GM_test_5N.xlsx".

For the neural network model with six hidden neurons, the predictions for the ten molecules are obtained with the following command:

docker run -it --rm -v ~/docker:/host 2cd64516692f demo2

The terminal output is then:

Run demo 2 | computing time: 0.18 s
results :

| Compound                     | M0  | M2  | M3  | n  | EACN | Estimated_EACN | minmin_EACN | maxmax_EACN |
|------------------------------|-----|-----|-----|----|------|----------------|-------------|-------------|
| Diocylether                  | 371.28 | 35.31 | 21.42 | 16  | 18.30 | 10.07          | 7.91        | 12.36       |
| 2,2,4,6,6-pentamethylheptane | 244.43 | 14.90 | -8.11 | 12  | 11.70 | 9.00           | 8.07        | 13.81       |
| Hemisqualane                 | 320.82 | 16.86 | -9.11 | 15  | 14.80 | 14.59          | 18.18       | 18.11       |
| Limonene                     | 197.00 | 24.05 | 5.78  | 10  | 1.00  | 1.12           | -1.14       | 5.76        |
| Caryophyllene                | 254.06 | 25.94 | 5.64  | 15  | 5.95  | 6.16           | -4.35       | 13.89       |
| Octyloctanoate               | 371.81 | 53.73 | 31.52 | 16  | 8.10  | 7.41           | 6.63        | 8.89        |
| Isopropyl myristate          | 392.03 | 55.25 | 31.87 | 17  | 7.25  | 8.86           | 7.31        | 9.57        |
| Rose oxide                   | 214.57 | 42.15 | 32.13 | 10  | -1.70 | -1.64          | -3.03       | 1.49        |
| Linalyl acetate              | 263.17 | 58.29 | 27.24 | 12  | -8.90 | -1.83          | -4.13       | 8.81        |
| beta-Ionone                  | 247.90 | 61.65 | 48.43 | 13  | -1.99 | -1.46          | -5.94       | 3.45        |

Writing results in file "~/Users/home/docker/result/NN_test_6N.xlsx"

In each case, the EACN predictions are the same as those reported as estimation errors in columns 5 and 6 of the Table 1 paper, or as estimations in columns 10 and 11 of Table S1 for the test molecules (entries 32, 60, 61, 88, 93, 102, 103, 109, 116 and 117). The predictions are also stored in the GM_test_5N.xlsx and NN_test_6N.xlsx files as explained in the 'loading and launching the Docker image' section.

Estimations of the EACN values for the 111-molecule training set

For the graph machine model with five hidden neurons, the estimations for the 111 molecules of the training set are obtained with the following command:

docker run -it --rm -v ~/docker:/host 2cd64516692f demo3
| Compound                  | M0     | M1     | EACN  | Estimated_EACN | minmin_EACN | maxmax_EACN |
|--------------------------|--------|--------|-------|----------------|-------------|-------------|
| hexane                   | 159.25 | 8.17   | -0.88 | 6              | 6.00        | 6.00        |
| heptane                  | 176.32 | 9.14   | -0.83 | 7              | 7.00        | 7.07        |
| octane                   | 198.88 | 10.04  | -0.80 | 8              | 8.00        | 8.08        |
| nonane                   | 218.16 | 11.07  | -0.71 | 9              | 9.00        | 9.07        |
| n-decane                 | 238.31 | 11.78  | -0.67 | 10             | 10.00       | 10.00       |
| undecane                 | 258.33 | 12.51  | -0.60 | 11             | 11.00       | 11.08       |
| dodecane                 | 278.75 | 13.46  | -0.53 | 12             | 12.00       | 12.09       |
| tridecane                | 298.90 | 14.22  | -0.47 | 13             | 13.00       | 13.09       |
| tetradecane              | 319.84 | 15.16  | -0.40 | 14             | 14.00       | 14.12       |
| pentadecane              | 339.96 | 16.11  | -0.33 | 15             | 15.00       | 15.15       |
| n-hexadecane             | 359.84 | 17.09  | -0.26 | 16             | 16.00       | 16.15       |
| heptadecane              | 379.87 | 18.07  | -0.19 | 17             | 17.00       | 17.15       |
| octadecane               | 399.94 | 19.05  | -0.12 | 18             | 18.00       | 18.23       |
| nonadecane               | 419.21 | 20.03  | -0.05 | 19             | 19.00       | 19.23       |
| eicosane                 | 448.35 | 21.03  | -0.00 | 20             | 20.00       | 20.23       |
| Cyclohexane              | 135.77 | 5.96   | -0.12 | 6              | 6.00        | 6.08        |
| Methylhexylcyclohexane   | 151.55 | 7.16   | -0.04 | 7              | 7.35        | 7.32        |
| Ethylhexylcyclohexane    | 171.51 | 8.19   | -0.00 | 8              | 8.00        | 8.08        |
| Propylhexylcyclohexane   | 191.55 | 9.18   | -0.00 | 9              | 9.00        | 9.07        |
| Butylhexylcyclohexane    | 210.95 | 10.17  | -0.03 | 10             | 10.00       | 10.07       |
| Decylhexylcyclohexane    | 332.16 | 14.95  | -0.24 | 16             | 16.00       | 16.15       |
| Dodecylicyclohexane      | 371.01 | 16.81  | 0.29  | 17              | 17.50       | 18.17       |
| 1-Decene                 | 192.38 | 18.77  | 3.59  | 8              | 8.00        | 8.08        |
| 1-Dodecene               | 234.26 | 21.31  | 3.17  | 10              | 10.00       | 10.15       |
| 1-Octadecene             | 394.66 | 27.89  | 3.60  | 18              | 18.00       | 18.23       |
| Dipropylether            | 171.36 | 26.39  | 20.59 | 6              | 6.00        | 6.08        |
| Dipentylether            | 213.38 | 28.47  | 21.23 | 8              | 8.00        | 8.08        |
| Dipentylether            | 225.44 | 29.29  | 21.23 | 8              | 8.00        | 8.08        |
| Cyclohexane              | 357.75 | 5.86   | -0.12 | 6              | 6.00        | 6.08        |
| Methylhexylcyclohexane   | 373.51 | 7.16   | -0.04 | 7              | 7.35        | 7.32        |
| Ethylhexylcyclohexane    | 393.47 | 8.19   | -0.00 | 8              | 8.00        | 8.08        |
| Propylhexylcyclohexane   | 413.51 | 9.18   | -0.00 | 9              | 9.00        | 9.07        |
| Butylhexylcyclohexane    | 433.91 | 10.17  | -0.03 | 10             | 10.00       | 10.07       |
| Decylhexylcyclohexane    | 453.67 | 14.95  | -0.24 | 16             | 16.00       | 16.15       |
| Dodecylicyclohexane      | 492.86 | 16.81  | 0.29  | 17              | 17.50       | 18.17       |
| 1-Decene                 | 320.64 | 21.31  | 3.17  | 8              | 8.00        | 8.08        |
| 1-Dodecene               | 365.87 | 21.31  | 3.17  | 10              | 10.00       | 10.15       |
| 1-Octadecene             | 434.66 | 21.31  | 3.17  | 18              | 18.00       | 18.23       |
| Dipropylether            | 171.36 | 26.39  | 20.59 | 6              | 6.00        | 6.08        |
| Dipentylether            | 213.38 | 28.47  | 21.23 | 8              | 8.00        | 8.08        |
| Dipentylether            | 225.44 | 29.29  | 21.23 | 8              | 8.00        | 8.08        |
| Cyclohexane              | 357.75 | 5.86   | -0.12 | 6              | 6.00        | 6.08        |
| Methylhexylcyclohexane   | 373.51 | 7.16   | -0.04 | 7              | 7.35        | 7.32        |
| Ethylhexylcyclohexane    | 393.47 | 8.19   | -0.00 | 8              | 8.00        | 8.08        |
| Propylhexylcyclohexane   | 413.51 | 9.18   | -0.00 | 9              | 9.00        | 9.07        |
| Butylhexylcyclohexane    | 433.91 | 10.17  | -0.03 | 10             | 10.00       | 10.07       |
| Decylhexylcyclohexane    | 453.67 | 14.95  | -0.24 | 16             | 16.00       | 16.15       |
| Dodecylicyclohexane      | 492.86 | 16.81  | 0.29  | 17              | 17.50       | 18.17       |
| 1-Decene                 | 320.64 | 21.31  | 3.17  | 8              | 8.00        | 8.08        |
| 1-Dodecene               | 365.87 | 21.31  | 3.17  | 10              | 10.00       | 10.15       |
| 1-Octadecene             | 434.66 | 21.31  | 3.17  | 18              | 18.00       | 18.23       |
The results are written in file "~/Users/home/docker/result/GM_limonene_5N.xlsx"

It is noteworthy to mention that the above estimations for the training set with both models are closer to the experimental values than the LOO predictions that are reported in Table S1. Indeed the previous values are the mean of the estimated values at the end of the training with the 25 models that have the best VLOO scores. For example, with the graph machine model the greater deviations listed in the first table are equal to 0.8, 0.5 and 0.6 for 1,2-dimethylcyclohexane, 3-Methyl-1-cyclohexene and α-Pinene respectively, whereas the larger deviations EACNexp reported in Table S1 are equal to ~2.7 for eucalyptol and ~2.1 for cedryl methyl ether.

### Prediction of EACN for a single molecule of the test set

EACN prediction of a given test molecule can be replicated for the default graph machine based model or for the default neural network based models. Limonene is given as an example. According to Table S1 in the SI_datasets.xlsx file, the following information can be extracted (entry 88) for this molecule:

| Compound | SMILES code | \( M^X_0 \) | \( M^X_2 \) | \( M^X_4 \) | \( N_C \) | EACNexp |
|----------|-------------|-------------|-------------|-------------|----------|---------|
| limonene | CC(=C)C1CCC(=CC1)C | 197.80 | 24.85 | 5.78 | 10 | 1.8 |

The command used to predict its EACN from SMILES is the following:

docker run -it --rm -v ~/docker:/host 2cd64516692f get GM 
"limonene;CC(=C)C1CCC(=CC1)C"

The output produced is then:

time: 0.31 s
results:

| ID | smiles | Estimated EACN | minmin EACN | maxmax EACN |
|----|--------|----------------|-------------|------------|
| limonene | CC(=C)C1CCC(=CC1)C | 1.10 | -1.22 | 5.47 |

The results are identical to those written for limonene in the file GM_test_5N.xlsx .

The commands for predicting its EACN from the descriptors used as inputs is the following:
The command below is then entered in the terminal window to predict the EACN value of ethyl hexanoate from the SMILES input:

docker run -it --rm -v ~/docker:/host 2cd64516692f get NN "ethyl hexanoate;CCCC[C:1]C(OCC)=O"

The messages returned are the following:

computing time: 0.32 s
results :

| ID | smiles   | Estimated_EACN | minmin_EACN | maxmax_EACN |
|----|----------|----------------|-------------|-------------|
| ethyl hexanoate | CCC[C:1]C(OCC)=O | -0.75 | -4.45 | 3.82 |

Writing results in file "~/docker/result/GM_ethyl hexanoate_5N.xlsx"

Similarly, this command line allows the EACN value prediction of the same molecule, but with the sigma-moment descriptors used as inputs of the NN estimator:

docker run -it --rm -v ~/docker:/host 2cd64516692f get NN "ethyl hexanoate;213.46,49.02,31.67,8"

The messages returned are the following:

computing time: 0.01 s
results :

| ID | M0 | M2 | M3 | n | Estimated_EACN | minmin_EACN | maxmax_EACN |
|----|----|----|----|---|----------------|-------------|-------------|
| ethyl hexanoate | 213.46 | 49.02 | 31.67 | 8 | -2.98 | -5.64 | -1.73 |

Writing results in file "~/docker/result/NN_ethyl hexanoate_6N.xlsx"

The EACN predictions are respectively equal to −0.7 and −3 for the GM and NN methods. In Figure 11d of the paper, the corresponding points (circle and diamond) are those of the ethyl alkanoate series with $N_c=8$. 

The values of the predicted EACN for limonene are equal to 1.1 for the GM method, and 1.1 for the NN method, as listed in Table 1 of the paper and in Table S1 of the SI (entry 88).

Prediction of EACN for other compounds

Prediction of the EACN of any liquid of molecule containing carbon, hydrogen, oxygen, nitrogen and chloride atoms can also be computed with similar command lines starting either from its SMILES code or its COSMO-RS σ-moments if available.

Ethyl hexanoate that belongs to the homologous set is used as an example. According to Table S2 in the SI_datasets.xlsx file, the following information can be extracted:

| Compound | SMILES code       | $M_0^{X[a]}$ | $M_2^{X[b]}$ | $M_3^{X[c]}$ | $N_c$ | EACN$_{exp}$ |
|----------|------------------|--------------|--------------|--------------|------|--------------|
| limonene | CCCCC[C:1]C(OCC)=O | 213.46       | 49.02        | 31.67        | 10   | 1.12         |

The command below is then entered in the terminal window to predict the EACN value of limonene:

```bash
docker run -it --rm -v ~/docker:/host 2cd64516692f get GM "limonene;213.46,49.02,31.67,10"
```

The messages returned are:

computing time: 0.0
results :

| ID | Estimated_EACN | minmin_EACN | maxmax_EACN |
|----|----------------|-------------|-------------|
| limonene | 1.12 | -1.14 | 5.76 |

Writing results in file "~/docker/result/GM_limonene_5N.xlsx"

The command below is then entered in the terminal window to predict the EACN value of limonene:

```bash
docker run -it --rm -v ~/docker:/host 2cd64516692f get NN "limonene;197.80,24.85,5.78,10"
```

The messages returned are:

computing time: 0.0
results :

| ID | Estimated_EACN | minmin_EACN | maxmax_EACN |
|----|----------------|-------------|-------------|
| limonene | 0.75 | -0.7 | 4.85 |

Writing results in file "~/docker/result/NN_limonene_6N.xlsx"
The prediction can also be obtained for a compound if only its SMILES code is known. Thus the predicted EACN for 1-(pentyloxy)heptane (not a member of Table S2) is computed with the SMILES in the following command:

```
docker run -it --rm -v ~/docker:/host 2cd64516692f get GM "1-(pentyloxy)heptane;CCCCCCCCCOCCCC"
```

The messages returned are the following:

```
computing time: 0.35 s
results :
| ID                        | smiles               |   Estimated_EACN |   minmin_EACN |   maxmax_EACN |
|---------------------------|----------------------|------------------|---------------|---------------|
| 1-(pentyloxy)heptane      | CCCCCCOCCCCC         | 6.20             | 5.63          | 6.76          |

Writing results in file "/Users/home/docker/result/GM_1-(pentyloxy)heptane_5N.xlsx"
```

The predicted EACN for 1-(pentyloxy)heptane is equal to 6.2; this value is quite good, since the same result (6.2) is computed using the appropriate Table 2 GM fit of the paper (0.89n–4.5 with n=12).

**Creating a graph machine image for a compound**

This command is useful to obtain the representation of a graph machine for a new compound, in particular when the predicted value has a large deviation or does not seem right, e.g. the hexan-2-one, a compound that belongs to the 2-ketone series discussed in the paper. In the following example, the command line to generate the graph machine image for hexan-2-one is:

```
docker run -it --rm -v ~/docker:/host 2cd64516692f draw GM "hexan-2-one;CCCCC(C)=O"
```

The messages returned are the following:

```
Writing image in file /Users/home/docker/result/hexan-2-one_5N.svg
```

The resulting svg file can be visualized with any browser. The svg output for the command inserted as a png image is displayed on next page.

```
hexan-2-one  CCCCC(C)=O
```

The svg file has been converted into a png file before insertion to reduce the document size. The written svg image is scalable and allows zooming on the connexion labels between the neurons easily.

A simpler representation of the graph machine built from the same compound but with zero hidden neuron can be obtained with the following command:
The EACN prediction for hexan-2-one is then computed by the following command:

docker run -it --rm -v ~/docker:/host 2cd64516692f get GM "hexan-2-one;CCCCC(C)=O"

The predicted EACN value returned is equal to −3.5, i.e. almost identical to the estimation value computed for octan-2-one (−3.4), so this predicted value is probably wrong. Indeed the data point with coordinates (N_C=6;EACN=−3.5) is well above the experimental dashed line of the ketone series in Figure 11a of the paper (or calculated with the experimental fit equation (−4.9) of the Table 2 paper. Seeking for an explanation of this deviation, we can compare the simplified GM representations of octan-2-one and decan-2-one with the hexan-2-one GM image. The representation for these two ketones are the following:

We can observe that for these two graph machine representations, the root node is not the node corresponding to the ketone carbon like in the previous hexan-2-one image. For consistency it would be advisable to build a hexan-2-one graph machine with an output node located in the middle of its directed graph. To illustrate this manual modification, we highlight it in the following SMILES code with the root node chosen by the algorithm colored in
green and the desired root node colored in red:

```
CCCC(C)=O
```

The command to compute a graph machine (with 0 neuron) with this particularized SMILES is then:

```
docker run -it --rm -v ~/docker:/host 2cd64516692f draw GM "hexan-2-one;CC[C:1]CC(C)=O" -H 0
```

The representation of the new graph machine for the hexan-2-one is displayed on next page.

With this new SMILES code input, the EACN prediction value for hexan-2-one computed by the following command:

```
docker run -it --rm -v ~/docker:/host 2cd64516692f get GM "hexan-2-one;CC[C:1]CC(C)=O"
```

is equal to –4.9, a value inline with the computed values for the six other ketones of the homologous series. Let us repeat that such a modification of the SMILES code would have not been necessary if a larger number of compounds with various positions of the ketone function were present in the training set.

References

(1) Bouton, F.; Durand, M.; Nardello-Rataj, V.; Borosy, A. P.; Quellet, C.; Aubry, J.-M. A QSPR Model for the Prediction of the “Fish-Tail” Temperature of C,E5/Water/Polar Hydrocarbon Oil Systems. *Langmuir* 2010, 26 (11), 7962–7970. https://doi.org/10.1021/la904836m.

(2) Kanei, N.; Tamura, Y.; Kunieda, H. Effect of Types of Perfume Compounds on the Hydrophilic–Lipophilic Balance Temperature. *J. Colloid Interface Sci.* 1999, 218 (1), 13–22. https://doi.org/10.1006/jcis.1999.6371.

(3) Kahlweit, M.; Strey, R.; Busse, G. Effect of Alcohols on the Phase Behavior of Microemulsions. *J. Phys. Chem.* 1991, 95 (13), 5344–5352. https://doi.org/10.1021/j100166a077.

(4) Kunieda, H.; Shinoda, K. Evaluation of the Hydrophilic–Lipophilic Balance (HLB) of Nonionic Surfactants. I. Multisurfactant Systems. *J. Colloid Interface Sci.* 1985, 107 (1), 107–121. https://doi.org/10.1016/0021-9797(85)90154-7.

(5) Queste, S.; Salager, J. L.; Strey, R.; Aubry, J. M. The EACN Scale for Oil Classification Revisited Thanks to Fish Diagrams. *J. Colloid Interface Sci.* 2007, 312 (1), 98–107. https://doi.org/10.1016/j.jcis.2006.07.004.

(6) Lukowicz, T.; Benazzouz, A.; Nardello-Rataj, V.; Aubry, J.-M. Rationalization and Prediction of the Equivalent Alkane Carbon Number (EACN) of Polar Hydrocarbon Oils with COSMO-RS ω-Moments. *Langmuir* 2015, 31 (41), 11220–11226. https://doi.org/10.1021/acs.langmuir.5b02545.

(7) Lukowicz, T.; Illous, E.; Nardello-Rataj, V.; Aubry, J.-M. Prediction of the equivalent alkane carbon number (EACN) of aprotic polar oils with COSMO-RS sigma-moments. *Colloids Surf. A Physicochem. Eng. Asp.* 2018, 536, 53–59. https://doi.org/10.1016/j.colsurfa.2017.07.068.

(8) Wormuth, K. R.; Kaler, E. W. Microemulsifying Polar Oils. *J. Phys. Chem.* 1989, 93 (12), 4855–4861. https://doi.org/10.1021/j100349a035.

(9) Deen, G. R.; Pedersen, J. S. Phase Behavior and Microstructure of C12E5 Nonionic Microemulsions with
Chlorinated Oils. *Langmuir* **2008**, *24* (7), 3111–3117. https://doi.org/10.1021/la703323n.

(10) Lukowicz, T. Synergistic Solubilisation of Fragrances in Binary Surfactant Systems and Prediction of Their EACN Value with COSMO-RS. PhD thesis, Lille 1, 2015.

(11) Ontiveros, J. F.; Pierlot, C.; Catté, M.; Molinier, V.; Pizzino, A.; Salager, J.-L.; Aubry, J.-M. Classification of Ester Oils According to Their Equivalent Alkane Carbon Number (EACN) and Asymmetry of Fish Diagrams of C10E4/Ester Oil/Water Systems. *J. Colloid Interface Sci.* **2013**, *403*, 67–76. https://doi.org/10.1016/j.jcis.2013.03.071.

(12) Bouton, F.; Durand, M.; Nardelllo-Rataj, V.; Serry, M.; Aubry, J.-M. Classification of Terpene Oils Using the Fish Diagrams and the Equivalent Alkane Carbon (EACN) Scale. *Colloids Surf. A Physicochem. Eng. Asp.* **2009**, *338* (1), 142–147. https://doi.org/10.1016/j.colsurfa.2008.05.027.

(13) Bouton, F. Influence of Terpenes and Terpenoids on the Phase Behavior of Micro- and Macro-Emulsions. PhD thesis, Lille 1, 2010.

(14) Lade, O.; Beizai, K.; Sottmann, T.; Strey, R. Polymerizable Nonionic Microemulsions: Phase Behavior of H2O–n-Alkyl Methacrylate–n-Alkyl Poly(Ethylene Glycol) Ether (CEj). *Langmuir* **2000**, *16* (9), 4122–4130. https://doi.org/10.1021/la991232i.