Electronic Supplementary Information for:

# Molecular Dynamics Based Descriptors for Predicting Supramolecular Gelation

Ruben Van Lommel,[1,2] Jianyu Zhao,[2] Wim M. De Borggraeve,[1] Frank De Proft[2] and Mercedes Alonso*[2]

1. Molecular Design and Synthesis, Department of Chemistry, KU Leuven, Celestijnenlaan 200F Leuven Chem&Tech, box 2404, 3001 Leuven, Belgium

2. Eenheid Algemene Chemie (ALGC), Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium

Correspondence: Mercedes.Alonso.Giner@vub.be

# Table of Contents

# Molecular Dynamics

All molecular dynamics simulations in this work were performed under an NPT ensemble with 5 gelator molecules in a periodic cubic simulation box together with the respective solvent to reach a concentration of 1.0 w/V% of the gelator molecules. The simulation box settings and average temperature, pressure and density during the 50 ns production run is provided in table below.

**Table S1:** Simulation box settings and average properties of the production run.

| Gelator | Solvent | Box edge (Å) | # solvent molecules | Temperature (K) | Pressure (bar) | Density (kg/m³) |
|---|---|---|---|---|---|---|
| 1 | Toluene | 76.15 | 2493 | 299.983 | 0.150 | 858.621 |
| 1 | Benzene | 76.15 | 2970 | 300.012 | 1.010 | 890.736 |
| 1 | Acetone | 76.15 | 3592 | 300.008 | 1.896 | 776.864 |
| 1 | Methanol | 76.15 | 6522 | 299.996 | 1.068 | 696.266 |
| 1 | Dimethylsulfoxide | 76.15 | 3718 | 299.998 | -0.103 | 1054.18 |
| 1 | Hexane | 76.15 | 2023 | 299.970 | 1.434 | 650.743 |
| 2 | Toluene | 70.65 | 1990 | 299.980 | 0.887 | 858.647 |
| 2 | Dibutylether | 70.65 | 1245 | 299.966 | 2.090 | 770.488 |
| 2 | Ethanol | 70.65 | 3611 | 299.964 | 0.450 | 739.477 |
| 2 | Dimethylsulfoxide | 70.65 | 2967 | 299.954 | 1.664 | 1054.57 |
| 2 | 1-propanol | 70.65 | 2818 | 299.966 | 1.394 | 766.899 |
| 3 | 1-propanol | 58.27 | 1584 | 299.980 | 1.026 | 768.161 |
| 3 | Dimethylsulfoxide | 58.27 | 1668 | 299.980 | 0.472 | 1056.46 |
| 3 | Nitromethane | 58.27 | 2212 | 299.992 | 0.932 | 1084.06 |
| 3 | Nitrobenzene | 58.27 | 1159 | 300.006 | 1.002 | 1157.82 |
| 3 | 1,2-dichlorobenzene | 58.27 | 1047 | 299.962 | 0.791 | 1219.64 |
| 3 | 1,3-dichlorobenzene | 58.27 | 1037 | 300.026 | 1.210 | 1208.59 |
| 4 | 1-propanol | 60.29 | 1677 | 299.997 | 0.819 | 768.426 |
| 4 | Dimethylsulfoxide | 60.29 | 1766 | 299.976 | 1.306 | 1056.01 |
| 4 | Dichloromethane | 60.29 | 1964 | 299.998 | 1.103 | 1134.11 |
| 4 | Hexane | 60.29 | 961 | 299.984 | -0.331 | 651.451 |
| 4 | Nitrobenzene | 60.29 | 1228 | 300.010 | 0.411 | 1157.29 |
| 4 | 1,2-dichlorobenzene | 60.29 | 1106 | 300.030 | 0.998 | 1218.56 |
| 5 | Ethanol | 52.30 | 1466 | 300.018 | 1.262 | 737.455 |
| 5 | 1-octanol | 52.30 | 545 | 299.986 | -0.276 | 825.939 |
| 5 | Dimethylsulfoxide | 52.30 | 1204 | 299.965 | 0.609 | 1053.31 |
| 5 | Water | 52.30 | 4739 | 300.024 | 1.181 | 1067.93 |
| 5 | Hexane | 52.30 | 655 | 300.003 | 1.110 | 649.221 |
| 5 | Toluene | 52.30 | 808 | 300.023 | 1.086 | 856.395 |
| 6 | Water | 67.91 | 10355 | 299.998 | 1.089 | 1071.11 |
| 6 | Dimethylsulfoxide | 67.91 | 2633 | 299.983 | 0.712 | 1056.63 |
| 6 | 1-octanol | 67.91 | 1191 | 299.996 | 0.995 | 828.892 |
| 6 | Acetonitrile | 67.91 | 3579 | 299.941 | 0.744 | 710.582 |
| 6 | Methyl *tert*-butyl ether | 67.91 | 1570 | 299.99 | 0.888 | 754.839 |
| 6 | Heptane | 67.91 | 1268 | 299.995 | 0.964 | 682.631 |
| 7 | Water | 69.56 | 11123 | 299.994 | 1.038 | 1070.78 |
| 7 | Dimethylsulfoxide | 69.56 | 2828 | 299.95 | 1.070 | 1056.68 |
| 7 | 1-octanol | 69.56 | 1280 | 300.001 | 1.196 | 828.504 |
| 7 | Acetonitrile | 69.56 | 3845 | 299.95 | 0.749 | 710.082 |
| 7 | Methyl *tert*-butyl ether | 69.56 | 1678 | 300.017 | 1.131 | 754.437 |
| 7 | Heptane | 69.56 | 1362 | 299.981 | 0.841 | 682.588 |
| 8 | Water | 69.55 | 11123 | 299.993 | 1.074 | 1070.83 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 8 | Dimethylsulfoxide | 69.55 | 2828 | 299.953 | 1.263 | 1056.84 |
| 8 | 1-octanol | 69.55 | 1280 | 300.014 | 1.126 | 828.533 |
| 8 | Acetonitrile | 69.55 | 3845 | 299.960 | 0.800 | 710.089 |
| 8 | Methyl *tert*-butyl ether | 69.55 | 1687 | 299.974 | 0.898 | 754.791 |
| 8 | Heptane | 69.55 | 1362 | 299.977 | 2.248 | 682.660 |
| 9 | Water | 69.76 | 11122 | 300.011 | 1.305 | 1070.81 |
| 9 | Dimethylsulfoxide | 69.76 | 2828 | 299.944 | 0.393 | 1056.83 |
| 9 | 1-octanol | 69.76 | 1280 | 300.008 | 1.141 | 828.56 |
| 9 | Acetonitrile | 69.76 | 3845 | 299.927 | 0.802 | 710.176 |
| 9 | Methyl *tert*-butyl ether | 69.76 | 1987 | 299.998 | 1.183 | 754.476 |
| 9 | Heptane | 69.76 | 1362 | 299.985 | 0.943 | 682.441 |
| 10 | Water | 69.64 | 11124 | 300.004 | 1.052 | 1070.95 |
| 10 | Dimethylsulfoxide | 69.64 | 2829 | 299.965 | 0.808 | 1056.89 |
| 10 | 1-octanol | 69.64 | 1280 | 299.959 | 0.983 | 829.164 |
| 10 | Acetonitrile | 69.64 | 3845 | 299.96 | 0.806 | 710.064 |
| 10 | Methyl *tert*-butyl ether | 69.64 | 1687 | 300.003 | 1.035 | 754.912 |
| 10 | Heptane | 69.64 | 1362 | 299.964 | 0.748 | 682.801 |
| 11 | Water | 71.12 | 11891 | 299.999 | 0.950 | 1070.55 |
| 11 | Dimethylsulfoxide | 71.12 | 3024 | 299.956 | 0.749 | 1056.36 |
| 11 | 1-octanol | 71.12 | 1368 | 299.993 | 0.977 | 828.390 |
| 11 | Acetonitrile | 71.12 | 4110 | 299.963 | 0.784 | 709.919 |
| 11 | Methyl *tert*-butyl ether | 71.12 | 1803 | 299.981 | 0.955 | 754.727 |
| 11 | Heptane | 71.12 | 1456 | 299.973 | 0.599 | 682.724 |

# How to calculate the MD based descriptors?

In the next section, the method to calculate the MD based descriptors derived in this work: rSASA, HB%, rH and F is explained in a detailed, tutorial like fashion.

### 1) Preparation

The calculation of the descriptors rSASA, rH and F requires the Solvent Accessible Surface Area (SASA), maximum end-to-end distance ($R_{max}$) and the volume (V) of a fully extended gelator molecule. To obtain a fully extended molecule, the molecule is built in an adequate building software (for example Gaussview[1]) and all dihedral angles of the backbone are set at 180°. No further optimization of the structure is required. Below a graphical representation together with the respective coordinates of the fully extended gelator molecules that are considered in this study can be found.

To obtain $R_{max}$, the distance is measured between the atoms that are furthest away from each other in the fully extended conformation.

The SASA of the extended conformation can be obtained through the gmx sasa implementation in the GROMACS software.[2] First, the extended conformer is saved as a pdb file (name.pdb). Then from this pdb file an index file (index.ndx) can be generated through the following command line in GROMACS:

gmx make_ndx -f name.pdb -o index.ndx

Following this, the SASA of the molecule can be computed by means of the command described below:

gmx sasa –f name.pdb –s name.pdb –n index.ndx –o sasa.xvg

Before the computation, the software will ask for which index group the SASA needs to be computed. Make sure to select the entire system in this case.

Similarly, the volume of the extended gelator molecule can be obtained through the next command:

gmx sasa –f name.pdb –s name.pdb –n index.ndx –tv volume.xvg

The SASA, $R_{max}$ and volume of the extended molecules considered in this work are provided below.

**2) Calculating rSASA**

rSASA is calculated through the following equation:

$$rSASA = \frac{S\bar{A}SA}{SASA_{max}}$$

$SASA_{max}$ is obtained by multiplying the SASA of a fully extended molecule (see above) with the total number of gelator molecules present in the simulation (in our case 5). To compute the $S\bar{A}SA$ a trajectory file of the simulation is necessary in .xtc (md.xtc) format, together with a gromacs file (md.gro) in which the solvent and gelator molecules are labeled distinctively as for example SOL and GEL. Next, an index file can be created through the following command in GROMACS:

gmx make_ndx –f md.gro –o indexmd.ndx

Then the evolution of the SASA during the simulation can be calculated through:

gmx sasa –f md.xtc –s md.gro –n indexmd.ndx –surface GEL –o sasamd.xvg

Following this, the average of the SASA can be taken from the sasamd.xvg file, straightforward through any mathematical operating software that is capable of calculating averages from a list of data points.

### 3) Calculating HB%

To obtain the HB%, first all hydrogen bond donor and hydrogen bond acceptor atoms need to be identified in the gelator molecule. This is done on the basis of prior chemical knowledge. Next, the indices of these atoms, which can be retrieved from the md.gro file used to run the simulation, are collected in an index file (indexHB.ndx) as follows:

[HBdonor1]
"index number"

[HBdonor2]
"index number"

[HBacceptor1]
"index number"

[HBacceptor2]
"index number"

…

Through the following command line in GROMACS, a distance histogram is provided for the distances between the hydrogen bond acceptor and donor atoms.

gmx distance –f md.xtc –s md.gro –n indexHB.ndx –oh HB.xvg –select 'com of group "HBdonor1" plus com of group "HBacceptor1"' 'com of group "HBdonor1" plus com of group "HBacceptor2"' 'com of group "HBdonor2" plus com of group "HBacceptor1"' 'com of group "HBdonor2" plus com of group "HBacceptor2"' –len 0.15

At the end of the output file generated by this command (HB.xvg), the probability of finding the corresponding atoms further or equal from each other than 3.0 Å is given. From this value, the probability of finding the atoms closer than 3.0 Å from each other can be calculated. Summation of these values for all hydrogen bond donor/hydrogen bond acceptor combinations and multiplying it with 100% will render the HB%.

### 4) Calculating rH

rH is calculated by the following equation:

$$rH = \frac{\bar{R}}{R_{max}}$$

The computation of $R_{max}$ from the extended conformation is explained above. To obtain $\bar{R}$, the indices of the atoms that are furthest away from each other in the gelator molecule need to be gathered in an index file (indexrH.ndx) as follows:

[gelator1]
"index1" "index2"

[gelator2]
"index1" "index2"

…

Following this, the evolution of the distance between these atoms during the simulation can be calculated via the following command in GROMACS:

gmx distance –f md.xtc –s md.gro –n indexrH.ndx –oall rH.xvg

From this command, the software will ask to specify from which groups the distance needs to be calculated. Sequentially provide all groups. From this, the average over time and all gelator molecules $\bar{R}$ can be calculated with any software that is capable of averaging over a number of datapoints using the rH.xvg output file.

**5) Calculating F**

The descriptor F is defined by the following equation:

$$F = \frac{\bar{R}_g}{R_h^{'}}$$

The computation of $R_h^{'}$ is obtained from the total number of gelator molecules present in the simulation (N, in our case 5) and the volume of an extended gelator molecule (V). The technical aspects of obtaining the latter property is provided above.

$$R_h^{'} = \sqrt[3]{\frac{3 \times V}{4 \times \pi}}$$

$\bar{R}_g$ is calculated from the molecular dynamics simulations through the next command line in GROMACS:

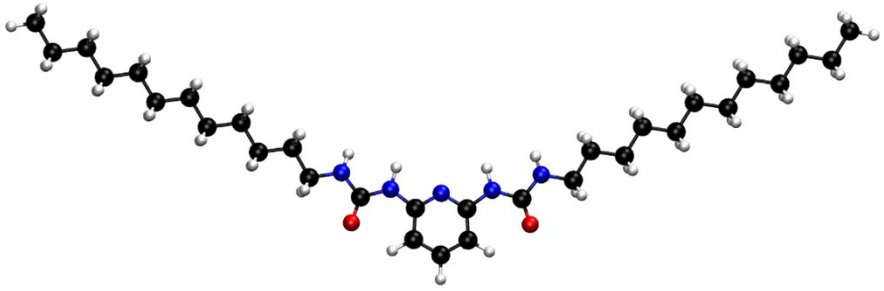
gmx gyrate –f md.xtc –s md.gro –n indexmd.ndx –o Rg.xvg

After this command make sure you select the group that corresponds to the gelator molecules from the index file. Subsequently the average over time value of the radius of gyration $\bar{R}_g$ can be obtained from the Rg.xvg file.

All descriptors that are calculated in this work can be retrieved from Table S2.

**Molecule 1**

SASA = 11.26 nm² $R_{max}$ = 38.46 Å V = 1.786 nm³

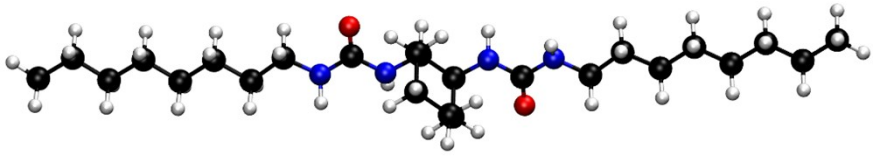


| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 5.888 | -1.7 | 0.247 | C | -15.014 | 2.534 | -1.541 |
| N | 4.45 | -1.475 | 0.218 | H | -15.365 | 1.972 | -0.667 |
| C | 3.608 | -2.561 | 0.153 | H | -15.216 | 1.902 | -2.415 |
| O | 4.01 | -3.702 | 0.096 | C | -15.812 | 3.83 | -1.661 |
| N | 2.27 | -2.203 | 0.132 | H | -15.462 | 4.393 | -2.535 |
| C | 1.229 | -4.441 | 0.119 | H | -15.611 | 4.463 | -0.788 |
| C | 0.023 | -5.124 | 0.054 | C | -17.317 | 3.601 | -1.781 |
| C | 1.158 | -3.048 | 0.07 | H | -17.665 | 3.039 | -0.907 |
| C | -1.188 | -4.457 | -0.064 | H | -17.516 | 2.969 | -2.653 |
| C | -1.126 | -3.063 | -0.126 | C | -18.102 | 4.905 | -1.9 |
| N | -2.243 | -2.234 | -0.255 | H | -17.787 | 5.469 | -2.782 |
| C | -3.573 | -2.61 | -0.354 | H | -19.175 | 4.723 | -1.984 |
| O | -3.962 | -3.756 | -0.32 | H | -17.937 | 5.539 | -1.025 |
| N | -4.424 | -1.535 | -0.47 | C | 6.638 | -0.378 | 0.318 |
| C | -5.855 | -1.78 | -0.584 | H | 6.329 | 0.174 | 1.215 |
| H | 6.148 | -2.241 | -0.664 | H | 6.359 | 0.236 | -0.544 |
| H | 6.169 | -2.344 | 1.09 | C | 8.151 | -0.578 | 0.35 |
| H | 4.118 | -0.656 | 0.706 | H | 8.419 | -1.203 | 1.21 |
| H | 2.024 | -1.23 | 0.024 | H | 8.46 | -1.133 | -0.543 |
| H | 0.027 | -6.207 | 0.097 | C | 8.926 | 0.736 | 0.422 |
| H | -2.135 | -4.967 | -0.108 | H | 8.615 | 1.292 | 1.314 |
| H | -2.018 | -1.257 | -0.134 | H | 8.661 | 1.361 | -0.439 |
| H | -4.075 | -0.712 | -0.939 | C | 10.439 | 0.534 | 0.453 |
| H | -6.161 | -2.323 | 0.311 | H | 10.702 | -0.093 | 1.314 |
| H | -6.077 | -2.427 | -1.441 | H | 10.748 | -0.024 | -0.439 |
| H | 2.18 | -4.941 | 0.203 | C | 11.219 | 1.844 | 0.525 |
| N | 0.014 | -2.377 | -0.055 | H | 10.911 | 2.401 | 1.418 |
| C | -6.617 | -0.468 | -0.701 | H | 10.957 | 2.471 | -0.336 |
| H | -6.263 | 0.088 | -1.579 | C | 12.732 | 1.638 | 0.557 |
| H | -6.397 | 0.15 | 0.176 | H | 12.994 | 1.011 | 1.417 |
| C | -8.123 | -0.688 | -0.822 | H | 13.04 | 1.08 | -0.336 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| H | -8.476 | -1.247 | 0.052 | C | 13.516 | 2.946 | 0.628 |
| H | -8.33 | -1.317 | -1.695 | H | 13.255 | 3.574 | -0.232 |
| C | -8.909 | 0.615 | -0.941 | H | 13.209 | 3.504 | 1.521 |
| H | -8.554 | 1.175 | -1.815 | C | 15.028 | 2.738 | 0.66 |
| H | -8.704 | 1.244 | -0.067 | H | 15.289 | 2.11 | 1.521 |
| C | -10.415 | 0.392 | -1.062 | H | 15.335 | 2.179 | -0.232 |
| H | -10.769 | -0.169 | -0.189 | C | 15.814 | 4.044 | 0.732 |
| H | -10.619 | -0.238 | -1.936 | H | 15.554 | 4.673 | -0.129 |
| C | -11.207 | 1.692 | -1.181 | H | 15.509 | 4.604 | 1.624 |
| H | -10.854 | 2.253 | -2.055 | C | 17.327 | 3.836 | 0.763 |
| H | -11.005 | 2.323 | -0.307 | H | 17.585 | 3.208 | 1.623 |
| C | -12.713 | 1.466 | -1.302 | H | 17.631 | 3.277 | -0.129 |
| H | -13.065 | 0.904 | -0.429 | C | 18.1 | 5.15 | 0.835 |
| H | -12.915 | 0.835 | -2.176 | H | 19.178 | 4.983 | 0.856 |
| C | -13.509 | 2.763 | -1.422 | H | 17.875 | 5.781 | -0.03 |
| H | -13.157 | 3.325 | -2.296 | H | 17.829 | 5.711 | 1.733 |
| H | -13.306 | 3.395 | -0.548 | | | | |

**Molecule 2**

SASA = 8.99 nm² $R_{max}$ = 30.37 Å V = 1.494 nm³

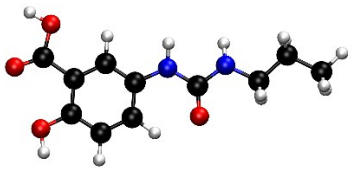


| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 7.345 | 0.184 | -1.148 | H | 10.378 | 9.724 | 4.89 |
| C | 7.989 | -0.735 | -0.13 | C | 12.162 | 10.615 | 4.065 |
| C | 6.074 | 0.803 | -0.603 | H | 11.63 | 11.237 | 3.336 |
| H | 7.107 | -0.398 | -2.078 | H | 13.079 | 10.282 | 3.569 |
| C | 7.019 | -1.798 | 0.345 | C | 12.51 | 11.447 | 5.296 |
| H | 8.339 | -0.131 | 0.748 | H | 13.056 | 10.848 | 6.03 |
| H | 8.89 | -1.225 | -0.583 | H | 13.129 | 12.308 | 5.038 |
| C | 5.104 | -0.259 | -0.128 | H | 11.603 | 11.819 | 5.782 |
| H | 5.569 | 1.466 | -1.353 | N | 4.708 | -1.101 | -1.266 |
| C | 5.747 | -1.18 | 0.889 | H | 4.563 | -0.592 | -2.129 |
| H | 7.504 | -2.42 | 1.141 | C | 3.821 | -2.134 | -1.029 |
| H | 6.766 | -2.48 | -0.509 | N | 3.488 | -2.891 | -2.148 |
| H | 4.176 | 0.229 | 0.268 | O | 3.332 | -2.35 | 0.058 |
| H | 5.026 | -1.99 | 1.171 | H | 4.229 | -3.054 | -2.816 |
| H | 5.983 | -0.6 | 1.82 | C | 2.553 | -3.996 | -1.956 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| N | 6.471 | 1.646 | 0.535 | H | 1.631 | -3.553 | -1.576 |
| H | 5.718 | 1.858 | 1.177 | H | 2.91 | -4.69 | -1.185 |
| C | 7.358 | 2.678 | 0.297 | C | 2.27 | -4.746 | -3.251 |
| N | 7.692 | 3.436 | 1.415 | H | 1.96 | -4.029 | -4.019 |
| O | 7.791 | 2.938 | -0.805 | H | 3.18 | -5.23 | -3.624 |
| H | 7.781 | 2.933 | 2.287 | C | 1.176 | -5.791 | -3.038 |
| C | 8.626 | 4.541 | 1.223 | H | 0.267 | -5.289 | -2.682 |
| H | 8.159 | 5.213 | 0.5 | H | 1.478 | -6.477 | -2.238 |
| H | 9.564 | 4.195 | 0.769 | C | 0.831 | -6.589 | -4.291 |
| C | 8.91 | 5.291 | 2.516 | H | 1.726 | -7.094 | -4.672 |
| H | 7.96 | 5.587 | 2.975 | H | 0.507 | -5.899 | -5.08 |
| H | 9.416 | 4.639 | 3.238 | C | -0.266 | -7.621 | -4.033 |
| C | 9.767 | 6.526 | 2.244 | H | -1.165 | -7.111 | -3.663 |
| H | 9.239 | 7.177 | 1.535 | H | 0.055 | -8.295 | -3.229 |
| H | 10.696 | 6.224 | 1.747 | C | -0.636 | -8.439 | -5.267 |
| C | 10.098 | 7.337 | 3.493 | H | 0.252 | -8.952 | -5.653 |
| H | 10.619 | 6.705 | 4.223 | H | -0.965 | -7.758 | -6.063 |
| H | 9.164 | 7.656 | 3.973 | C | -1.734 | -9.464 | -4.989 |
| C | 10.955 | 8.561 | 3.176 | H | -2.622 | -8.946 | -4.609 |
| H | 10.425 | 9.197 | 2.454 | H | -1.405 | -10.133 | -4.186 |
| H | 11.876 | 8.237 | 2.677 | C | -2.102 | -10.28 | -6.225 |
| C | 11.305 | 9.396 | 4.404 | H | -1.231 | -10.818 | -6.61 |
| H | 11.831 | 8.774 | 5.137 | H | -2.88 | -11.014 | -6.007 |
| H | -2.468 | -9.628 | -7.024 | H | 8.033 | 0.956 | -1.42 |

**Molecule 3**

SASA = 4.47 nm² $R_{max}$ = 12.13 Å V = 0.683 nm³

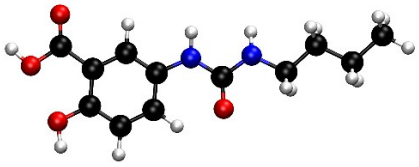


| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | -1.928 | -4.36 | 0.276 | H | -3.868 | -5.297 | -0.785 |
| C | -0.785 | -3.736 | 0.742 | N | 1.046 | -2.193 | 0.484 |
| C | -0.093 | -2.822 | -0.055 | H | 1.315 | -1.333 | 0.031 |
| C | -0.566 | -2.568 | -1.338 | C | 1.562 | -2.405 | 1.754 |
| C | -1.701 | -3.219 | -1.829 | O | 1.081 | -3.167 | 2.564 |
| C | -2.406 | -4.115 | -1.01 | N | 2.685 | -1.653 | 2.013 |
| H | -2.457 | -5.059 | 0.916 | H | 3.269 | -1.401 | 1.232 |
| H | -0.43 | -3.957 | 1.744 | C | 3.324 | -1.786 | 3.318 |
| H | -0.043 | -1.875 | -1.978 | H | 3.608 | -2.834 | 3.44 |
| C | -2.173 | -2.917 | -3.209 | H | 2.603 | -1.537 | 4.103 |
| O | -3.122 | -3.4 | -3.759 | C | 4.552 | -0.891 | 3.454 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| O | -1.388 | -2.001 | -3.82 | H | 5.26 | -1.142 | 2.659 |
| H | -1.767 | -1.872 | -4.698 | H | 4.237 | 0.143 | 3.298 |
| O | -3.521 | -4.723 | -1.472 | C | 5.205 | -1.046 | 4.824 |
| H | 5.524 | -2.078 | 4.997 | H | 6.083 | -0.406 | 4.918 |
| H | 4.507 | -0.773 | 5.62 | | | | |

**Molecule 4**

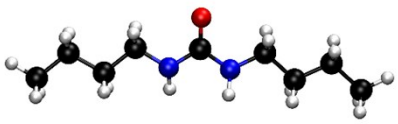
SASA = 4.85 nm² $R_{max}$ = 15.03 Å V = 0.762 nm³



| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | -3.116 | -1.94 | -0.627 | H | 0.82 | -0.862 | 1.134 |
| C | -1.796 | -2.198 | -0.301 | C | 0.942 | -2.761 | 0.386 |
| C | -0.969 | -1.185 | 0.183 | O | 0.352 | -3.744 | -0.005 |
| C | -1.494 | 0.096 | 0.32 | N | 2.264 | -2.786 | 0.768 |
| C | -2.825 | 0.366 | -0.01 | H | 2.797 | -1.939 | 0.661 |
| C | -3.651 | -0.66 | -0.495 | C | 2.986 | -4.053 | 0.693 |
| H | -3.745 | -2.742 | -0.999 | H | 2.949 | -4.39 | -0.346 |
| H | -1.404 | -3.201 | -0.429 | H | 2.475 | -4.798 | 1.31 |
| H | -0.873 | 0.897 | 0.685 | C | 4.434 | -3.917 | 1.149 |
| C | -3.35 | 1.75 | 0.16 | H | 4.934 | -3.163 | 0.532 |
| O | -2.74 | 2.698 | 0.567 | H | 4.436 | -3.552 | 2.179 |
| O | -4.649 | 1.843 | -0.201 | C | 5.187 | -5.243 | 1.069 |
| H | -4.904 | 2.76 | -0.037 | H | 5.158 | -5.619 | 0.039 |
| O | -4.938 | -0.4 | -0.814 | H | 4.672 | -5.987 | 1.686 |
| H | -5.349 | -1.211 | -1.126 | C | 6.638 | -5.112 | 1.525 |
| N | 0.366 | -1.504 | 0.502 | H | 6.687 | -4.76 | 2.559 |
| H | 7.164 | -6.067 | 1.469 | H | 7.178 | -4.394 | 0.903 |

**Molecule 5**

SASA = 4.67 nm² $R_{max}$ = 13.23 Å V = 0.699 nm³

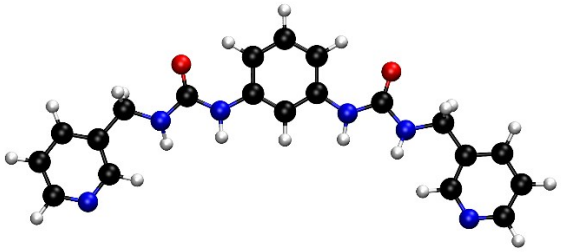


| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 0.019 | 0.445 | 0.003 | C | 6.079 | -1.065 | -0.432 |
| O | 0.049 | 1.654 | 0.092 | H | 7.058 | -0.584 | -0.451 |
| N | 1.156 | -0.333 | -0.116 | H | 6.067 | -1.77 | 0.404 |
| H | 1.127 | -1.225 | 0.358 | H | 5.969 | -1.641 | -1.354 |
| N | -1.154 | -0.285 | 0.011 | C | -2.42 | 0.427 | -0.09 |
| H | -1.118 | -1.177 | -0.462 | H | -2.458 | 1.044 | -0.998 |
| C | 2.454 | 0.326 | -0.135 | H | -2.47 | 1.111 | 0.758 |
| H | 2.603 | 0.937 | 0.766 | C | -3.589 | -0.547 | -0.065 |
| H | 2.452 | 1.012 | -0.983 | H | -3.543 | -1.131 | 0.86 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| C | 3.574 | -0.695 | -0.271 | H | -3.494 | -1.26 | -0.894 |
| H | 3.528 | -1.407 | 0.563 | C | -4.937 | 0.162 | -0.168 |
| H | 3.417 | -1.273 | -1.188 | H | -5.031 | 0.875 | 0.657 |
| C | 4.954 | -0.042 | -0.295 | H | -4.966 | 0.752 | -1.091 |
| H | 5.095 | 0.541 | 0.621 | C | -6.111 | -0.813 | -0.143 |
| H | 4.999 | 0.67 | -1.125 | H | -6.113 | -1.394 | 0.783 |
| H | -7.067 | -0.292 | -0.215 | H | -6.049 | -1.517 | -0.977 |

**Molecule 6**

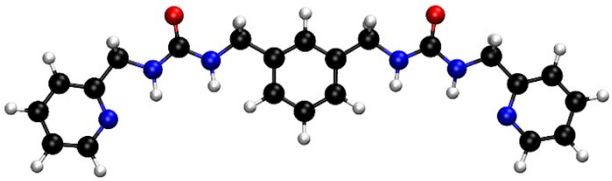
SASA = 7.10 nm² $R_{max}$ = 18.38 Å V = 1.133 nm³



| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 4.892 | -1.768 | -3.335 | H | 6.82 | -3.773 | -5.014 |
| C | 5.787 | -1.054 | -2.537 | C | 5.298 | -0.01 | -1.555 |
| C | 7.147 | -1.357 | -2.661 | H | 5.572 | -0.29 | -0.537 |
| C | 7.533 | -2.344 | -3.559 | H | 5.761 | 0.955 | -1.773 |
| C | 6.549 | -2.996 | -4.304 | N | 3.846 | 0.172 | -1.565 |
| N | 5.247 | -2.721 | -4.203 | H | 3.309 | -0.655 | -1.781 |
| C | 0.972 | 1.826 | 0.013 | C | 3.269 | 1.099 | -0.712 |
| C | 1.355 | 2.78 | 0.963 | O | 3.927 | 1.875 | -0.037 |
| C | -0.385 | 1.575 | -0.235 | N | 1.886 | 1.063 | -0.739 |
| C | 0.36 | 3.468 | 1.651 | H | 1.462 | 0.481 | -1.446 |
| C | -1.367 | 2.275 | 0.465 | H | 2.4 | 2.978 | 1.139 |
| C | -0.988 | 3.23 | 1.419 | H | -0.686 | 0.845 | -0.972 |
| C | -7.529 | 2.358 | 0.292 | H | 0.648 | 4.21 | 2.386 |
| C | -7.481 | 1.317 | -0.637 | H | -1.743 | 3.777 | 1.974 |
| C | -8.788 | 2.868 | 0.626 | N | -2.748 | 2.069 | 0.27 |
| N | -8.554 | 0.78 | -1.227 | H | -2.974 | 1.311 | -0.355 |
| C | -9.913 | 2.322 | 0.023 | C | -3.811 | 2.694 | 0.897 |
| C | -9.745 | 1.283 | -0.895 | O | -3.696 | 3.516 | 1.794 |
| H | 7.871 | -0.825 | -2.054 | N | -5.041 | 2.268 | 0.425 |
| H | 3.826 | -1.556 | -3.274 | H | -5.078 | 1.914 | -0.52 |
| H | 8.577 | -2.607 | -3.685 | C | -6.262 | 2.929 | 0.892 |
| H | -6.292 | 2.85 | 1.981 | H | -8.864 | 3.679 | 1.342 |
| H | -6.163 | 3.984 | 0.637 | H | -10.906 | 2.687 | 0.257 |
| H | -6.521 | 0.887 | -0.918 | H | -10.606 | 0.836 | -1.383 |

**Molecule 7**

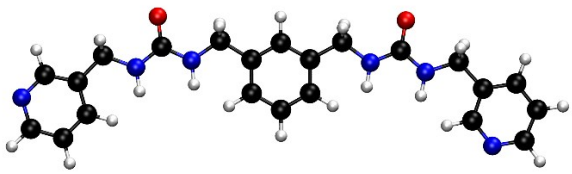
SASA = 7.64 nm² $R_{max}$ = 21.57 Å V = 1.205 nm³



| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 5.518 | -5.315 | -0.442 | C | -1.642 | 3.121 | 0.481 |
| C | 5.954 | -6.646 | -0.447 | H | -2.033 | 2.854 | 1.469 |
| C | 7.306 | -6.916 | -0.647 | H | -2.384 | 2.803 | -0.258 |
| H | 5.243 | -7.45 | -0.285 | N | -1.536 | 4.567 | 0.406 |
| H | 7.667 | -7.939 | -0.649 | H | -0.623 | 4.978 | 0.278 |
| C | 4.048 | -4.995 | -0.238 | C | -2.628 | 5.371 | 0.596 |
| H | 3.455 | -5.462 | -1.028 | O | -3.781 | 4.886 | 0.802 |
| H | 3.723 | -5.422 | 0.715 | N | -2.413 | 6.71 | 0.555 |
| N | 3.763 | -3.572 | -0.261 | H | -1.503 | 7.122 | 0.373 |
| H | 4.529 | -2.936 | -0.092 | C | -3.497 | 7.652 | 0.753 |
| C | 2.478 | -3.139 | -0.089 | H | -3.959 | 7.504 | 1.737 |
| O | 1.527 | -3.955 | 0.083 | H | -4.287 | 7.49 | 0.007 |
| N | 2.279 | -1.789 | -0.122 | C | -2.981 | 9.067 | 0.643 |
| H | 3.07 | -1.164 | -0.199 | C | -3.852 | 10.153 | 0.796 |
| C | 0.972 | -1.206 | 0.134 | C | -1.179 | 10.492 | 0.286 |
| H | 0.662 | -1.524 | 1.134 | H | -4.903 | 9.982 | 0.999 |
| H | 0.259 | -1.625 | -0.581 | H | -0.119 | 10.574 | 0.084 |
| C | 2.098 | 1.054 | -0.221 | C | 7.678 | -4.555 | -0.816 |
| C | 2.033 | 2.452 | -0.26 | H | 8.324 | -3.695 | -0.953 |
| C | 0.946 | 0.312 | 0.053 | C | -3.343 | 11.445 | 0.687 |
| C | 0.824 | 3.11 | -0.029 | H | -3.997 | 12.302 | 0.802 |
| H | 2.928 | 3.027 | -0.474 | C | -1.98 | 11.622 | 0.425 |
| C | -0.265 | 0.981 | 0.283 | H | -1.549 | 12.611 | 0.331 |
| C | -0.338 | 2.375 | 0.246 | C | 8.192 | -5.85 | -0.834 |
| H | 0.779 | 4.195 | -0.066 | H | 9.25 | -6.017 | -0.987 |
| H | -1.164 | 0.405 | 0.488 | N | -1.669 | 9.237 | 0.394 |
| N | 6.367 | -4.288 | -0.625 | H | 3.042 | 0.549 | -0.411 |

**Molecule 8**

SASA = 7.45 nm² $R_{max}$ = 21.94 Å V = 1.189 nm³

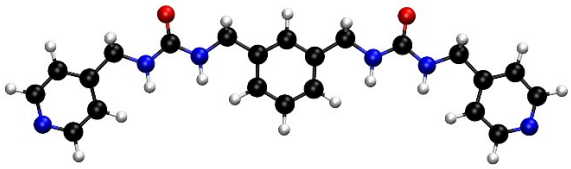


| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | -5.754 | -5.247 | 2.811 | N | 1.916 | 3.704 | -0.262 |
| C | -6.409 | -4.115 | 3.314 | H | 1.174 | 4.213 | 0.2 |
| C | -7.642 | -4.261 | 3.951 | C | 2.967 | 4.376 | -0.818 |
| H | -5.96 | -3.132 | 3.202 | O | 3.876 | 3.777 | -1.461 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| H | -8.17 | -3.401 | 4.346 | N | 2.987 | 5.732 | -0.647 |
| C | -4.404 | -5.108 | 2.134 | H | 2.263 | 6.191 | -0.114 |
| H | -3.665 | -5.71 | 2.669 | C | 4.034 | 6.543 | -1.244 |
| H | -4.471 | -5.499 | 1.114 | H | 4.003 | 6.372 | -2.323 |
| N | -3.913 | -3.741 | 2.059 | H | 5.007 | 6.191 | -0.886 |
| H | -4.504 | -3 | 2.409 | C | 3.909 | 8.026 | -0.959 |
| C | -2.71 | -3.451 | 1.477 | C | 4.816 | 8.97 | -1.443 |
| O | -1.982 | -4.352 | 0.971 | H | 5.642 | 8.655 | -2.073 |
| N | -2.341 | -2.137 | 1.463 | C | 4.643 | 10.317 | -1.116 |
| H | -2.93 | -1.436 | 1.887 | H | 5.331 | 11.071 | -1.478 |
| C | -1.136 | -1.712 | 0.768 | C | 3.564 | 10.683 | -0.311 |
| H | -1.204 | -2.006 | -0.283 | H | 3.393 | 11.716 | -0.036 |
| H | -0.299 | -2.262 | 1.208 | C | -8.19 | -5.538 | 4.068 |
| C | -1.734 | 0.644 | 1.535 | H | -9.145 | -5.697 | 4.554 |
| C | -1.461 | 2.016 | 1.579 | H | -2.616 | 0.247 | 2.026 |
| C | -0.872 | -0.218 | 0.852 | C | -6.38 | -6.485 | 2.974 |
| C | -0.326 | 2.529 | 0.951 | H | -5.919 | -7.391 | 2.598 |
| H | -2.136 | 2.684 | 2.105 | N | -7.573 | -6.638 | 3.587 |
| C | 0.263 | 0.307 | 0.218 | N | 2.675 | 9.781 | 0.161 |
| C | 0.549 | 1.673 | 0.266 | C | 2.851 | 8.485 | -0.164 |
| H | -0.117 | 3.595 | 0.987 | H | 2.114 | 7.79 | 0.225 |
| H | 0.926 | -0.361 | -0.324 | H | 1.732 | 2.038 | -1.484 |
| C | 1.773 | 2.266 | -0.413 | H | 2.677 | 1.798 | -0.015 |

**Molecule 9**

SASA = 8.17 nm² $R_{max}$ = 20.84 Å V = 1.272 nm³

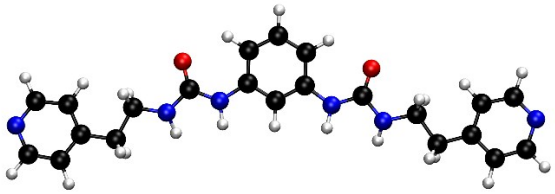


| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 6.294 | -4.82 | -0.905 | H | 0.544 | -2.094 | -0.2 |
| C | 6.964 | -6.049 | -0.874 | H | 2.149 | 2.924 | -1.201 |
| C | 8.242 | -6.143 | -1.421 | H | -0.086 | 3.724 | -0.486 |
| C | 1.817 | 0.897 | -0.547 | H | -1.04 | -0.225 | 0.912 |
| C | 1.451 | 2.235 | -0.738 | C | -2.075 | 2.321 | 0.699 |
| C | 0.923 | 0.006 | 0.049 | H | -2.203 | 2.149 | 1.773 |
| C | 0.195 | 2.686 | -0.334 | H | -2.86 | 1.761 | 0.184 |
| C | -0.34 | 0.468 | 0.451 | N | -2.292 | 3.733 | 0.431 |
| C | -0.713 | 1.8 | 0.268 | H | -1.526 | 4.274 | 0.058 |
| C | -4.695 | 7.902 | 0.473 | C | -3.467 | 4.344 | 0.762 |
| C | -3.525 | 8.516 | 0.021 | O | -4.429 | 3.708 | 1.283 |
| C | -3.537 | 9.883 | -0.26 | N | -3.553 | 5.684 | 0.505 |
| C | 6.951 | -3.732 | -1.487 | H | -2.785 | 6.171 | 0.065 |
| C | -5.839 | 8.698 | 0.617 | C | -4.758 | 6.427 | 0.819 |
| C | -5.769 | 10.056 | 0.314 | H | -4.953 | 6.317 | 1.89 |
| C | 8.232 | -3.909 | -2.01 | H | -5.601 | 5.976 | 0.286 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| N | -4.636 | 10.654 | -0.12 | H | -2.614 | 7.941 | -0.108 |
| N | 8.88 | -5.094 | -1.985 | H | -2.643 | 10.384 | -0.61 |
| H | 6.5 | -6.921 | -0.424 | H | 2.794 | 0.55 | -0.866 |
| H | 8.783 | -7.081 | -1.409 | H | 6.473 | -2.76 | -1.531 |
| C | 4.892 | -4.705 | -0.337 | H | -6.773 | 8.267 | 0.961 |
| H | 4.231 | -5.402 | -0.859 | H | -6.638 | 10.694 | 0.416 |
| H | 4.91 | -4.997 | 0.718 | H | 8.762 | -3.083 | -2.466 |
| N | 4.318 | -3.377 | -0.458 | N | 2.6 | -1.842 | -0.148 |
| H | 4.869 | -2.646 | -0.884 | H | 3.201 | -1.131 | -0.543 |
| C | 3.059 | -3.118 | 0.009 | C | 1.274 | -1.451 | 0.3 |
| O | 2.362 | -4.018 | 0.558 | H | 1.21 | -1.657 | 1.373 |

**Molecule 10**

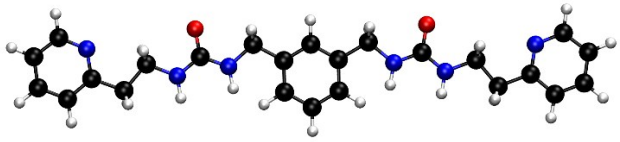
SASA = 7.54 nm² $R_{max}$ = 22.15 Å V = 1.223 nm³



| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 8.945 | -4.23 | 0.255 | H | 2.309 | 2.686 | 0.352 |
| C | 7.559 | -4.188 | 0.182 | C | -1.495 | 2.025 | -0.075 |
| C | 6.914 | -2.954 | 0.242 | H | -0.948 | -0.05 | -0.178 |
| C | 7.714 | -1.822 | 0.369 | C | -1.033 | 3.335 | 0.067 |
| C | 9.095 | -1.972 | 0.436 | H | 0.696 | 4.553 | 0.326 |
| N | 9.715 | -3.148 | 0.38 | H | -1.727 | 4.159 | 0.059 |
| H | 9.465 | -5.182 | 0.214 | N | -2.853 | 1.685 | -0.219 |
| H | 6.984 | -5.103 | 0.084 | C | -3.934 | 2.543 | -0.256 |
| H | 7.27 | -0.833 | 0.426 | H | -3.033 | 0.707 | -0.377 |
| H | 9.733 | -1.1 | 0.54 | O | -3.861 | 3.749 | -0.178 |
| C | 5.417 | -2.868 | 0.113 | N | -5.148 | 1.902 | -0.408 |
| H | 5.106 | -3.254 | -0.86 | H | -5.041 | 1.012 | -0.873 |
| H | 4.933 | -3.464 | 0.891 | C | -6.446 | 2.546 | -0.479 |
| C | 4.951 | -1.407 | 0.212 | H | -6.501 | 3.26 | -1.311 |
| H | 5.409 | -0.829 | -0.591 | H | -6.647 | 3.098 | 0.448 |
| H | 5.269 | -0.968 | 1.161 | C | -7.541 | 1.492 | -0.649 |
| N | 3.512 | -1.25 | 0.095 | H | -7.349 | 0.921 | -1.561 |
| H | 3.032 | -1.904 | -0.506 | H | -7.499 | 0.8 | 0.197 |
| C | 2.997 | 0.031 | 0.172 | C | -8.897 | 2.143 | -0.724 |
| O | 3.697 | 1.004 | 0.352 | C | -9.038 | 3.528 | -0.699 |
| N | 1.62 | 0.062 | 0.048 | C | -10.054 | 1.385 | -0.879 |
| H | 1.148 | -0.825 | 0.127 | C | -10.307 | 4.081 | -0.824 |
| C | 0.784 | 1.193 | 0.081 | H | -8.172 | 4.17 | -0.575 |
| C | 1.256 | 2.5 | 0.224 | C | -11.278 | 2.035 | -0.996 |
| C | -0.585 | 0.968 | -0.067 | H | -10.01 | 0.302 | -0.9 |
| C | 0.334 | 3.538 | 0.214 | N | -11.417 | 3.36 | -0.972 |
| H | -10.438 | 5.158 | -0.802 | H | -12.191 | 1.461 | -1.113 |

**Molecule 11**

SASA = 7.92 nm² $R_{max}$ = 26.66 Å V = 1.315 nm³



| Atom | X | Y | Z | Atom | X | Y | Z |
|---|---|---|---|---|---|---|---|
| C | 4.289 | -2.045 | 0.393 | H | 7.239 | 1.894 | -4.037 |
| C | 3.967 | -2.804 | 1.523 | H | 7.271 | 0.425 | -4.981 |
| C | 3.273 | -1.561 | -0.441 | H | 7.568 | -0.285 | -2.248 |
| C | 2.628 | -3.084 | 1.821 | H | 5.694 | -1.032 | -1.108 |
| C | 1.927 | -1.819 | -0.13 | H | 3.276 | -1.264 | -2.566 |
| C | 1.605 | -2.589 | 1 | H | 3.202 | 0.217 | -1.64 |
| C | -3.805 | -6.398 | 6.477 | H | 4.744 | -3.173 | 2.159 |
| C | -3.915 | -7.189 | 7.629 | H | 2.384 | -3.675 | 2.678 |
| C | -6.136 | -6.189 | 6.334 | H | 1.15 | -1.434 | -0.756 |
| C | -5.182 | -7.474 | 8.142 | H | -0.315 | -3.446 | 0.544 |
| C | -6.305 | -6.961 | 7.49 | H | -0.39 | -1.985 | 1.502 |
| C | 9.726 | 1.603 | -5.125 | H | 0.915 | -3.984 | 3.047 |
| C | 11.12 | 1.7 | -5.233 | H | -0.407 | -5.19 | 4.604 |
| C | 11.678 | 2.395 | -6.307 | H | -3.12 | -5.705 | 3.909 |
| C | 9.439 | 2.85 | -7.088 | H | -3.075 | -4.266 | 4.899 |
| C | 10.825 | 2.976 | -7.246 | H | 5.313 | -1.833 | 0.165 |
| N | 8.93 | 2.175 | -6.043 | C | 9.106 | 0.833 | -3.944 |
| N | -4.905 | -5.936 | 5.859 | H | 9.408 | 1.288 | -3.024 |
| C | 7.571 | 0.876 | -4.06 | H | 9.437 | -0.183 | -3.965 |
| N | 6.979 | 0.142 | -2.934 | C | -2.418 | -6.046 | 5.913 |
| C | 5.65 | 0.047 | -2.827 | H | -1.865 | -5.508 | 6.654 |
| O | 4.91 | 0.585 | -3.692 | H | -1.897 | -6.946 | 5.66 |
| N | 5.106 | -0.621 | -1.804 | H | -3.039 | -7.571 | 8.11 |
| C | 3.645 | -0.754 | -1.7 | H | -6.991 | -5.797 | 5.826 |
| C | 0.137 | -2.904 | 1.349 | H | -5.289 | -8.073 | 9.021 |
| N | 0.077 | -3.72 | 2.571 | H | -7.286 | -7.158 | 7.87 |
| C | -1.118 | -4.098 | 3.037 | H | 11.75 | 1.245 | -4.499 |
| O | -2.162 | -3.735 | 2.433 | H | 12.74 | 2.48 | -6.407 |
| N | -1.224 | -4.853 | 4.137 | H | 8.781 | 3.295 | -7.804 |
| C | -2.563 | -5.173 | 4.652 | H | 11.228 | 3.513 | -8.079 |

**Table S2:** Molecular descriptors calculated as time averages of a 50 ns molecular dynamics simulation.

| Molecule | Solvent | $\bar{SASA}$ nm² | rSASA | $\bar{R}$ (nm) | rH | HB% | $\bar{R}_g$ (nm) | F |
|---|---|---|---|---|---|---|---|---|
| 1 | Toluene | 51.25 | 0.9102 | 2.57 | 0.67 | 126.96 | 2.73 | 2.12 |
| 1 | Benzene | 50.77 | 0.9017 | 2.59 | 0.67 | 136.20 | 3.27 | 2.54 |
| 1 | Acetone | 53.42 | 0.9489 | 2.41 | 0.63 | 28.91 | 3.09 | 2.40 |
| 1 | Methanol | 51.82 | 0.9204 | 2.35 | 0.61 | 34.92 | 3.21 | 2.49 |
| 1 | Dimethylsulfoxide | 54.59 | 0.9696 | 2.37 | 0.62 | 0.00 | 3.13 | 2.44 |
| 1 | Hexane | 36.05 | 0.6403 | 2.58 | 0.67 | 868.50 | 1.24 | 0.97 |
| 2 | Toluene | 35.23 | 0.7841 | 1.71 | 0.56 | 396.27 | 2.41 | 1.98 |
| 2 | Dibutylether | 39.24 | 0.8733 | 1.60 | 0.53 | 200.84 | 2.56 | 2.11 |
| 2 | Ethanol | 42.65 | 0.9491 | 1.55 | 0.51 | 0.00 | 3.05 | 2.52 |
| 2 | Dimethylsulfoxide | 43.05 | 0.9580 | 1.56 | 0.51 | 0.03 | 2.57 | 2.12 |
| 2 | 1-propanol | 43.03 | 0.9575 | 1.67 | 0.55 | 0.00 | 2.87 | 2.36 |
| 3 | 1-propanol | 22.87 | 1.0221 | 1.03 | 0.85 | 6.96 | 2.46 | 2.63 |
| 3 | Dimethylsulfoxide | 23.12 | 1.0335 | 1.04 | 0.86 | 3.49 | 2.37 | 2.54 |
| 3 | Nitromethane | 22.03 | 0.9848 | 1.05 | 0.86 | 33.31 | 2.24 | 2.40 |
| 3 | Nitrobenzene | 22.19 | 0.9917 | 1.04 | 0.86 | 122.22 | 2.04 | 2.19 |
| 3 | 1,2-dichlorobenzene | 19.24 | 0.8599 | 1.04 | 0.86 | 418.46 | 1.88 | 2.01 |
| 3 | 1,3-dichlorobenzene | 15.05 | 0.6728 | 1.05 | 0.86 | 670.97 | 1.97 | 2.11 |
| 4 | 1-propanol | 24.05 | 0.9924 | 1.15 | 0.76 | 7.57 | 2.49 | 2.57 |
| 4 | Dimethylsulfoxide | 24.66 | 1.0176 | 1.14 | 0.76 | 3.43 | 2.31 | 2.39 |
| 4 | Dichloromethane | 18.80 | 0.7760 | 1.10 | 0.73 | 584.39 | 1.78 | 1.83 |
| 4 | Hexane | 15.70 | 0.6480 | 1.12 | 0.75 | 760.34 | 0.78 | 0.81 |
| 4 | Nitrobenzene | 24.05 | 0.9924 | 1.13 | 0.75 | 43.88 | 2.44 | 2.52 |
| 4 | 1,2-dichlorobenzene | 18.64 | 0.7694 | 1.13 | 0.75 | 614.60 | 1.31 | 1.35 |
| 5 | Ethanol | 21.76 | 0.9327 | 1.05 | 0.79 | 2.30 | 2.12 | 2.26 |
| 5 | 1-octanol | 21.39 | 0.9169 | 1.06 | 0.80 | 23.77 | 2.25 | 2.39 |
| 5 | Dimethylsulfoxide | 21.57 | 0.9247 | 1.04 | 0.78 | 7.97 | 2.12 | 2.25 |
| 5 | Water | 20.16 | 0.8642 | 1.03 | 0.78 | 15.59 | 1.87 | 1.98 |
| 5 | Hexane | 13.42 | 0.5752 | 1.05 | 0.79 | 427.05 | 0.63 | 0.67 |
| 5 | Toluene | 15.58 | 0.6676 | 1.05 | 0.79 | 315.08 | 1.17 | 1.24 |
| 6 | Water | 25.99 | 0.7322 | 1.77 | 0.96 | 58.36 | 1.92 | 1.73 |
| 6 | Dimethylsulfoxide | 33.87 | 0.9542 | 1.78 | 0.97 | 0.78 | 2.89 | 2.61 |
| 6 | 1-octanol | 34.02 | 0.9584 | 1.70 | 0.92 | 0.09 | 2.94 | 2.66 |
| 6 | Acetonitrile | 32.37 | 0.9119 | 1.78 | 0.97 | 72.74 | 2.74 | 2.48 |
| 6 | Methyl *tert*-butyl ether | 21.43 | 0.6037 | 1.78 | 0.97 | 675.07 | 1.10 | 1.00 |
| 6 | Heptane | 18.92 | 0.5329 | 1.79 | 0.97 | 1009.99 | 0.81 | 0.74 |
| 7 | Water | 23.51 | 0.6155 | 1.51 | 0.70 | 358.65 | 1.08 | 0.96 |
| 7 | Dimethylsulfoxide | 37.56 | 0.9835 | 1.62 | 0.75 | 0.21 | 2.74 | 2.42 |
| 7 | 1-octanol | 32.37 | 0.8476 | 1.86 | 0.86 | 227.53 | 2.32 | 2.05 |
| 7 | Acetonitrile | 30.14 | 0.7892 | 1.49 | 0.69 | 429.41 | 2.51 | 2.22 |
| 7 | Methyl *tert*-butyl ether | 24.46 | 0.6406 | 1.68 | 0.78 | 632.62 | 1.32 | 1.17 |
| 7 | Heptane | 17.78 | 0.4655 | 1.64 | 0.76 | 911.77 | 0.73 | 0.64 |
| 8 | Water | 29.81 | 0.7999 | 1.52 | 0.69 | 201.61 | 2.16 | 1.92 |
| 8 | Dimethylsulfoxide | 36.96 | 0.9917 | 1.63 | 0.75 | 13.94 | 2.84 | 2.53 |
| 8 | 1-octanol | 32.87 | 0.8818 | 1.80 | 0.82 | 162.40 | 3.08 | 2.74 |
| 8 | Acetonitrile | 33.12 | 0.8887 | 1.61 | 0.74 | 204.82 | 2.84 | 2.53 |
| 8 | Methyl *tert*-butyl ether | 28.06 | 0.7529 | 1.79 | 0.82 | 491.73 | 2.46 | 2.19 |
| 8 | Heptane | 19.04 | 0.5108 | 1.01 | 0.46 | 809.37 | 0.91 | 0.81 |
| 9 | Water | 25.09 | 0.6140 | 1.47 | 0.71 | 323.04 | 1.55 | 1.35 |
| 9 | Dimethylsulfoxide | 37.05 | 0.9068 | 1.52 | 0.73 | 0.80 | 2.77 | 2.41 |
| 9 | 1-octanol | 37.49 | 0.9173 | 1.53 | 0.74 | 0.00 | 3.32 | 2.88 |
| 9 | Acetonitrile | 35.07 | 0.8583 | 1.55 | 0.74 | 99.65 | 2.81 | 2.44 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 9 | Methyl tert-butyl ether | 25.41 | 0.6217 | 1.43 | 0.68 | 566.88 | 1.85 | 1.61 |
| 9 | Heptane | 16.95 | 0.4149 | 1.10 | 0.53 | 918.63 | 0.70 | 0.61 |
| 10 | Water | 24.39 | 0.6473 | 1.99 | 0.90 | 106.82 | 1.57 | 1.38 |
| 10 | Dimethylsulfoxide | 36.56 | 0.9703 | 1.93 | 0.87 | 2.35 | 3.03 | 2.67 |
| 10 | 1-octanol | 34.14 | 0.9061 | 1.95 | 0.88 | 35.41 | 2.95 | 2.60 |
| 10 | Acetonitrile | 33.26 | 0.8828 | 1.91 | 0.86 | 148.34 | 2.67 | 2.35 |
| 10 | Methyl tert-butyl ether | 22.33 | 0.5926 | 1.81 | 0.82 | 777.27 | 0.97 | 0.86 |
| 10 | Heptane | 18.81 | 0.4993 | 1.69 | 0.76 | 976.81 | 0.86 | 0.75 |
| 11 | Water | 33.77 | 0.8527 | 1.72 | 0.64 | 190.65 | 2.48 | 2.13 |
| 11 | Dimethylsulfoxide | 40.31 | 1.0179 | 2.01 | 0.75 | 1.76 | 2.99 | 2.57 |
| 11 | 1-octanol | 40.56 | 1.0241 | 1.87 | 0.70 | 0.00 | 3.01 | 2.59 |
| 11 | Acetonitrile | 37.33 | 0.9425 | 1.93 | 0.72 | 177.35 | 2.64 | 2.27 |
| 11 | Methyl *tert*-butyl ether | 29.14 | 0.7358 | 2.09 | 0.78 | 533.95 | 1.74 | 1.50 |
| 11 | Heptane | 20.27 | 0.5118 | 1.46 | 0.55 | 991.31 | 0.78 | 0.68 |

# Dataset partitioning

To ensure that the machine learning models are not over fitted, the data is partitioned into a training, validation and test set as indicated in Table S3. Note, that for the decision tree model 44 data points are used for model training, while 15 data points are left out to validate the model. This partitioning is determined by a stratified random approach over the experimental response. On the other hand, for the construction of the artificial neural network a JMP version of the random k-fold (with k = 5) cross-validation approach was employed to maximally utilize the data. Additionally, for the artificial neural network, the data is balanced by oversampling the data points with a gel or soluble response with respect to data points that were classified as a precipitate (Table S3). In both cases, all data points originating from compound **11** are used as the test set.

**Table S3:** Data partitioning for machine learning methods and experimental outcomes (G = gel, P = precipitate and S = soluble).

| Molecule | Solvent | Training/Validation /Test | k-fold | Experimental outcome | Frequency |
|---|---|---|---|---|---|
| 1 | Toluene | Training | 3 | G | 3 |
| 1 | Benzene | Training | 2 | G | 2 |
| 1 | Acetone | Training | 1 | P | 1 |
| 1 | Methanol | Training | 2 | P | 1 |
| 1 | Dimethylsulfoxide | Training | 1 | S | 2 |
| 1 | Hexane | Training | 4 | P | 1 |
| 2 | Toluene | Training | 4 | G | 3 |
| 2 | Dibutylether | Training | 5 | G | 2 |
| 2 | Ethanol | Validation | 1 | P | 1 |
| 2 | Dimethylsulfoxide | Training | 2 | G | 3 |
| 2 | 1-propanol | Training | 4 | P | 1 |
| 3 | 1-propanol | Training | 5 | S | 2 |

| | | | | | |
|---|---|---|---|---|---|
| 3 | Dimethylsulfoxide | Training | 3 | S | 2 |
| 3 | Nitromethane | Validation | 5 | G | 2 |
| 3 | Nitrobenzene | Training | 1 | G | 3 |
| 3 | 1,2-dichlorobenzene | Validation | 2 | G | 2 |
| 3 | 1,3-dichlorobenzene | Training | 4 | G | 3 |
| 4 | 1-propanol | Training | 2 | S | 2 |
| 4 | Dimethylsulfoxide | Training | 2 | S | 2 |
| 4 | Dichloromethane | Validation | 5 | P | 1 |
| 4 | Hexane | Training | 3 | P | 1 |
| 4 | Nitrobenzene | Training | 3 | G | 2 |
| 4 | 1,2-dichlorobenzene | Training | 1 | G | 3 |
| 5 | Ethanol | Validation | 5 | S | 2 |
| 5 | 1-octanol | Training | 3 | S | 2 |
| 5 | Dimethylsulfoxide | Training | 4 | S | 2 |
| 5 | Water | Training | 3 | P | 1 |
| 5 | Hexane | Training | 1 | P | 1 |
| 5 | Toluene | Training | 1 | S | 2 |
| 6 | Water | Validation | 1 | G | 3 |
| 6 | Dimethylsulfoxide | Validation | 5 | S | 2 |
| 6 | 1-octanol | Validation | 4 | S | 1 |
| 6 | Acetonitrile | Training | 2 | P | 1 |
| 6 | Methyl tert-butyl ether | Validation | 4 | P | 1 |
| 6 | Heptane | Validation | 2 | P | 1 |
| 7 | Water | Training | 3 | P | 1 |
| 7 | Dimethylsulfoxide | Training | 4 | S | 2 |
| 7 | 1-octanol | Training | 4 | P | 1 |
| 7 | Acetonitrile | Validation | 5 | P | 1 |
| 7 | Methyl tert-butyl ether | Validation | 5 | P | 1 |
| 7 | Heptane | Training | 1 | P | 1 |
| 8 | Water | Training | 3 | P | 1 |
| 8 | Dimethylsulfoxide | Training | 1 | S | 2 |
| 8 | 1-octanol | Training | 2 | P | 1 |
| 8 | Acetonitrile | Training | 4 | P | 1 |
| 8 | Methyl tert-butyl ether | Training | 3 | P | 1 |
| 8 | Heptane | Training | 2 | P | 1 |
| 9 | Water | Training | 2 | P | 1 |
| 9 | Dimethylsulfoxide | Training | 4 | S | 2 |
| 9 | 1-octanol | Training | 1 | P | 1 |
| 9 | Acetonitrile | Validation | 3 | P | 1 |
| 9 | Methyl tert-butyl ether | Training | 5 | P | 1 |
| 9 | Heptane | Training | 5 | P | 1 |
| 10 | Water | Training | 4 | P | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 10 | Dimethylsulfoxide | Validation | 2 | S | 2 |
| 10 | 1-octanol | Training | 3 | S | 2 |
| 10 | Acetonitrile | Training | 1 | P | 1 |
| 10 | Methyl tert-butyl ether | Validation | 5 | P | 1 |
| 10 | Heptane | Training | 5 | P | 1 |
| 11 | Water | Test | Test | G | N.A. |
| 11 | Dimethylsulfoxide | Test | Test | S | N.A. |
| 11 | 1-octanol | Test | Test | S | N.A. |
| 11 | Acetonitrile | Test | Test | P | N.A. |
| 11 | Methyl tert-butyl ether | Test | Test | P | N.A. |
| 11 | Heptane | Test | Test | P | N.A. |

# Decision tree model

### Optimization

The decision tree model is constructed by consecutive splits which recursively partition the data according to a relationship between the descriptors and the experimental results. The optimum model was retrieved by following the evolution of the entropy $R^2$ value of the validation data over the total number of splits (Figure S2). The total number of splits resulting in the maximum entropy $R^2$ value of the validation data set is deemed the optimal model, which in this study was equal to 5. Initially, the minimum size split was set to a count of 5 and subsequently decreased by 1 till no adequate splits were possible anymore *i.e.* they resulted in a lower entropy $R^2$ value for the training data.

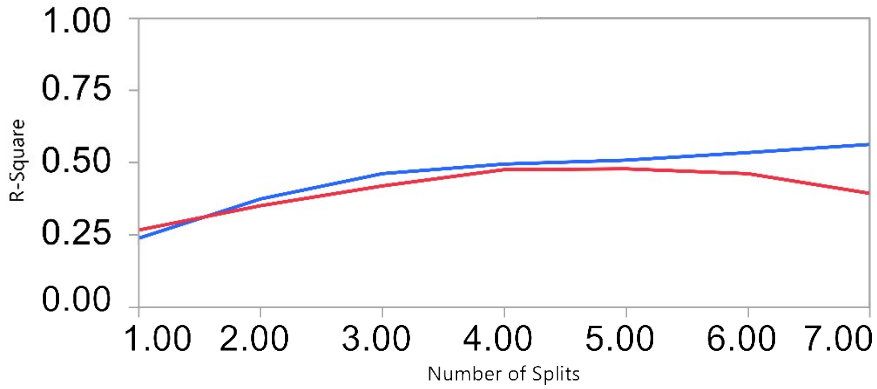



**Figure S1:** Split history. The red curve describes the evolution of the entropy $R^2$ of the validation data while the blue curve describes the evolution of the entropy $R^2$ of the training data.

**Performance**

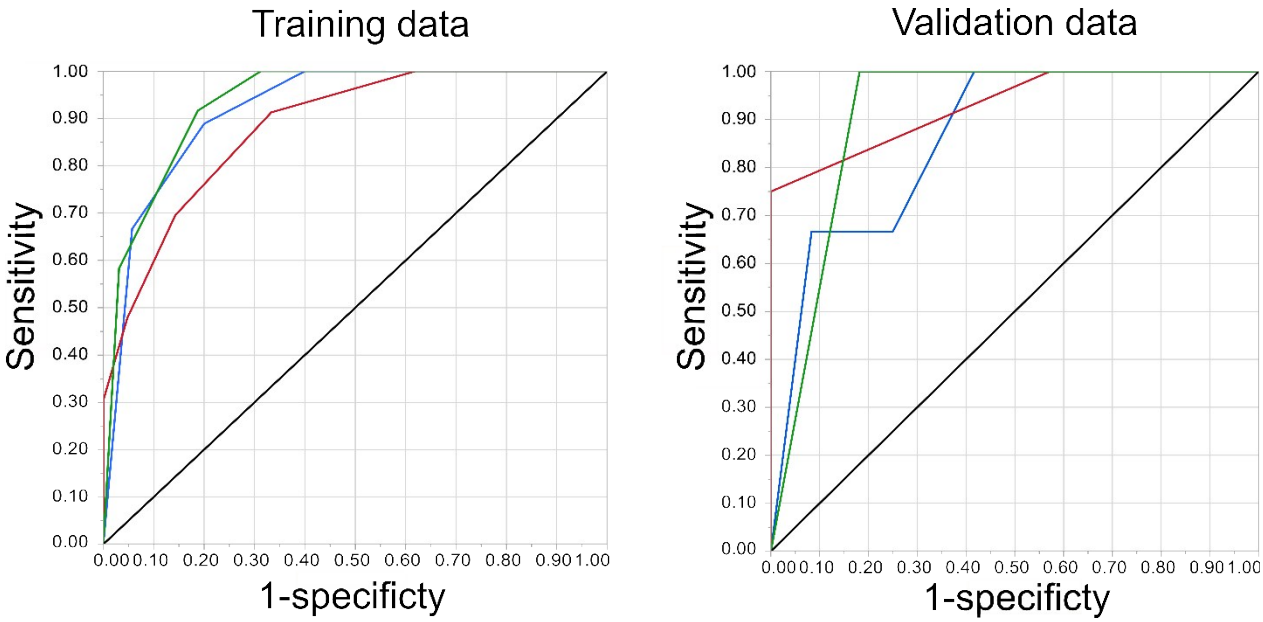



**Figure S2:** Receiver operating characteristics (ROC) for the training data (left) and validation data (right) of the optimized decision tree model. Predictions on precipitates are visualized by a red line, gels by a blue line and soluble by a green line.

**Table S4:** Confusion matrix for the optimized decision tree model.

| | **Training** | | | | **Validation** | | |
|---|---|---|---|---|---|---|---|
| | **Predicted count** | | | | **Predicted count** | | |
| **Actual** | *G* | *P* | *S* | **Actual** | *G* | *P* | *S* |
| *G* | 6 | 1 | 2 | *G* | 2 | 0 | 1 |
| *P* | 2 | 21 | 0 | *P* | 1 | 7 | 0 |
| *S* | 0 | 5 | 7 | *S* | 0 | 2 | 2 |

# Artificial neural network

**Optimization**

The artificial neural network is defined by a set of hyperparameters such as the number of neurons, activation function, number of hidden layers and the penalty method to optimize the weights. In this study we restricted the architecture of the neural network to 1 hidden layer and selected the weight decay method where the penalty function $p(\beta_i)$ is given as:

$$\sum \frac{\beta_i^2}{1+\beta_i^2}$$

A manual grid search over the other hyperparameters was performed to find the optimal network. The total number of neurons is varied between 0 and 5, while three different transformation functions were

considered being a linear identity function, a hyperbolic tangent function and a Gaussian function. The model with the lowest misclassified data of the validation set , *i.e.* the lowest misclassification rate, is considered to have the optimal settings. From Table S5 we can discern that two architectures have perfect classification of the validation data, thus to make a differentiation between the two models we looked at the misclassification rate on the training data. From this we opted for the model with 5 hyperbolic tangent neurons as this model has a slightly better misclassification rate on the training data compared to the model with 4 Gaussian neurons. Hence, the former model should be trained slightly better.

**Table S5**: Hyperparameter optimization of the artificial neural network by a manual grid search. The optimal model is marked green.

| Activation | # neurons | Misclassification rate training data | Misclassification rate validation data |
|---|---|---|---|
| Linear | 1 | 0.4324 | 0.1579 |
| Linear | 2 | 0.3514 | 0.1053 |
| Linear | 3 | 0.3684 | 0.2353 |
| Linear | 4 | 0.3553 | 0.1765 |
| Linear | 5 | 0.3649 | 0.1131 |
| tanH | 1 | 0.3194 | 0.6190 |
| tanH | 2 | 0.1486 | 0.4211 |
| tanH | 3 | 0.1622 | 0.0526 |
| tanH | 4 | 0.1711 | 0.1176 |
| tanH | 5 | 0.0263 | 0.0000 |
| Gaussian | 1 | 0.4189 | 0.3684 |
| Gaussian | 2 | 0.1892 | 0.1053 |
| Gaussian | 3 | 0.2237 | 0.1765 |
| Gaussian | 4 | 0.0270 | 0.0000 |
| Gaussian | 5 | 0.1447 | 0.1176 |

**Performance**

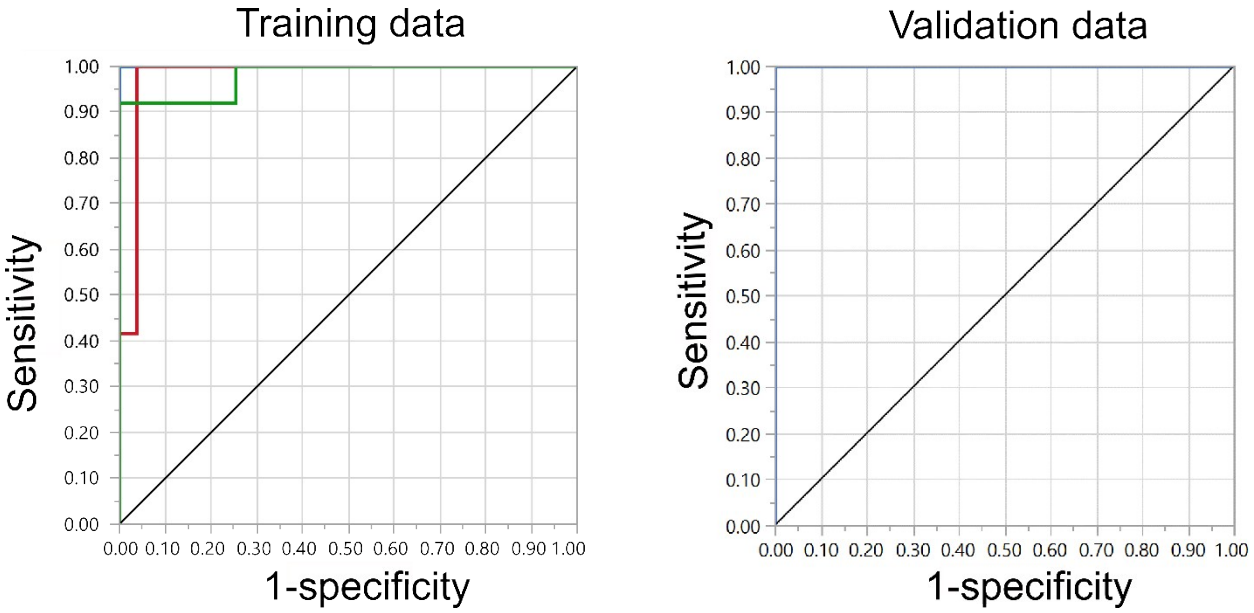



**Figure S3:** Receiver operating characteristics (ROC) for the training data (left) and validation data (right) of the optimized artificial neural network. Predictions on precipitates are visualized by a red line, gels by a blue line and soluble by a green line. The three lines on the ROC-curve of the validation data fully overlap.

**Table S6:** Confusion matrix for the optimized artificial neural network. Note that the frequency of each data point is taken into consideration for the counts (see Table S3).

| | **Training** | | | | **Validation** | | |
|---|---|---|---|---|---|---|---|
| | **Predicted count** | | | | **Predicted count** | | |
| **Actual** | *G* | *P* | *S* | **Actual** | *G* | *P* | *S* |
| *G* | 27 | 0 | 0 | *G* | 4 | 0 | 0 |
| *P* | 0 | 24 | 0 | *P* | 0 | 7 | 0 |
| *S* | 0 | 2 | 23 | *S* | 0 | 0 | 6 |

## Measures of fit

A myriad of evaluation metrics exist to quantitatively assess the quality of a predictive model and summarize discrepancies between observed and predicted outcomes. It is important to note, that different measures of fit can suggest different qualities for the same model depending on the data (dataset size, balance,…) that was used to build the model. As a result, it is crucial to recognize the features of the used data and adopt a correct measure of fit. A valid strategy to determine the quality of the predictive model, is to calculate multiple measures of fit and verify if they show a similar quality for the model. Below a definition is provided for all measures of fit that were used in this work.

The **balanced accuracy (BA)** is calculated by the average of the proportion correct predictions of each class in a classification model.[3] The main difference between the balanced accuracy and the overall accuracy is that with the BA, imbalance of the data is taken into consideration in the metric. As an example the BA is

calculated below for the confusion matrix provided in Table S6 (training set).

$$BA = \frac{\left(\frac{27}{27} + \frac{24}{24} + \frac{23}{25}\right)}{3} = 0.97$$

**Entropy $R^2$** is defined by one minus the ratio of the negative log-likelihoods of the fitted model over a hypothetical constant probability (random) model. Perfect predictive models have an entropy $R^2$ value of 1, while random models have a value of 0.[4]

**Misclassification Rate (MR)** is the probability for which the prediction does not correspond with the observed outcome.[4] As an example the MR is calculated below with the confusion matrix provided in Table S6 (training set).

$$MR = \frac{2}{76} = 0.03$$

**Cohen's Kappa (K)** classically measures inter-observer agreement.[5] However, in the case of a classification model it can also be used as a measure of fit when the observers (also called raters) are chosen as an "actual" observer and a "predictive" observer. To calculate K, the relative observed agreement between the two raters ($P_0$) and the hypothetical probability of chance agreement ($P_e$) is necessary. Again, a calculation is provided based on the data in Table S2 (training set).

$$P_0 = \frac{27 + 24 + 23}{76} = 0.97$$

$$P_e = \frac{27}{76} \cdot \frac{27}{76} + \frac{24}{76} \cdot \frac{26}{76} + \frac{25}{76} \cdot \frac{23}{76} = 0.334$$

Next, K can be calculated as follows, with a substantial agreement between two observers indicated by a value of K larger than 0.61. A value of 0 indicates an agreement that is equivalent to chance.

$$K = 1 - \frac{1 - P_0}{1 - P_e} = 0.96$$

The **area under the receiver operating characteristics curve (AUROC)** is often used as a measure to assess the quality of a machine learning predictor. AUROC values of 1 indicate perfect classification, values of 0.5 indicate classification similar to random models, while values below 0.5 indicate models which perform worse than random. To calculate the AUROC, receiver operating characteristic curves need to be plotted for every class in the prediction model, which is accomplished by depicting the sensitivity (true positive rate) on the y-axis and the 1-specificy (false positive rate) on the x-axis (Figure S3 and Figure S4).[4]

## References and notes


1. Gaussview, Version 6.1, R. Dennington, T. A. Keith, J. M. Millam, Semichem Inc., Shawnee Mission, KS, 2016
2. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.
3. K. H. Brodersen, C. S. Ong, K. E. Stephan and J. M. Buhmann, in *2010 20th International Conference on Pattern Recognition,* 2010, 3121-3124
4. JMP®, Version Pro 14. SAS Institute Inc., Cary, NC, 1989-2019.
5. M. Banerjee, M. Capozzoli, L. McSweeney and D. Sinha, *Canadian Journal of Statistics*, 1999, **27**, 3-23.