An Intuitively Understandable Quality Measure for Theoretical Vibrational Spectra

Henning Henschel*†‡ and David van der Spoel*†

†Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-75124 Uppsala, Sweden
‡present address: Department of Medicinal Chemistry, Uppsala University, Husargatan 3, Box 574, SE-75123 Uppsala, Sweden

E-mail: henning.henschel@ilk.uu.se; david.vanderspoel@icm.uu.se

Contents

S1 Comparison of Correlation Coefficients S2
S2 Directions of Comparison and Wavenumber Ranges S6
S3 Performance Depending on Database Size S11
S4 Additional Measures for Correct Matches S15
S5 List of Compounds S45
S1 Comparison of Correlation Coefficients

Figure S1: Pearson’s and Spearman’s correlation coefficients of the comparison of theoretical with experimental spectra. The linear fit in (a) is given by $y = (0.5219 \pm 0.0014) + (0.3740 \pm 0.0028) \cdot x$, in (b) by $y = (0.4553 \pm 0.0001) + (0.4625 \pm 0.0002) \cdot x$. 

(a) Comparison of corresponding spectra, wavenumber range 550–3846 cm$^{-1}$.

(b) Comparison over complete databases, wavenumber range 550–3846 cm$^{-1}$.
(a) Comparison of corresponding spectra, wavenumber range 550–2000 cm\(^{-1}\).

(b) Comparison over complete databases, wavenumber range 550–2000 cm\(^{-1}\).

Figure S2: Pearson’s and Spearman’s correlation coefficients of the comparison of theoretical with experimental spectra. The linear fit in (a) is given by \( y = (0.2485 \pm 0.0015) + (0.8325 \pm 0.0038) \cdot x \), in (b) by \( y = (0.21545 \pm 0.00001) + (0.9965 \pm 0.0002) \cdot x \).
Figure S3: Pearson’s and Spearman’s correlation coefficients of the comparison of theoretical with experimental spectra. The linear fit in (a) is given by \( y = (0.1129 \pm 0.0012) + (0.9382 \pm 0.0034) \cdot x \), in (b) by \( y = (2.17 \pm 0.03) \cdot 10^{-3} + (1.1593 \pm 0.0002) \cdot x \).
Figure S4: Pearson’s and Spearman’s correlation coefficients of the comparison of theoretical with experimental spectra. The linear fit in (a) is given by $y = (0.0804 \pm 0.0021) + (0.6100 \pm 0.0033) \cdot x$, in (b) by $y = (8.92 \pm 0.06) \cdot 10^{-3} + (0.7338 \pm 0.0001) \cdot x$. 
S2  Directions of Comparison and Wavenumber Ranges

The comparison direction theoretical spectra versus experimental database (below also ,,forward comparison") means the theoretical spectrum for one compound is compared to all experimental spectra by calculating it’s correlation coefficient with all experimental spectra. The theoretical spectrum is assigned to the compound with the experimental spectrum giving the highest correlation coefficient – in principle all theoretical spectra could be assigned to the same compound. In the comparison experimental spectra versus theoretical database (also ,,reverse comparison"), one experimental spectrum is compared to all all calculated spectra, correspondingly. Comparisons were done using four different wavenumber ranges: 550–3846 cm$^{-1}$ (,,full”), 550–2000 cm$^{-1}$ (,,medium”), 550–1650 cm$^{-1}$ (,,short”), and 2000–3846 cm$^{-1}$ (,,inverse”).

Table S1: Numbers of compounds correctly identified using various theoretical methods in combination with different statistical measures. The labels ,,full”, ,,medium”, ,,short”, and ,,high” refer to different spectroscopical ranges used in the calculations: 550–3846 cm$^{-1}$, 550–2000 cm$^{-1}$, 550–1650 cm$^{-1}$, and 2000–3846 cm$^{-1}$, respectively. Comparisons were made over 670 compounds in the database. Numbers in parentheses are the mean and median rank of the correct compound in the respective comparison.

| Level of Theory | Theory vs. Experimental Database | Spearman $\rho$ |
|----------------|-------------------------------|----------------|
|                | Pearson $r$                   | full | medium | short | high |
| B3LYP6-31G(2df,p) | 145 126 114 18 | 62 167 227 3 |
|                | (44.5; 9) (22.2; 9) (27.5; 10) (133.8; 84) | (75.5; 28) (33.3; 6) (18.0; 3) (230.0; 193) |
| B3LYP/aug-cc-pVTZ | 147 143 153 21 | 65 201 292 4 |
|                | (38.3; 11) (24.4; 9) (25.9; 8) (110.5; 67) | (69.4; 23) (25.0; 4) (12.7; 2) (222.8; 181) |
| CGenFF          | 22 5 1 14 | 15 16 15 3 |
|                | (108.8; 60) (180.1; 142) (195.7; 158) (130.0; 60) | (169.3; 104) (161.8; 102) (145.5; 95) (258.7; 232) |
| GAFF-BCC       | 1 5 3 1 | 3 8 2 1 |
|                | (282.8; 241) (215.5; 166) (219.0; 180) (350.1; 354) | (243.4; 214) (169.4; 101) (178.3; 121) (303.1; 291) |
| GAFF-ESP       | 1 0 1 0 | 0 3 4 0 |
|                | (313.7; 288) (252.8; 213) (257.8; 226) (357.1; 357) | (262.8; 230) (196.7; 152) (220.5; 180) (313.9; 309) |
| OPLS           | 2 1 1 5 | 6 4 4 1 |
|                | (230.5; 187) (275.1; 222) (303.6; 296) (216.5; 174) | (219.6; 180) (289.8; 265) (258.8; 236) (275.2; 250) |

| Level of Theory | Experiment vs. Theoretical Database | Spearman $\rho$ |
|----------------|-------------------------------|----------------|
|                | Pearson $r$                   | full | medium | short | high |
| B3LYP6-31G(2df,p) | 146 131 125 15 | 175 271 278 17 |
|                | (54.0; 13) (27.2; 8) (33.5; 12) (245.7; 106) | (44.5; 6) (12.7; 2) (14.1; 2) (185.5; 148) |
| B3LYP/aug-cc-pVTZ | 115 135 148 21 | 192 306 316 15 |
|                | (45.6; 14) (25.6; 10) (28.6; 10) (128.4; 85) | (37.7; 4) (11.1; 2) (11.8; 2) (183.5; 142) |
| CGenFF          | 14 8 4 5 | 30 23 17 7 |
|                | (143.4; 88) (152.1; 112) (177.5; 135) (199.6; 143) | (103.7; 46) (103.3; 48) (127.6; 72) (221.6; 180) |
| GAFF-BCC       | 2 3 3 0 | 8 3 6 3 |
|                | (273.7; 217) (230.8; 182) (229.5; 194) (313.9; 291) | (181.5; 124) (214.0; 136) (196.8; 158) (252.5; 209) |
Figure S5: Ranks of the spectrum of the correct compound for each method, wavenumber range 550–3846 cm\(^{-1}\).
Figure S6: Ranks of the spectrum of the correct compound for each method, wavenumber range 550–2000 cm$^{-1}$. 
Figure S7: Ranks of the spectrum of the correct compound for each method, wavenumber range 550–1650 cm$^{-1}$. 

(a) Pearson, forward comparison
(b) Pearson, reverse comparison
(c) Spearman, forward comparison
(d) Spearman, reverse comparison
Figure S8: Ranks of the spectrum of the correct compound for each method, wavenumber range 2000–3846 cm$^{-1}$.
S3  Performance Depending on Database Size

The following figures show how the fraction of correctly matched compounds varies with database size. For each database size 150 different subsamples of compounds were chosen randomly, and the matching algorithm applied as for the complete database. The shaded area in the figures corresponds to the standard deviation of the correctly matched fraction over these 150 subsamples.

(a) Pearson, forward comparison
(b) Pearson, reverse comparison
(c) Spearman, forward comparison
(d) Spearman, reverse comparison

Figure S9: Performance of methods on subsamples of the database, wavenumber range 550–3846 cm\(^{-1}\).
Figure S10: Performance of methods on subsamples of the database, wavenumber range 550–2000 cm⁻¹.
Figure S11: Performance of methods on subsamples of the database, wavenumber range 550–1650 cm\(^{-1}\).
Figure S12: Performance of methods on subsamples of the database, wavenumber range 2000–3846 cm⁻¹.
S4   Additional Measures for Correct Matches

In total we are presenting three different measures for quantifying the performance of the different methods in matching compounds. The immediate measure for this is naturally the number of correct matches $m$ – how many spectra were assigned to the correct compound or, for class matches, not necessarily the correct compound, but as a compound belonging to the same class as the correct compound. Dividing this number by the number of compounds in the database $N$, or the class $n$, respectively, gives the fraction of matches as $F = m/N$, and $F = m/n$, respectively. Especially in case of class matches, however, the fraction is a poor measure for the performance of a method, as it does not balance for the probability of randomly matching a compound correctly.

In order to account for this, we create a third measure, which gives the probability a random assignment of the spectra to compounds would have to create at least the present number of matches. The probability of generating a certain number of compound matches $k$ over the database can be expressed as

$$p_k = \binom{N}{m} p_0^k (1 - p_0)^{N-k}$$  \hspace{1cm} (1)$$

where $p_0 = 1/N$ is the elementary probability of matching any one compound correctly.

In the case of class matches, however, the situation is somewhat more intricate. We start again by defining the elementary probability $p_c = n/N$, this time for assigning any spectrum randomly to a compound of the class in question. The probability that a certain number $c$ of spectra is assigned to compounds of the given class is then, noting that in principle all spectra can be assigned to the same compound:

$$p_c = \binom{N}{c} p_c^c (1 - p_c)^{N-c}$$  \hspace{1cm} (2)$$

However, this gives only the number of spectra that are assigned to a compound of a given class, not the number of spectra that are correctly assigned to any (i.e. not necessarily the correct one) compound of the class. The probability that of the $c$ spectra assigned to compounds of a class
a certain number $k \leq c$ is correctly assigned to the class can be described as checking for the
spectra that actually represent compounds of the class, the probability that $k$ of these are assigned
to compounds of the same class:

$$p_k(c) = \binom{n}{k} \frac{c!}{(c-k)!} \frac{(N-c)!}{(N+k-n-c)!} \frac{(N-n)!}{N!}$$  \hspace{1cm} (3)$$

Combining this with the expression for $p_c$, and summing over all values for $c$ that can give $k$ correct
matches gives:

$$p_k = \sum_{c=k}^{N+k-n} \binom{N}{c} p_c^e (1-p_e)^{N-c} \binom{n}{k} \frac{c!}{(c-k)!} \frac{(N-c)!}{(N+k-n-c)!} \frac{(N-n)!}{N!}$$

$$= \sum_{c=k}^{N+k-n} \frac{N!}{c!(N-c)!} p_c^e (1-p_e)^{N-c} \frac{n!}{k!(n-k)!} \frac{c!}{(c-k)!} \frac{(N-c)!}{(N+k-n-c)!} \frac{(N-n)!}{N!}$$

$$= \sum_{c=k}^{N+k-n} p_c^e (1-p_e)^{N-c} \frac{n!}{k!(n-k)!(c-k)!(N+k-n-c)!}$$  \hspace{1cm} (4)$$

In either case, the probability of generating a certain number of correct matches is then summed
over all values equal to or larger than the actual number of matches and presented as negative
decadic logarithm a measure for predictive power (pP):

$$pP = -\log_{10} \left( \sum_{k=m}^{n} p_k \right)$$  \hspace{1cm} (5)$$

For practical reasons, we have capped the possible values of pP at 300.0.

Table S2: Correlation of theoretical spectra against database of experimental spectra in the range
550–3846 cm$^{-1}$ for B3LYP/6-31G(2df,p) sorted according to compound classification.

| Class       | N   | Pearson r | Spearman ρ |
|-------------|-----|-----------|-------------|
|             | Molecules | Classes | Molecules | Classes |
|             | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| alcohol     | 76   | 18       | 0.237    | 49       | 0.645 | 27.2  | 2        | 0.026  | 65       | 0.855 | 49.2 |
| aldehyde    | 12   | 3        | 0.250    | 5        | 0.417 | 5.9   | 5        | 0.417  | 8        | 0.667 | 11.3 |
| alkane      | 82   | 3        | 0.037    | 25       | 0.305 | 5.0   | 6        | 0.073  | 44       | 0.537 | 18.7 |
| alkene      | 174  | 43       | 0.247    | 80       | 0.460 | 8.0   | 16       | 0.092  | 106      | 0.609 | 21.5 |
| alkylbromide| 20   | 1        | 0.050    | 4        | 0.200 | 2.6   | 1        | 0.050  | 12       | 0.600 | 13.3 |

S16
Table S3: Correlation of experimental spectra against database of theoretical spectra in the range 550–3846 cm\(^{-1}\) for B3LYP/6-31G(2df,p) sorted according to compound classification.

| Class              | N     | Molecules Matched | Molecules Fraction | Classes Matched | Classes Fraction | Power |
|--------------------|-------|-------------------|-------------------|-----------------|-----------------|-------|
| alkylchloride      | 41    | 4                 | 0.098             | 19              | 0.463           | 12.2  |
| alkylfluoride      | 8     | 2                 | 0.250             | 6               | 0.750           | 10.1  |
| alkyne             | 27    | 3                 | 0.111             | 10              | 0.370           | 7.3   |
| amide              | 10    | 4                 | 0.400             | 5               | 0.500           | 6.8   |
| amine              | 41    | 10                | 0.244             | 16              | 0.390           | 9.0   |
| aromatic           | 172   | 65                | 0.378             | 121             | 0.703           | 33.7  |
| allylchloride      | 23    | 4                 | 0.174             | 8               | 0.348           | 6.2   |
| allylfluoride      | 15    | 2                 | 0.154             | 8               | 0.615           | 10.6  |
| cycloalkane        | 39    | 2                 | 0.051             | 12              | 0.308           | 5.9   |
| cycloalkene        | 11    | 6                 | 0.545             | 6               | 0.545           | 8.1   |
| halogenated compound | 105  | 14                | 0.133             | 62              | 0.590           | 23.3  |
| heterocyclic       | 57    | 35                | 0.614             | 70              | 0.702           | 29.3  |
| inorganic          | 3     | 0                 | 0.000             | 0               | 0.000           | 0.0   |
| ketone             | 33    | 8                 | 0.242             | 23              | 0.697           | 22.3  |
| nitro              | 10    | 2                 | 0.200             | 7               | 0.700           | 10.7  |
| phenol             | 18    | 7                 | 0.389             | 12              | 0.667           | 14.6  |
| thiol              | 16    | 4                 | 0.250             | 4               | 0.250           | 3.3   |
| all molecules      | 670   | 145               | 0.216             | —               | —               | 259.6 |

| Class              | N     | Molecules Matched | Molecules Fraction | Classes Matched | Classes Fraction | Power |
|--------------------|-------|-------------------|-------------------|-----------------|-----------------|-------|
| alcohol            | 76    | 13                | 0.171             | 60              | 0.709           | 41.5  |
| aldehyde           | 12    | 4                 | 0.333             | 7               | 0.583           | 9.4   |
| alkane             | 82    | 1                 | 0.012             | 12              | 0.146           | 0.5   |
| alkenne            | 174   | 41                | 0.236             | 111             | 0.638           | 24.9  |
| alkythromide       | 20    | 0                 | 0.000             | 2               | 0.100           | 0.9   |
| alkylchloride      | 41    | 4                 | 0.098             | 21              | 0.512           | 14.6  |
| alkylfluoride      | 8     | 1                 | 0.125             | 4               | 0.500           | 5.9   |
| alkyne             | 27    | 0                 | 0.000             | 3               | 0.111           | 1.0   |
| amide              | 10    | 4                 | 0.400             | 7               | 0.700           | 10.7  |
| amine              | 41    | 10                | 0.244             | 16              | 0.390           | 9.0   |
| aromatic           | 172   | 69                | 0.401             | 137             | 0.797           | 48.8  |
| allylchloride      | 23    | 4                 | 0.174             | 11              | 0.478           | 10.1  |
| allylfluoride      | 13    | 6                 | 0.462             | 12              | 0.923           | 19.4  |
| cycloalkane        | 39    | 3                 | 0.077             | 27              | 0.692           | 24.1  |
| cycloalkene        | 11    | 5                 | 0.455             | 9               | 0.818           | 14.3  |
| halogenated compound | 105  | 17                | 0.162             | 63              | 0.600           | 24.2  |
| heterocyclic       | 57    | 38                | 0.667             | 43              | 0.754           | 33.3  |
| inorganic          | 3     | 0                 | 0.000             | 0               | 0.000           | 0.0   |
| ketone             | 33    | 9                 | 0.273             | 27              | 0.818           | 29.4  |
| nitro              | 10    | 3                 | 0.300             | 6               | 0.600           | 8.7   |
| phenol             | 18    | 5                 | 0.278             | 12              | 0.667           | 14.6  |
| thiol              | 16    | 1                 | 0.062             | 3               | 0.188           | 2.2   |
| all molecules      | 670   | 146               | 0.218             | —               | —               | 261.8 |

| Class              | N     | Molecules Matched | Molecules Fraction | Classes Matched | Classes Fraction | Power |
|--------------------|-------|-------------------|-------------------|-----------------|-----------------|-------|
| alcohol            | 76    | 14                | 0.184             | 71              | 0.934           | 60.1  |
| aldehyde           | 12    | 11                | 0.917             | 11              | 0.917           | 18.1  |
| alkane             | 82    | 8                 | 0.088             | 19              | 0.232           | 2.4   |
| alkenne            | 174   | 53                | 0.305             | 124             | 0.713           | 34.9  |
| alkythromide       | 20    | 0                 | 0.000             | 1               | 0.050           | 0.3   |
| alkylchloride      | 41    | 6                 | 0.146             | 22              | 0.537           | 15.8  |
| alkylfluoride      | 8     | 2                 | 0.250             | 6               | 0.750           | 10.1  |
| alkyne             | 27    | 8                 | 0.296             | 15              | 0.556           | 13.9  |
| amide              | 10    | 5                 | 0.500             | 7               | 0.700           | 10.7  |
| amine              | 41    | 28                | 0.683             | 35              | 0.854           | 36.0  |
| aromatic           | 172   | 69                | 0.401             | 159             | 0.924           | 76.5  |
| allylchloride      | 23    | 8                 | 0.348             | 12              | 0.522           | 11.6  |
| allylfluoride      | 13    | 5                 | 0.385             | 8               | 0.615           | 10.6  |
| cycloalkane        | 39    | 4                 | 0.103             | 6               | 0.154           | 1.6   |
| cycloalkene        | 11    | 3                 | 0.273             | 3               | 0.273           | 3.2   |
| halogenated compound | 105  | 22                | 0.210             | 63              | 0.600           | 24.2  |
| heterocyclic       | 57    | 20                | 0.351             | 28              | 0.491           | 14.9  |
| inorganic          | 3     | 2                 | 0.667             | 2               | 0.667           | 4.2   |
| ketone             | 33    | 10                | 0.303             | 23              | 0.697           | 22.3  |
| nitro              | 10    | 1                 | 0.100             | 5               | 0.500           | 6.8   |
| phenol             | 18    | 10                | 0.556             | 14              | 0.778           | 18.5  |
| thiol              | 16    | 3                 | 0.188             | 3               | 0.188           | 2.2   |
| all molecules      | 670   | 175               | 0.261             | —               | —               | 300.0 |
Table S4: Correlation of theoretical spectra against database of experimental spectra in the range 550–3846 cm\(^{-1}\) for B3LYP/aug-cc-pVTZ sorted according to compound classification.

| Class           | N  | Molecules | Pearson \(r\) | Spearman \(\rho\) |
|-----------------|----|-----------|----------------|-------------------|
| N               |    | Matched   | Fraction       | Matched           | Fraction         | Power  |
| alcohol         | 76 | 11        | 0.145          | 48                | 0.632            | 26.1   |
| aldehyde        | 12 | 1         | 0.083          | 3                 | 0.250            | 3.0    |
| alkane          | 82 | 4         | 0.049          | 18                | 0.220            | 2.0    |
| alkenes         | 174| 49        | 0.282          | 99                | 0.569            | 17.2   |
| alkyl bromide   | 20 | 2         | 0.100          | 5                 | 0.250            | 3.6    |
| alkyl chloride  | 41 | 2         | 0.049          | 13                | 0.317            | 6.2    |
| alkyl fluoride  | 8  | 1         | 0.125          | 6                 | 0.750            | 10.1   |
| alkyne          | 27 | 4         | 0.148          | 13                | 0.481            | 11.1   |
| amides          | 10 | 0         | 0.000          | 5                 | 0.500            | 6.8    |
| alkenes         |    | 17        | 0.415          | 28                | 0.683            | 24.1   |
| aromatic        | 172| 65        | 0.378          | 141               | 0.820            | 53.1   |
| chlorinated     | 23 | 3         | 0.130          | 10                | 0.435            | 8.8    |
| aromatic        | 172| 13        | 0.000          | 5                 | 0.385            | 5.5    |
| cycloalkene     | 39 | 6         | 0.154          | 14                | 0.359            | 7.7    |
| alkyl bromide   | 27 | 1         | 0.000          | 2                 | 0.100            | 0.9    |
| amides          | 10 | 1         | 0.333          | 1                 | 0.333            | 1.9    |
| ketones         | 33 | 5         | 0.152          | 15                | 0.455            | 11.0   |
| nitro           | 10 | 0         | 0.000          | 4                 | 0.400            | 5.0    |
| phenols         | 18 | 2         | 0.111          | 9                 | 0.500            | 9.5    |
| thiol           | 16 | 4         | 0.250          | 6                 | 0.375            | 5.9    |
| all molecules   | 670| 143       | 0.213          | —                 | —                | 255.0  |

Table S5: Correlation of experimental spectra against database of theoretical spectra in the range 550–3846 cm\(^{-1}\) for B3LYP/aug-cc-pVTZ sorted according to compound classification.

| Class           | N  | Molecules | Pearson \(r\) | Spearman \(\rho\) |
|-----------------|----|-----------|----------------|-------------------|
| N               |    | Matched   | Fraction       | Matched           | Fraction         | Power  |
| alcohol         | 76 | 11        | 0.118          | 52                | 0.664            | 30.8   |
| aldehyde        | 12 | 2         | 0.167          | 3                 | 0.250            | 3.0    |
| alkane          | 82 | 1         | 0.012          | 9                 | 0.110            | 0.2    |
| alkenes         | 174| 40        | 0.230          | 119               | 0.684            | 30.8   |
| alkyl bromide   | 20 | 1         | 0.050          | 2                 | 0.100            | 0.9    |
| alkyl chloride  | 41 | 1         | 0.024          | 13                | 0.317            | 6.2    |
| alkyl fluoride  | 8  | 1         | 0.125          | 4                 | 0.500            | 5.9    |
| alkyne          | 27 | 1         | 0.037          | 9                 | 0.333            | 6.2    |
| amides          | 10 | 0         | 0.000          | 5                 | 0.500            | 6.8    |
| aromatic        | 172| 63        | 0.366          | 134               | 0.779            | 45.7   |
| aromatic        | 172| 23        | 0.348          | 13                | 0.565            | 13.1   |
| aromatic        | 172| 13        | 0.000          | 6                 | 0.462            | 7.1    |
| aromatic        | 172| 39        | 0.103          | 26                | 0.667            | 22.5   |
| aromatic        | 172| 11        | 0.455          | 8                 | 0.727            | 12.1   |
| aromatic        | 172| 10        | 0.095          | 51                | 0.486            | 14.5   |
| aromatic        | 172| 57        | 0.491          | 36                | 0.632            | 24.0   |
| aromatic        | 172| 3         | 0         | 0.000          | 0                 | 0.000  |
| Class        | N | Pearson $r$ | Spearman $\rho$ |
|-------------|---|-------------|----------------|
|              |   | Molecules  | Classes       | Molecules  | Classes       |
|              |   | Matched   | Fraction | Matched   | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| ketone       | 33| 5 0.152   | 17 0.515 | 12 0.364 | 22 0.667 | 20.7 |
| nitro        | 10| 0 0.000   | 9 0.900  | 1 0.100  | 6 0.600  | 8.7  |
| phenol       | 18| 3 0.167   | 3 0.500  | 9 0.500  | 12 0.667 | 14.6 |
| thiol        | 16| 1 0.062   | 2 0.125  | 5 0.312  | 6 0.375  | 5.9  |
| all molecules| 670| 115 0.172 | — — | 192 0.287 | — — | 300.0 |

Table S6: Correlation of theoretical spectra against database of experimental spectra in the range 550–3846 cm$^{-1}$ for CGenFF sorted according to compound classification.

| Class        | N | Pearson $r$ | Spearman $\rho$ |
|-------------|---|-------------|----------------|
|              |   | Molecules  | Classes       | Molecules  | Classes       |
|              |   | Matched   | Fraction | Matched   | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| alcohol      | 76| 1 0.013   | 14 0.184 | 3 0.039  | 67 0.882 | 52.6 |
| aldehyde     | 12| 0 0.000   | 0 0.000  | 0 0.000  | 0 0.000  | 0.0  |
| alkane       | 82| 2 0.024   | 18 0.220 | 0 0.000  | 2 0.024  | 0.0  |
| alkene       | 174| 5 0.029 | 96 0.552 | 3 0.017 | 136 0.782 | 46.0 |
| alkylbromide | 20| 0 0.000   | 0 0.000  | 0 0.000  | 0 0.000  | 0.0  |
| alkylchloride| 41| 0 0.000   | 2 0.049  | 0 0.000  | 1 0.125  | 1.0  |
| alkylfluoride| 8 | 0 0.000   | 0 0.000  | 0 0.000  | 1 0.125  | 1.0  |
| alkyne       | 27| 2 0.074   | 12 0.444 | 3 0.111  | 4 0.148  | 1.7  |
| amide        | 10| 0 0.000   | 0 0.000  | 2 0.200  | 4 0.400  | 5.0  |
| amine        | 41| 0 0.000   | 4 0.098  | 1 0.024  | 4 0.098  | 0.6  |
| aromatic     | 172| 9 0.052 | 137 0.797 | 3 0.017 | 55 0.320 | 14.0 |
| arylchloride | 23| 0 0.000   | 1 0.043  | 0 0.000  | 0 0.000  | 0.0  |
| arylfluoride | 13| 0 0.000   | 0 0.000  | 0 0.000  | 0 0.000  | 0.0  |
| cycloalkane  | 39| 1 0.026   | 2 0.051  | 0 0.000  | 0 0.000  | 0.0  |
| cycloalkene  | 11| 0 0.000   | 1 0.091  | 0 0.000  | 0 0.000  | 0.0  |
| halogenated compound | 105| 0 0.000 | 11 0.105 | 2 0.019 | 6 0.057 | 0.0  |
| heterocyclic | 57| 6 0.105   | 8 0.140  | 1 0.018  | 5 0.088  | 0.3  |
| inorganic    | 3 | 0 0.000   | 0 0.000  | 0 0.000  | 0 0.000  | 0.0  |
| ketone       | 33| 3 0.091   | 8 0.242  | 0 0.000  | 3 0.091  | 0.7  |
| nitro        | 10| 0 0.000   | 0 0.000  | 0 0.000  | 2 0.200  | 2.0  |
| phenol       | 18| 2 0.111   | 4 0.222  | 0 0.000  | 3 0.167  | 1.9  |
| thiol        | 16| 0 0.000   | 7 0.438  | 1 0.062  | 11 0.688 | 14.2 |
| all molecules| 670| 22 0.033 | — — | 15 0.022 | — — | 12.6 |

Table S7: Correlation of experimental spectra against database of theoretical spectra in the range 550–3846 cm$^{-1}$ for CGenFF sorted according to compound classification.
Table S8: Correlation of theoretical spectra against database of experimental spectra in the range 550–3846 cm\(^{-1}\) for GAFF-BCC sorted according to compound classification.

| Class                | N   | Pearson \(r\) | Spearman \(\rho\) |
|----------------------|-----|---------------|------------------|
|                      | Molecules | Classes     | Molecules | Classes   |
|                      | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| amide                | 10    | 2        | 0.200   | 4        | 0.400 | 5.0    | 2        | 0.200   | 2        | 0.200 | 2.0   |
| amine                | 41    | 0        | 0.000   | 2        | 0.049 | 0.1    | 0        | 0.000   | 5        | 0.122 | 1.0   |
| aromatic             | 172   | 2        | 0.012   | 104      | 0.605 | 21.1   | 6        | 0.035   | 110      | 0.640 | 25.2  |
| arylchloride         | 23    | 0        | 0.000   | 4        | 0.174 | 2.1    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| arythiourea          | 13    | 0        | 0.000   | 1        | 0.077 | 0.6    | 0        | 0.000   | 1        | 0.077 | 0.6   |
| cycloalkane          | 39    | 0        | 0.000   | 21       | 0.538 | 15.6   | 1        | 0.026   | 3        | 0.077 | 0.4   |
| cycloalkene          | 11    | 0        | 0.000   | 1        | 0.091 | 0.8    | 0        | 0.000   | 1        | 0.091 | 0.8   |
| halogenated compound | 105   | 0        | 0.000   | 20       | 0.390 | 0.7    | 4        | 0.038   | 12       | 0.114 | 0.0   |
| heterocyclic         | 57    | 0        | 0.000   | 13       | 0.228 | 3.1    | 2        | 0.035   | 15       | 0.263 | 4.2   |
| inorganic            | 3     | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| ketone               | 33    | 2        | 0.061   | 17       | 0.515 | 13.5   | 3        | 0.091   | 9        | 0.273 | 4.7   |
| nitro                | 10    | 0        | 0.000   | 1        | 0.100 | 0.9    | 0        | 0.000   | 1        | 0.100 | 0.9   |
| phenol               | 18    | 0        | 0.000   | 1        | 0.056 | 0.4    | 2        | 0.111   | 7        | 0.389 | 6.6   |
| thiol                | 16    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 1        | 0.062 | 0.5   |
| all molecules        | 670   | 14       | 0.021   | —        | —     | 11.4   | 30       | 0.045   | —        | —     | 33.1  |

| Class                | N   | Pearson \(r\) | Spearman \(\rho\) |
|----------------------|-----|---------------|------------------|
|                      | Molecules | Classes     | Molecules | Classes   |
|                      | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| alcohol              | 76    | 0        | 0.000   | 22       | 0.289 | 4.6    | 0        | 0.000   | 3        | 0.039 | 0.0   |
| aldehyde             | 12    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| alkane               | 82    | 1        | 0.012   | 3        | 0.037 | 0.0    | 0        | 0.000   | 5        | 0.061 | 0.0   |
| alkeene              | 174   | 0        | 0.000   | 15       | 0.086 | 0.0    | 1        | 0.006   | 109      | 0.626 | 23.5  |
| alkylbromide         | 20    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| alkylchloride        | 41    | 0        | 0.000   | 3        | 0.073 | 0.3    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| alkylfluoride        | 8     | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| alkyne               | 27    | 0        | 0.000   | 2        | 0.074 | 0.5    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| amide                | 10    | 0        | 0.000   | 1        | 0.100 | 0.9    | 0        | 0.000   | 3        | 0.300 | 3.4   |
| amine                | 41    | 0        | 0.000   | 3        | 0.073 | 0.3    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| aromatic             | 172   | 0        | 0.000   | 35       | 0.203 | 0.0    | 1        | 0.006   | 38       | 0.221 | 0.1   |
| arylchloride         | 23    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| arythiourea          | 13    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| cycloalkane          | 39    | 0        | 0.000   | 2        | 0.051 | 0.2    | 1        | 0.026   | 3        | 0.077 | 0.4   |
| cycloalkene          | 11    | 0        | 0.000   | 1        | 0.091 | 0.8    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| halogenated compound | 105   | 0        | 0.000   | 8        | 0.076 | 0.0    | 0        | 0.000   | 2        | 0.019 | 0.0   |
| heterocyclic         | 57    | 0        | 0.000   | 11       | 0.193 | 2.1    | 0        | 0.000   | 4        | 0.070 | 0.1   |
| inorganic            | 3     | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| ketone               | 33    | 0        | 0.000   | 3        | 0.091 | 0.7    | 0        | 0.000   | 3        | 0.091 | 0.7   |
| nitro                | 10    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| phenol               | 18    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| thiol                | 16    | 0        | 0.000   | 0        | 0.000 | 0.0    | 0        | 0.000   | 0        | 0.000 | 0.0   |
| all molecules        | 670   | 1        | 0.001   | —        | —     | 0.2    | 3        | 0.004   | —        | —     | 1.1   |

S20
Table S9: Correlation of experimental spectra against database of theoretical spectra in the range 550–3846 cm$^{-1}$ for GAFF-BCC sorted according to compound classification.

| Class         | N  | Molecules Matched | Fraction | Molecules Matched | Fraction | Pearson $r$ | Spearman $\rho$ |
|---------------|----|-------------------|----------|-------------------|----------|-------------|-----------------|
|               |    |                   |          |                   |          |             |                 |
| alcohol       | 76 | 0                  | 0.000    | 4                 | 0.053    | 0.0000      | 0.013           |
| aldehyde      | 12 | 1                  | 0.083    | 4                 | 0.333    | 0.0000      | 0.000           |
| alkane        | 82 | 0                  | 0.000    | 3                 | 0.037    | 0.0000      | 0.024           |
| alkenes       | 174| 0                  | 0.000    | 119               | 0.684    | 30.8        | 0.000           |
| alkythromide  | 20 | 0                  | 0.000    | 1                 | 0.050    | 0.0000      | 0.000           |
| alkylchloride | 41 | 0                  | 0.000    | 4                 | 0.098    | 0.6         | 0.000           |
| alkylfluoride | 8  | 0                  | 0.000    | 2                 | 0.250    | 2.4         | 0.000           |
| alkyne        | 27 | 0                  | 0.000    |                   | 0.000    | 0.0000      | 0.000           |
| amide         | 10 | 0                  | 0.000    |                   | 0.000    | 0.0000      | 0.000           |
| amine         | 41 | 0                  | 0.000    |                   | 0.000    | 0.0000      | 0.000           |
| aromatic      | 172| 0                  | 0.000    | 24                | 0.140    | 0.0         | 0.000           |
| arylchloride  | 23 | 0                  | 0.000    |                   | 0.000    | 0.0000      | 0.000           |
| arylfluoride  | 13 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| cycloalkane   | 39 | 0                  | 0.000    | 3                 | 0.077    | 0.4         | 0.000           |
| cycloalkene   | 11 | 0                  | 0.000    | 1                 | 0.091    | 0.8         | 0.000           |
| halogenated   | 105| 0                  | 0.000    | 25                | 0.238    | 1.7         | 1.010           |
| heterocyclic  | 57 | 1                  | 0.018    | 10                | 0.175    | 1.7         | 0.000           |
| inorganic     | 3  | 0                  | 0.000    |                   | 0.000    | 0.0000      | 0.000           |
| ketone        | 33 | 1                  | 0.030    | 8                 | 0.242    | 3.8         | 0.000           |
| nitro         | 10 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| phenol        | 18 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| thiol         | 16 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| all molecules | 670| 2                  | 0.003    |                   | 0.000    | 0.6         | 0.012           |

Table S10: Correlation of theoretical spectra against database of experimental spectra in the range 550–3846 cm$^{-1}$ for GAFF-ESP sorted according to compound classification.

| Class         | N  | Molecules Matched | Fraction | Molecules Matched | Fraction | Pearson $r$ | Spearman $\rho$ |
|---------------|----|-------------------|----------|-------------------|----------|-------------|-----------------|
|               |    |                   |          |                   |          |             |                 |
| alcohol       | 76 | 0                  | 0.000    | 28                | 0.368    | 8.2         | 0.000           |
| aldehyde      | 12 | 0                  | 0.000    |                   | 0.000    | 0           | 0.000           |
| alkane        | 82 | 1                  | 0.012    | 10                | 0.122    | 0.3         | 0.000           |
| alkenes       | 174| 0                  | 0.000    | 17                | 0.098    | 0.0         | 0.000           |
| alkythromide  | 20 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| alkylchloride | 41 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| alkylfluoride | 8  | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| alkyne        | 27 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| amide         | 10 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| amine         | 41 | 0                  | 0.000    | 4                 | 0.098    | 0.6         | 0.000           |
| aromatic      | 172| 0                  | 0.000    | 32                | 0.186    | 0.0         | 0.000           |
| arylchloride  | 23 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| arylfluoride  | 13 | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
| cycloalkane   | 39 | 0                  | 0.000    | 8                 | 0.205    | 2.8         | 0.000           |
| cycloalkene   | 11 | 0                  | 0.000    | 1                 | 0.091    | 0.8         | 0.000           |
| halogenated   | 105| 0                  | 0.000    | 13                | 0.124    | 0.1         | 0.000           |
| heterocyclic  | 57 | 0                  | 0.000    | 4                 | 0.070    | 0.1         | 0.000           |
| inorganic     | 3  | 0                  | 0.000    |                   | 0.000    | 0.0         | 0.000           |
Class | N | Pearson r | Spearman ρ
|----|----|----------|----------|
|    |    | Molecules | Classes | Molecules | Classes |
|    |    | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| ketone | 33 | 0 | 0.000 | 1 | 0.030 | 0.1 | 0 | 0.000 | 3 | 0.091 | 0.7 |
| nitro | 10 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| phenol | 18 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| thiol | 16 | 0 | 0.000 | 2 | 0.125 | 1.3 | 0 | 0.000 | 1 | 0.062 | 0.5 |
| all molecules | 670 | 1 | 0.001 | — | — | 0.2 | 0 | 0.000 | — | — | 0.0 |

Table S11: Correlation of experimental spectra against database of theoretical spectra in the range 550–3846 cm\(^{-1}\) for GAFF-ESP sorted according to compound classification.

Class | N | Pearson r | Spearman ρ
|----|----|----------|----------|
|    |    | Molecules | Classes | Molecules | Classes |
|    |    | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| alcohol | 76 | 1 | 0.013 | 8 | 0.105 | 0.2 | 0 | 0.000 | 18 | 0.237 | 2.7 |
| aldehyde | 12 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 2 | 0.167 | 1.7 |
| alkane | 82 | 0 | 0.000 | 12 | 0.146 | 0.5 | 1 | 0.012 | 24 | 0.293 | 4.5 |
| alkene | 174 | 0 | 0.000 | 98 | 0.563 | 16.6 | 1 | 0.006 | 52 | 0.299 | 0.9 |
| alkylbromide | 20 | 0 | 0.000 | 1 | 0.050 | 0.3 | 0 | 0.000 | 10 | 0.500 | 10.1 |
| alkylchloride | 41 | 0 | 0.000 | 5 | 0.122 | 1.0 | 0 | 0.000 | 5 | 0.122 | 1.0 |
| alkylfluoride | 8 | 1 | 0.125 | 2 | 0.250 | 2.4 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| alkyne | 27 | 0 | 0.000 | 0 | 0.000 | 0.0 | 1 | 0.037 | 1 | 0.037 | 0.2 |
| amide | 10 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| amine | 41 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| aromatic | 172 | 0 | 0.000 | 49 | 0.285 | 0.7 | 2 | 0.012 | 62 | 0.360 | 2.8 |
| arylchloride | 23 | 0 | 0.000 | 1 | 0.043 | 0.3 | 1 | 0.043 | 4 | 0.174 | 2.1 |
| arylfluoride | 13 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| cycloalkane | 39 | 0 | 0.000 | 4 | 0.103 | 0.7 | 0 | 0.000 | 1 | 0.026 | 0.0 |
| cycloalkene | 11 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 1 | 0.091 | 0.8 |
| halogenated compound | 105 | 1 | 0.010 | 25 | 0.238 | 1.7 | 2 | 0.019 | 31 | 0.295 | 3.6 |
| heterocyclic | 57 | 0 | 0.000 | 10 | 0.175 | 1.7 | 2 | 0.035 | 12 | 0.211 | 2.6 |
| inorganic | 3 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| ketone | 33 | 0 | 0.000 | 8 | 0.242 | 3.8 | 0 | 0.000 | 3 | 0.091 | 0.7 |
| nitro | 10 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| phenol | 18 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| thiol | 16 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| all molecules | 670 | 1 | 0.001 | — | — | 0.2 | 8 | 0.012 | — | — | 5.0 |

Table S12: Correlation of theoretical spectra against database of experimental spectra in the range 550–3846 cm\(^{-1}\) for OPLS sorted according to compound classification.
Table S13: Correlation of experimental spectra against database of theoretical spectra in the range 550–3846 cm\(^{-1}\) for OPLS sorted according to compound classification.

| Class                | N  | Pearson | Spearman |                  |
|----------------------|----|---------|----------|-----------------|
|                      |    | Molecules | Classes | Molecules | Classes |
|                      |    | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| amide                | 10 | 0      | 0.000    | 0      | 0.000    | 0.0   | 2       | 0.200    | 3       | 0.300    | 3.4   |
| amine                | 41 | 0      | 0.000    | 1      | 0.024    | 0.0   | 0       | 0.000    | 3       | 0.073    | 0.3   |
| aromatic             | 172| 0      | 0.000    | 134    | 0.779    | 45.7  | 2       | 0.012    | 43      | 0.250    | 0.2   |
| aryl chloride        | 23 | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| arythioflouride       | 13 | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| cycloalkane          | 39 | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| cycloalkene          | 11 | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| halogenated compound | 105| 1      | 0.010    | 16     | 0.152    | 0.2   | 1       | 0.010    | 6       | 0.057    | 0.0   |
| heterocyclic         | 57 | 0      | 0.000    | 15     | 0.263    | 4.2   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| inorganic            | 3  | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| ketone               | 33 | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| nitro                | 10 | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| phenol               | 18 | 0      | 0.000    | 0      | 0.000    | 0.0   | 0       | 0.000    | 0       | 0.000    | 0.0   |
| thiol                | 16 | 0      | 0.000    | 1      | 0.062    | 0.5   | 2       | 0.125    | 11      | 0.688    | 14.2  |
| all molecules        | 670| 2      | 0.003    | —      | —       | 0.6   | 6       | 0.009    | —       | —       | 3.2   |

S23
Table S14: Correlation of theoretical spectra against database of experimental spectra in the range 550–2000 cm$^{-1}$ for B3LYP/6-31G(2df,p) sorted according to compound classification.

| Class             | N      | Molecules          | Pearson r | Spearman ρ |
|-------------------|--------|--------------------|-----------|------------|
|                   |        | Matched Fraction   | Matched Fraction | Power |
| alcohol           | 76     | 13 0.171           | 60 0.799   | 41.5      |
| aldehyde          | 12     | 2 0.167            | 4 0.333   | 4.3       |
| alkane            | 82     | 3 0.037            | 10 0.122  | 0.3       |
| alkene            | 174    | 50 0.287           | 131 0.753 | 41.2      |
| alkyllithium      | 20     | 0 0.000            | 6 0.300   | 4.7       |
| alkylchlooride    | 41     | 2 0.049            | 14 0.341  | 7.1       |
| alkylfluoride     | 8      | 1 0.125            | 3 0.375   | 4.0       |
| alkyne            | 27     | 3 0.111            | 5 0.185   | 2.4       |
| amide             | 10     | 2 0.200            | 5 0.500   | 6.8       |
| amine             | 41     | 5 0.122            | 8 0.395   | 2.5       |
| aromatic          | 172    | 46 0.267           | 104 0.605 | 21.1      |
| vinyl             | 12     | 2 0.154            | 8 0.615   | 10.6      |
| cyano              | 39     | 7 0.179            | 12 0.308  | 5.9       |
| cyklene           | 11     | 5 0.455            | 5 0.455   | 6.3       |
| halogenated compound | 105 | 9 0.086               | 46 0.438  | 11.1      |
| heterocyclic      | 57     | 27 0.474           | 34 0.596  | 21.5      |
| inorganic         | 3      | 0 0.000            | 0 0.000   | 0.0       |
| ketone            | 33     | 3 0.091            | 13 0.394  | 8.6       |
| nitro             | 10     | 2 0.200            | 6 0.600   | 8.7       |
| phenol            | 18     | 5 0.278            | 12 0.667  | 14.6      |
| thiol             | 16     | 0 0.000            | 3 0.188   | 2.2       |
| all molecules     | 670    | 126 0.188          | —          | 217.2     |

Table S15: Correlation of experimental spectra against database of theoretical spectra in the range 550–2000 cm$^{-1}$ for B3LYP/6-31G(2df,p) sorted according to compound classification.

| Class             | N      | Molecules          | Pearson r | Spearman ρ |
|-------------------|--------|--------------------|-----------|------------|
|                   |        | Matched Fraction   | Matched Fraction | Power |
| alcohol           | 76     | 15 0.197           | 60 0.799   | 41.5      |
| aldehyde          | 12     | 2 0.167            | 7 0.583   | 9.4       |
| alkane            | 82     | 5 0.061            | 48 0.585  | 22.6      |
| alkene            | 174    | 49 0.282           | 125 0.718 | 35.8      |
| alkyllithium      | 20     | 0 0.000            | 1 0.050   | 0.3       |
| alkylchlooride    | 41     | 0 0.000            | 8 0.395   | 2.5       |
| alkylfluoride     | 8      | 1 0.125            | 3 0.375   | 4.0       |
| alkyne            | 27     | 1 0.037            | 2 0.074   | 0.5       |
| amide             | 10     | 4 0.400            | 6 0.600   | 8.7       |
| amine             | 41     | 5 0.122            | 9 0.220   | 3.2       |
| aromatic          | 172    | 54 0.314           | 139 0.808 | 50.9      |
| vinyl             | 12     | 2 0.087            | 4 0.174   | 2.1       |
| cyano              | 39     | 4 0.308            | 11 0.846  | 17.0      |
| cyklene           | 11     | 5 0.128            | 6 0.154   | 1.6       |
| halogenated compound | 105 | 11 0.105           | 34 0.324  | 4.8       |
| heterocyclic      | 57     | 25 0.439           | 33 0.579  | 20.3      |
| inorganic         | 3      | 0 0.000            | 0 0.000   | 0.0       |
| Class | N | Pearson r | Spearman ρ |
|-------|---|------------|-------------|
|       | Molecules | Matched | Fraction | Matched | Fraction | Power |
|       | Classes |         |          |         |          |       |
| ketone | 33 | 4 | 0.121 | 0.576 | 16.2 |
| nitro | 10 | 3 | 0.300 | 0.500 | 6.8 |
| phenol | 18 | 3 | 0.167 | 0.611 | 12.9 |
| thiol | 16 | 0 | 0.000 | 0.062 | 0.5 |
| all molecules | 670 | 131 | 0.196 | — | 228.2 |
| alcohol | 76 | 5 | 0.066 | 0.671 | 29.6 |
| aldehyde | 12 | 0 | 0.000 | 0.083 | 0.7 |
| alkane | 82 | 3 | 0.037 | 0.305 | 5.0 |
| alkene | 174 | 63 | 0.362 | 0.810 | 51.2 |
| alkythromide | 20 | 3 | 0.150 | 0.500 | 10.1 |
| alkylchloride | 41 | 1 | 0.024 | 0.195 | 2.5 |
| alkylthioamide | 8 | 1 | 0.125 | 0.250 | 2.4 |
| alkyn | 27 | 4 | 0.148 | 0.519 | 12.4 |
| amide | 10 | 0 | 0.000 | 0.300 | 3.4 |
| amine | 41 | 7 | 0.171 | 0.488 | 13.4 |
| aromatic | 172 | 52 | 0.302 | 0.721 | 36.3 |
| arylchloride | 23 | 1 | 0.043 | 0.087 | 0.7 |
| arylnitro | 13 | 0 | 0.000 | 0.308 | 4.1 |
| cycloalkane | 39 | 12 | 0.308 | 0.538 | 15.6 |
| cycloalkene | 11 | 7 | 0.636 | 0.636 | 10.0 |
| halogenated compound | 105 | 9 | 0.086 | 0.400 | 8.7 |
| heterocyclic | 57 | 25 | 0.439 | 0.561 | 19.2 |
| inorganic | 3 | 0 | 0.000 | 0.000 | 0.0 |
| ketone | 33 | 1 | 0.030 | 0.485 | 12.2 |
| nitro | 10 | 0 | 0.000 | 0.400 | 5.0 |
| phenol | 18 | 2 | 0.111 | 0.556 | 11.2 |
| thiol | 16 | 6 | 0.375 | 0.625 | 12.4 |
| all molecules | 670 | 139 | 0.207 | — | 246.0 |

Table S16: Correlation of theoretical spectra against database of experimental spectra in the range 550–2000 cm\(^{-1}\) for B3LYP/aug-cc-pVTZ sorted according to compound classification.

| Class | N | Pearson r | Spearman ρ |
|-------|---|------------|-------------|
|       | Molecules | Matched | Fraction | Matched | Fraction | Power |
|       | Classes |         |          |         |          |       |
| alcohol | 76 | 6 | 0.079 | 0.658 | 28.4 |
| aldehyde | 12 | 0 | 0.000 | 0.000 | 0.7 |
| alkane | 82 | 9 | 0.110 | 0.634 | 26.8 |
| alkene | 174 | 60 | 0.345 | 0.759 | 42.1 |
| alkythromide | 20 | 2 | 0.100 | 0.200 | 2.6 |
| alkylchloride | 41 | 2 | 0.049 | 0.244 | 3.8 |
| alkylthioamide | 8 | 1 | 0.125 | 0.375 | 4.0 |
| alkyn | 27 | 3 | 0.111 | 0.333 | 6.2 |

Table S17: Correlation of experimental spectra against database of theoretical spectra in the range 550–2000 cm\(^{-1}\) for B3LYP/aug-cc-pVTZ sorted according to compound classification.
Table S18: Correlation of theoretical spectra against database of experimental spectra in the range 550–2000 cm\(^{-1}\) for CGenFF sorted according to compound classification.

| Class                  | N   | Pearson r | Spearman ρ |
|------------------------|-----|-----------|------------|
|                        | Molecules | Classes | Molecules | Classes |
|                        | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| amide                  | 10    | 0.000    | 5       | 0.066    | 0.0 | 6      | 0.013 | 23      | 0.303 | 5.1  |
| amine                  | 41    | 0.146    | 24      | 0.585    | 18.4 | 0.659 | 36     | 0.878   | 37.9  |
| aromatic               | 172   | 0.320    | 134     | 0.779    | 45.7 | 0.733 | 162    | 0.942   | 81.3  |
| arylchloride           | 23    | 0.043    | 5       | 0.217    | 3.0 | 0.783 | 18     | 0.783   | 21.9  |
| arylfluoride           | 13    | 0.000    | 5       | 0.385    | 5.5 | 0.846 | 13     | 1.000   | 22.3  |
| cycloalkane            | 39    | 0.179    | 12      | 0.308    | 5.9 | 0.385 | 22     | 0.564   | 16.9  |
| cycloalkene            | 11    | 0.545    | 6       | 0.545    | 8.1 | 0.818 | 9      | 0.818   | 14.3  |
| halogenated compound   | 105   | 0.076    | 41      | 0.390    | 8.2 | 0.448 | 82     | 0.781   | 44.7  |
| heterocyclic           | 57    | 0.368    | 30      | 0.526    | 17.0 | 0.737 | 45     | 0.789   | 36.8  |
| inorganic              | 3     | 0.000    | 0       | 0.000    | 0.0 | 1.000 | 1      | 0.333   | 1.9   |
| ketone                 | 33    | 0.061    | 12      | 0.364    | 7.6 | 0.485 | 24     | 0.727   | 24.0  |
| nitro                  | 10    | 0.000    | 7       | 0.700    | 10.7 | 0.900 | 10     | 1.000   | 18.3  |
| phenol                 | 18    | 0.167    | 10      | 0.556    | 11.2 | 0.667 | 15     | 0.833   | 20.7  |
| thiol                  | 16    | 0.250    | 4       | 0.250    | 3.3 | 0.438 | 8      | 0.500   | 8.9   |
| all molecules          | 670   | 0.201    | —       | —       | 237.1 | 0.457 | —      | —      | 300.0 |

S26
Table S19: Correlation of experimental spectra against database of theoretical spectra in the range 550–2000 cm\(^{-1}\) for CGenFF sorted according to compound classification.

| Class               | N   | Molecules Matched | Pearson r | Molecules Matched | Spearman ρ |
|---------------------|-----|-------------------|-----------|-------------------|------------|
|                     |     | Fraction          |           | Fraction          |            |
|                     |     |                   | Matched   |                   | Power      |
| alcohol             | 76  | 0.000             | 13        | 0.171             | 1.1        |
| aldehyde            | 12  | 0.000             | 0         | 0.000             | 0.0        |
| alkane              | 82  | 0.012             | 4         | 0.049             | 0.0        |
| alkeno              | 174 | 0.011             | 72        | 0.414             | 5.2        |
| alkylbromide        | 20  | 0.000             | 0         | 0.000             | 0.0        |
| alkylchloride       | 41  | 0.000             | 2         | 0.049             | 0.1        |
| alkylfluoride       | 8   | 0.000             | 0         | 0.000             | 0.0        |
| alkyne              | 27  | 0.000             | 0         | 0.000             | 0.0        |
| amide               | 10  | 0.100             | 2         | 0.200             | 2.0        |
| amine               | 41  | 0.000             | 4         | 0.098             | 0.6        |
| aromatic            | 172 | 0.006             | 94        | 0.547             | 15.2       |
| aryalkylide         | 23  | 0.000             | 3         | 0.130             | 1.4        |
| aryalkylfluoride    | 13  | 0.000             | 0         | 0.000             | 0.0        |
| cycloalkane         | 39  | 0.026             | 2         | 0.051             | 0.2        |
| cycloalkene         | 11  | 0.000             | 0         | 0.000             | 0.0        |
| halogenated compound| 105 | 0.000             | 17        | 0.162             | 0.3        |
| heterocyclic        | 57  | 0.018             | 8         | 0.140             | 1.0        |
| inorganic           | 3   | 0.000             | 0         | 0.000             | 0.0        |
| ketone              | 33  | 0.030             | 10        | 0.303             | 5.6        |
| nitro               | 10  | 0.000             | 0         | 0.000             | 0.0        |
| phenol              | 18  | 0.056             | 5         | 0.278             | 4.0        |
| thiol               | 16  | 0.000             | 0         | 0.000             | 0.0        |
| all molecules       | 670 | 0.012             | —         | —                 | 5.0        |

Table S20: Correlation of theoretical spectra against database of experimental spectra in the range 550–2000 cm\(^{-1}\) for GAFF-BCC sorted according to compound classification.

| Class               | N   | Molecules Matched | Pearson r | Molecules Matched | Spearman ρ |
|---------------------|-----|-------------------|-----------|-------------------|------------|
|                     |     | Fraction          |           | Fraction          |            |
|                     |     |                   | Matched   |                   | Power      |
| alcohol             | 76  | 0.013             | 7         | 0.092             | 0.1        |
| aldehyde            | 12  | 0.083             | 1         | 0.083             | 0.7        |
| alkane              | 82  | 0.024             | 16        | 0.195             | 1.4        |
| alkeno              | 174 | 0.006             | 33        | 0.190             | 0.0        |
| alkylbromide        | 20  | 0.000             | 1         | 0.050             | 0.3        |
| alkylchloride       | 41  | 0.000             | 4         | 0.098             | 0.6        |
| alkylfluoride       | 8   | 0.125             | 1         | 0.125             | 1.0        |
| alkyne              | 27  | 0.000             | 1         | 0.037             | 0.2        |
| amide               | 10  | 0.000             | 0         | 0.000             | 0.0        |
| amine               | 41  | 0.000             | 1         | 0.024             | 0.0        |
| aromatic            | 172 | 0.006             | 65        | 0.378             | 3.5        |
| aryalkylide         | 23  | 0.000             | 0         | 0.000             | 0.0        |
| aryalkylfluoride    | 13  | 0.000             | 0         | 0.000             | 0.0        |
| cycloalkane         | 39  | 0.000             | 1         | 0.026             | 0.0        |
| cycloalkene         | 11  | 0.000             | 2         | 0.182             | 1.9        |
| halogenated compound| 105 | 0.010             | 24        | 0.229             | 1.5        |
| heterocyclic        | 57  | 0.018             | 13        | 0.228             | 3.1        |
| inorganic           | 3   | 0.000             | 0         | 0.000             | 0.0        |
| Class     | N | Pearson $r$ | Spearman $\rho$ |
|-----------|---|-------------|-----------------|
|           |   | Molecules Matched | Fractions Matched | Fractions | Power | Molecules Matched | Fractions Matched | Fractions | Power |
| ketone    | 33 | 0.000 | 4 | 0.121 | 1.1 | 1 | 0.030 | 2 | 0.063 | 0.3 |
| nitro     | 10 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| phenol    | 18 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| thiol     | 16 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| all molecules | 670 | 5 | 0.007 | — | — | 2.4 | 8 | 0.012 | — | 5.0 |

Table S21: Correlation of experimental spectra against database of theoretical spectra in the range 550–2000 cm$^{-1}$ for GAFF-BCC sorted according to compound classification.

| Class     | N | Pearson $r$ | Spearman $\rho$ |
|-----------|---|-------------|-----------------|
|           |   | Molecules Matched | Fractions Matched | Fractions | Power | Molecules Matched | Fractions Matched | Fractions | Power |
| alcohol   | 76 | 0.000 | 18 | 0.237 | 2.7 | 0 | 0.000 | 58 | 0.763 | 38.7 |
| aldehyde  | 12 | 0.000 | 1 | 0.083 | 0.7 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| alkane    | 82 | 2 | 0.024 | 40 | 0.488 | 15.2 | 0 | 0.000 | 25 | 0.305 | 5.0 |
| alkene    | 174 | 0 | 0.000 | 61 | 0.351 | 2.3 | 0 | 0.000 | 40 | 0.230 | 0.1 |
| alkylbromide | 20 | 0 | 0.000 | 10 | 0.500 | 10.1 | 0 | 0.000 | 9 | 0.450 | 8.6 |
| alkylchloride | 41 | 1 | 0.024 | 8 | 0.195 | 2.5 | 1 | 0.042 | 19 | 0.463 | 12.2 |
| alkylfluoride | 8 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 1 | 0.125 | 1.0 |
| alkyne    | 27 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| amide     | 10 | 0 | 0.000 | 2 | 0.200 | 2.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| amine     | 41 | 0 | 0.000 | 1 | 0.024 | 0.0 | 0 | 0.000 | 1 | 0.024 | 0.0 |
| aromatic  | 172 | 0 | 0.000 | 22 | 0.128 | 0.0 | 0 | 0.000 | 29 | 0.169 | 0.0 |
| arylchloride | 23 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| arylfluoride | 13 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| cycloalkane | 39 | 0 | 0.000 | 5 | 0.128 | 1.1 | 0 | 0.000 | 2 | 0.051 | 0.2 |
| cycloalkene | 11 | 0 | 0.000 | 4 | 0.364 | 4.7 | 0 | 0.000 | 1 | 0.091 | 0.8 |
| halogenated compound | 105 | 1 | 0.010 | 37 | 0.352 | 6.1 | 1 | 0.010 | 37 | 0.352 | 6.1 |
| heterocyclic | 57 | 1 | 0.018 | 9 | 0.158 | 1.3 | 0 | 0.000 | 4 | 0.070 | 0.1 |
| inorganic | 3 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| ketone    | 33 | 1 | 0.050 | 10 | 0.303 | 5.6 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| nitro     | 10 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| phenol    | 18 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 1 | 0.056 | 0.4 |
| thiol     | 16 | 0 | 0.000 | 0 | 0.000 | 0.0 | 0 | 0.000 | 0 | 0.000 | 0.0 |
| all molecules | 670 | 3 | 0.004 | — | — | 1.1 | 3 | 0.004 | — | 1.1 |

Table S22: Correlation of theoretical spectra against database of experimental spectra in the range 550–2000 cm$^{-1}$ for GAFF-ESP sorted according to compound classification.
| Class                | N   | Pearson $r$ Molecules |             |             | Spearman $\rho$ Molecules |             |             |
|---------------------|-----|-----------------------|-------------|-------------|---------------------------|-------------|-------------|
|                      |     | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| amide               | 10  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| amine               | 41  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| aromatic            | 172 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| arylchloride        | 23  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| arytfuoride         | 13  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| cycloalkane         | 39  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| cyctalkene          | 11  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| halogenated compound | 105 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| heterocyclic        | 57  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| inorganic           | 3   | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| ketone              | 33  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| nitro               | 10  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| phenol              | 18  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| thiol               | 16  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |
| all molecules       | 670 | 0.000  | —      | —      | —      | —    | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000 |

Table S23: Correlation of experimental spectra against database of theoretical spectra in the range 550–2000 cm$^{-1}$ for GAFF-ESP sorted according to compound classification.
Table S24: Correlation of theoretical spectra against database of experimental spectra in the range 550–2000 cm\(^{-1}\) for OPLS sorted according to compound classification.

| Class             | N  | Pearson \(r\) Molecules Fraction | Pearson \(r\) Classes Fraction | Spearman \(\rho\) Molecules Fraction | Spearman \(\rho\) Classes Fraction | Power |
|-------------------|----|---------------------------------|--------------------------------|-------------------------------------|-----------------------------------|-------|
|                   |    | Matched                        |                  | Matched                            |                                  |       |
| alcohol           | 76 | 0.000                          | 0.000            | 0.079                              | 0.1                               |       |
| aldehyde          | 12 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| alkanes           | 82 | 0.000                          | 0.232            | 0.144                              | 0.000                             | 0.085 |
| alkenes           | 174| 0.000                          | 25               | 0.000                              | 0.000                             | 0.155 |
| alkylbromides     | 20 | 0.000                          | 0.000            | 0.024                              | 0.000                             | 0.024 |
| alkylchlorides    | 41 | 0.000                          | 1.1              | 0.000                              | 0.000                             | 0.172 |
| alkylfluorides    | 8  | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| alkyne            | 27 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| amides            | 10 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| amines            | 41 | 0.000                          | 1.024            | 0.000                              | 0.000                             | 0.000 |
| aromatics         | 12 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| arylchlorides     | 23 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| arylfluorides     | 12 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| cycloalkanes      | 39 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| cycloalkenes      | 11 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| halogenated       | 23 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| heterocycles      | 41 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| inorganics        | 3  | 0.000                          | 0.333            | 0.333                              | 0.000                             | 0.000 |
| ketones           | 33 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| nitriles          | 10 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| phenols           | 18 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| thiols            | 16 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| all molecules     | 670| 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |

Table S25: Correlation of experimental spectra against database of theoretical spectra in the range 550–2000 cm\(^{-1}\) for OPLS sorted according to compound classification.

| Class             | N  | Pearson \(r\) Molecules Fraction | Pearson \(r\) Classes Fraction | Spearman \(\rho\) Molecules Fraction | Spearman \(\rho\) Classes Fraction | Power |
|-------------------|----|---------------------------------|--------------------------------|-------------------------------------|-----------------------------------|-------|
|                   |    | Matched                        |                  | Matched                            |                                  |       |
| alcohol           | 76 | 0.000                          | 0.000            | 0.079                              | 0.1                               |       |
| aldehyde          | 12 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| alkanes           | 82 | 0.000                          | 0.232            | 0.144                              | 0.000                             | 0.085 |
| alkenes           | 174| 0.000                          | 25               | 0.000                              | 0.000                             | 0.155 |
| alkylbromides     | 20 | 0.000                          | 0.000            | 0.024                              | 0.000                             | 0.024 |
| alkylchlorides    | 41 | 0.000                          | 1.1              | 0.000                              | 0.000                             | 0.172 |
| alkylfluorides    | 8  | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| alkyne            | 27 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| amides            | 10 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| amines            | 41 | 0.000                          | 1.024            | 0.000                              | 0.000                             | 0.000 |
| aromatics         | 12 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| arylchlorides     | 23 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| arylfluorides     | 12 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| cycloalkanes      | 39 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| cycloalkenes      | 11 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| halogenated       | 23 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| heterocycles      | 41 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| inorganics        | 3  | 0.000                          | 0.333            | 0.333                              | 0.000                             | 0.000 |
| ketones           | 33 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| nitriles          | 10 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| phenols           | 18 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| thiols            | 16 | 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| all molecules     | 670| 0.000                          | 0.000            | 0.000                              | 0.000                             | 0.000 |
| Class    | N  | Pearson r | Spearman ρ |
|----------|----|-----------|------------|
| Molecules | Matched | Fraction | Matched | Fraction | Power | Molecules | Matched | Fraction | Matched | Fraction | Power |
| ketone   | 33  | 0.000    | 0.000    | 0.000       | 0.000 | 0.000    | 0.000    | 2.061   | 0.300   |
| nitro    | 10  | 0.000    | 0.000    | 0.000       | 0.000 | 0.000    | 0.000    | 1.000   | 0.900   |
| phenol   | 18  | 0.000    | 0.000    | 0.000       | 0.000 | 0.000    | 0.000    | 0.000   | 0.000   |
| thiol    | 16  | 0.000    | 0.000    | 0.000       | 0.000 | 0.000    | 0.000    | 0.000   | 0.000   |
| all molecules | 670 | 2.000 | —       | —         | 0.600 | 7.010   | —       | 4.100   | —       |

Table S26: Correlation of theoretical spectra against database of experimental spectra in the range 550–1650 cm$^{-1}$ for B3LYP/6-31G(2df,p) sorted according to compound classification.

| Class         | N | Pearson r | Spearman ρ |
|---------------|---|-----------|------------|
| Molecules     | Matched | Fraction | Matched | Fraction | Power | Molecules | Matched | Fraction | Matched | Fraction | Power |
| alcohol       | 76 | 0.010    | 0.046    | 0.605    | 23.9  | 0.289    | 0.580    | 0.763   | 38.7   |
| aldehyde      | 12 | 0.250    | 0.250    | 3.00     |       | 0.417    | 0.541    | 0.417   | 5.9    |
| alkane        | 82 | 0.037    | 0.134    | 0.40     |       | 0.049    | 0.293    | 0.45    |
| aldehyde      | 174| 0.276    | 0.736    | 38.4     |       | 0.431    | 0.759    | 42.1    |
| alkylbromide  | 20 | 0.050    | 0.250    | 3.60     |       | 0.000    | 0.100    | 0.90    |
| alkylchloride | 41 | 0.049    | 0.317    | 6.20     |       | 0.171    | 0.390    | 9.00    |
| alkylfluoride | 8  | 0.000    | 0.000    | 0.00     |       | 0.259    | 0.556    | 13.9    |
| aldehyde      | 27 | 0.073    | 0.111    | 1.00     |       | 0.171    | 0.538    | 8.80    |
| amide         | 10 | 0.098    | 0.171    | 0.90     |       | 0.259    | 0.556    | 13.9    |
| aromatic      | 172| 0.250    | 0.529    | 13.60    |       | 0.529    | 0.797    | 48.8    |
| alkylchloride | 23 | 0.043    | 0.043    | 0.30     |       | 0.035    | 0.435    | 8.80    |
| alkylfluoride | 13 | 0.154    | 0.462    | 7.10     |       | 0.538    | 0.737    | 12.1    |
| alkene        | 39 | 0.179    | 0.308    | 5.90     |       | 0.308    | 0.436    | 10.8    |
| alkylfluoride | 11 | 0.455    | 0.455    | 6.30     |       | 0.455    | 0.874    | 64.2    |
| aldehyde      | 20 | 0.065    | 0.182    | 3.00     |       | 0.065    | 0.352    | 15.2    |
| aldehyde      | 57 | 0.544    | 0.544    | 18.10    |       | 0.737    | 12.2     | 32.2    |
| alcohol       | 3  | 0.000    | 0.000    | 0.00     |       | 0.000    | 1.000    | 1.90    |
| ketone        | 33 | 0.030    | 0.182    | 2.30     |       | 0.273    | 0.394    | 8.60    |
| nitro         | 10 | 0.100    | 0.100    | 0.90     |       | 0.600    | 1.000    | 18.3    |
| phenol        | 18 | 0.111    | 0.222    | 2.90     |       | 0.500    | 0.667    | 14.6    |
| thiol         | 16 | 0.000    | 0.062    | 0.50     |       | 0.312    | 0.562    | 10.60   |
| all molecules | 670| 0.170    | —       | —       | 191.2 | 0.339    | —       | 300.0   |

Table S27: Correlation of experimental spectra against database of theoretical spectra in the range 550–1650 cm$^{-1}$ for B3LYP/6-31G(2df,p) sorted according to compound classification.
| Class              | N  | Molecules | Pearson r | Spearman ρ |
|--------------------|----|-----------|-----------|------------|
|                    |    | Matched   | Fraction  | Matched   | Fraction  | Power     | Matched   | Fraction  | Matched   | Fraction  | Power     |
| amide              | 10 | 0         | 0.000     | 2         | 0.200     | 2.0       | 5         | 0.500     | 7         | 0.700     | 10.7      |
| amine              | 41 | 4         | 0.098     | 7         | 0.171     | 1.9       | 23        | 0.561     | 30        | 0.732     | 27.2      |
| aromatic           | 172| 57        | 0.331     | 133       | 0.773     | 44.7      | 114       | 0.663     | 165       | 0.959     | 86.4      |
| arylchloride       | 23 | 2         | 0.087     | 4         | 0.174     | 2.1       | 15        | 0.652     | 16        | 0.696     | 18.1      |
| arylylfluoride     | 13 | 6         | 0.462     | 12        | 0.923     | 19.4      | 13        | 1.000     | 13        | 1.000     | 22.3      |
| cycloalkane        | 39 | 3         | 0.077     | 4         | 0.103     | 0.7       | 5         | 0.500     | 7         | 0.636     | 10.0      |
| cycloalkene        | 11 | 4         | 0.364     | 4         | 0.364     | 4.7       | 7         | 0.636     | 7         | 0.636     | 10.0      |
| halogenated compound | 105 | 12        | 0.114     | 34        | 0.324     | 4.8       | 46        | 0.438     | 68        | 0.648     | 28.9      |
| heterocyclic       | 57 | 24        | 0.421     | 31        | 0.544     | 18.1      | 42        | 0.737     | 46        | 0.807     | 38.4      |
| inorganic          | 3  | 0         | 0.000     | 0         | 0.000     | 0.0       | 0         | 0.000     | 0         | 0.000     | 0.0       |
| ketone             | 33 | 1         | 0.030     | 5         | 0.152     | 1.7       | 12        | 0.364     | 14        | 0.424     | 9.8       |
| nitro              | 10 | 3         | 0.300     | 9         | 0.500     | 6.8       | 9         | 0.900     | 10        | 1.000     | 18.3      |
| phenol             | 18 | 5         | 0.278     | 10        | 0.556     | 11.2      | 9         | 0.500     | 14        | 0.778     | 18.5      |
| thiol              | 16 | 0         | 0.000     | 1         | 0.062     | 0.5       | 4         | 0.250     | 4         | 0.250     | 3.3       |
| all molecules      | 670| 125       | 0.187     | —         | —         | 215.0     | 278       | 0.415     | —         | —         | 300.0     |

Table S28: Correlation of theoretical spectra against database of experimental spectra in the range 550–1650 cm⁻¹ for B3LYP/aug-cc-pVTZ sorted according to compound classification.
Table S29: Correlation of experimental spectra against database of theoretical spectra in the range 550–1650 cm\(^{-1}\) for B3LYP/aug-cc-pVTZ sorted according to compound classification.

| Class            | N  | Molecules |  | Classes |  | Power |
|------------------|----|-----------|---|---------|---|-------|
|                  |    | Matched   | Fraction | Matched | Fraction |       |
|                  |    | Pearson r | Spearman ρ |       |       |       |
| alcohol          | 76 | 0.066     | 0.632    | 26.1    | 0.711    | 33.3  |
| aldehyde         | 12 | 0.333     | 0.333    | 4.3     | 0.750    | 15.4  |
| alkane           | 82 | 0.122     | 0.634    | 26.8    | 0.585    | 22.6  |
| alkene           | 174| 0.351     | 0.753    | 41.2    | 0.920    | 75.3  |
| alkythromide     | 20 | 0.100     | 0.200    | 2.6     | 0.450    | 8.6   |
| alkylchloride    | 41 | 0.024     | 0.220    | 3.2     | 0.585    | 22.6  |
| alkylfluoride    | 8  | 0.125     | 0.125    | 1.0     | 0.500    | 5.9   |
| alkyne           | 27 | 0.111     | 0.370    | 7.3     | 0.630    | 17.0  |
| amide            | 10 | 0.300     | 0.400    | 5.0     | 0.400    | 5.0   |
| amine            | 41 | 0.146     | 0.463    | 12.2    | 0.902    | 40.0  |
| aromatic         | 172| 0.355     | 0.756    | 41.8    | 0.948    | 82.9  |
| arylchloride     | 23 | 0.130     | 0.304    | 5.1     | 0.826    | 23.9  |
| arylfluoride     | 13 | 0.000     | 0.385    | 5.5     | 1.000    | 22.3  |
| cycloalkane      | 39 | 0.231     | 0.359    | 7.7     | 0.641    | 21.0  |
| cycloalkene      | 11 | 0.636     | 0.636    | 10.0    | 0.818    | 14.3  |
| halogenated compound | 105| 0.067     | 0.390    | 8.2     | 0.790    | 46.0  |
| heterocyclic     | 57 | 0.421     | 0.561    | 19.2    | 0.789    | 36.8  |
| inorganic        | 3  | 0.000     | 0.000    | 0.0     | 0.667    | 4.2   |
| ketone           | 33 | 0.091     | 0.364    | 7.6     | 0.636    | 19.2  |
| nitro            | 10 | 0.000     | 0.900    | 15.4    | 1.000    | 18.3  |
| phenol           | 18 | 0.167     | 0.389    | 6.6     | 0.778    | 18.5  |
| thiol            | 16 | 0.312     | 0.312    | 4.6     | 0.625    | 12.4  |
| all molecules    | 670| 0.221     | —        | —       | 266.4    |       |

Table S30: Correlation of theoretical spectra against database of experimental spectra in the range 550–1650 cm\(^{-1}\) for CGenFF sorted according to compound classification.

| Class            | N  | Molecules |  | Classes |  | Power |
|------------------|----|-----------|---|---------|---|-------|
|                  |    | Matched   | Fraction | Matched | Fraction |       |
|                  |    | Pearson r | Spearman ρ |       |       |       |
| alcohol          | 76 | 0.000     | 0.013    | 0.0     | 0.250    | 3.2   |
| aldehyde         | 12 | 0.000     | 0.000    | 0.0     | 0.000    | 0.0   |
| alkane           | 82 | 0.000     | 0.037    | 0.0     | 0.073    | 0.0   |
| alkene           | 174| 0.006     | 0.247    | 0.2     | 0.494    | 10.5  |
| alkythromide     | 20 | 0.000     | 0.000    | 0.0     | 0.050    | 0.3   |
| alkylchloride    | 41 | 0.000     | 0.049    | 0.1     | 0.049    | 0.1   |
| alkylfluoride    | 8  | 0.000     | 0.000    | 0.0     | 0.125    | 1.0   |
| alkyne           | 27 | 0.000     | 0.100    | 0.9     | 0.200    | 2.0   |
| amide            | 10 | 0.000     | 0.100    | 0.9     | 0.200    | 2.0   |
| amine            | 41 | 0.000     | 0.098    | 0.6     | 0.103    | 0.7   |
| aromatic         | 172| 0.006     | 0.488    | 10.2    | 0.343    | 2.1   |
| arylchloride     | 23 | 0.000     | 0.000    | 0.0     | 0.000    | 0.0   |
| arylfluoride     | 13 | 0.000     | 0.000    | 0.0     | 0.000    | 0.0   |
| cycloalkane      | 39 | 0.000     | 0.026    | 0.0     | 0.091    | 0.8   |
| cycloalkene      | 11 | 0.000     | 0.000    | 0.0     | 0.086    | 0.0   |
| halogenated compound | 105| 0.000     | 0.114    | 0.0     | 0.158    | 1.3   |
| heterocyclic     | 57 | 0.018     | 0.088    | 0.3     | 0.000    | 0.0   |
| inorganic        | 3  | 0.000     | 0.000    | 0.0     | 0.000    | 0.0   |

S33
Table S31: Correlation of experimental spectra against database of theoretical spectra in the range 550–1650 cm\(^{-1}\) for CGenFF sorted according to compound classification.

| Class          | N  | Pearson \(r\) | Spearman \(\rho\) |
|----------------|----|----------------|-------------------|
|                | Molecules | Classes | Molecules | Classes |
|                | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| ketone         | 33    | 0        | 0.000   | 3       | 0.091 | 0.7    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| nitro          | 10    | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| phenol         | 18    | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| thiol          | 16    | 0        | 0.000   | 0       | 0.000 | 0.0    | 1        | 0.062   | 6       | 0.375 | 5.9   |
| all molecules  | 670   | 1        | 0.001   | —       | —     | 0.2    | 15       | 0.022   | —       | —     | 12.6  |

Table S32: Correlation of theoretical spectra against database of experimental spectra in the range 550–1650 cm\(^{-1}\) for GAFF-BCC sorted according to compound classification.

| Class          | N  | Pearson \(r\) | Spearman \(\rho\) |
|----------------|----|----------------|-------------------|
|                | Molecules | Classes | Molecules | Classes |
|                | Matched | Fraction | Matched | Fraction | Power | Matched | Fraction | Matched | Fraction | Power |
| alcohol        | 76    | 0        | 0.000   | 12      | 0.158 | 0.8    | 3        | 0.039   | 32      | 0.421 | 11.0  |
| aldehyde       | 12    | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| alkane         | 82    | 2        | 0.012   | 4       | 0.049 | 0.0    | 1        | 0.012   | 13      | 0.159 | 0.7   |
| alkene         | 174   | 2        | 0.011   | 85      | 0.489 | 10.1   | 3        | 0.017   | 65      | 0.374 | 3.2   |
| alkylbromide   | 20    | 0        | 0.000   | 0       | 0.000 | 0.0    | 3        | 0.150   | 3       | 0.150 | 1.7   |
| alkylchloride  | 41    | 0        | 0.000   | 5       | 0.122 | 1.0    | 2        | 0.049   | 5       | 0.122 | 1.0   |
| alkylfluoride  | 8     | 0        | 0.000   | 0       | 0.000 | 0.0    | 1        | 0.125   | 1       | 0.125 | 1.0   |
| alkyne         | 27    | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 1       | 0.037 | 0.2   |
| amide          | 10    | 0        | 0.000   | 0       | 0.000 | 0.0    | 1        | 0.100   | 2       | 0.200 | 2.0   |
| amine          | 41    | 0        | 0.000   | 12      | 0.293 | 5.4    | 2        | 0.049   | 14      | 0.341 | 7.1   |
| aromatic       | 172   | 0        | 0.000   | 96      | 0.558 | 16.3   | 3        | 0.017   | 85      | 0.494 | 10.7  |
| arylchloride   | 23    | 0        | 0.000   | 1       | 0.043 | 0.3    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| arylfluoride   | 13    | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| cycloalkene    | 57    | 0        | 0.000   | 7       | 0.132 | 0.7    | 0        | 0.000   | 3       | 0.053 | 0.1   |
| halogenated compound | 105 | 0     | 0.000   | 19      | 0.181 | 0.5    | 4        | 0.038   | 23      | 0.219 | 1.2   |
| heterocyclic   | 57    | 0        | 0.000   | 7       | 0.123 | 0.7    | 0        | 0.000   | 3       | 0.053 | 0.1   |
| inorganic      | 3     | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| ketone         | 33    | 0        | 0.000   | 2       | 0.061 | 0.3    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| nitro          | 10    | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 0       | 0.000 | 0.0   |
| phenol         | 18    | 0        | 0.000   | 4       | 0.222 | 2.9    | 2        | 0.111   | 4       | 0.222 | 2.9   |
| thiol          | 16    | 0        | 0.000   | 0       | 0.000 | 0.0    | 0        | 0.000   | 1       | 0.062 | 0.5   |
| all molecules  | 670   | 4        | 0.006   | —       | —     | 1.7    | 17       | 0.025   | —       | —     | 15.0  |

S34
| Class               | N   | Pearson r |          | Spearman ρ |          |
|---------------------|-----|-----------|----------|------------|----------|
|                     |     | Molecules | Classes  | Molecules  | Classes  |
|                     |     | Matched   | Fraction | Matched    | Fraction |
|                     |     |           |          |           | Power    |
|                     |     |           |          |           |          |
| amide               | 10  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| amine               | 41  | 0         | 0.000    | 1         | 0.024    | 0.0     |
| aromatic            | 172 | 1         | 0.006    | 71        | 0.413    | 5.2     |
| arylchloride        | 23  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| aryfluoride         | 13  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| cycloalkane         | 39  | 0         | 0.000    | 1         | 0.026    | 0.0     |
| cycloalkene         | 11  | 0         | 0.000    | 2         | 0.182    | 1.9     |
| halogenated compound| 105 | 0         | 0.000    | 1         | 0.181    | 0.5     |
| heterocyclic        | 57  | 1         | 0.018    | 11        | 0.193    | 2.1     |
| inorganic           | 3   | 0         | 0.000    | 0         | 0.000    | 0.0     |
| ketone              | 33  | 0         | 0.000    | 2         | 0.061    | 0.3     |
| nitro               | 10  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| phenol              | 18  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| thiol               | 16  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| all molecules       | 670 | 3         | 0.004    | —         | —        | 1.1     |

Table S33: Correlation of experimental spectra against database of theoretical spectra in the range 550–1650 cm\(^{-1}\) for GAFF-BCC sorted according to compound classification.

| Class               | N   | Pearson r |          | Spearman ρ |          |
|---------------------|-----|-----------|----------|------------|----------|
|                     |     | Molecules | Classes  | Molecules  | Classes  |
|                     |     | Matched   | Fraction | Matched    | Fraction |
|                     |     |           |          |           | Power    |
| alcohol             | 76  | 0         | 0.000    | 18        | 0.237    | 2.7     |
| aldehyde            | 12  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| alkane              | 82  | 2         | 0.024    | 39        | 0.476    | 14.4    |
| alkenes             | 174 | 1         | 0.006    | 51        | 0.293    | 0.7     |
| alkylthiomide       | 20  | 0         | 0.000    | 10        | 0.500    | 10.1    |
| alkylchloride       | 41  | 1         | 0.024    | 9         | 0.220    | 3.2     |
| alkylchloride       | 8   | 0         | 0.000    | 0         | 0.000    | 0.0     |
| alkynes             | 27  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| amide               | 10  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| amine               | 41  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| aromatic            | 172 | 0         | 0.000    | 45        | 0.262    | 0.3     |
| arylchloride        | 23  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| aryfluoride         | 13  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| cycloalkane         | 39  | 0         | 0.000    | 3         | 0.077    | 0.4     |
| cycloalkene         | 11  | 0         | 0.000    | 3         | 0.273    | 3.2     |
| halogenated compound| 105 | 1         | 0.010    | 35        | 0.333    | 5.2     |
| heterocyclic        | 57  | 0         | 0.000    | 3         | 0.053    | 0.1     |
| inorganic           | 3   | 0         | 0.000    | 0         | 0.000    | 0.0     |
| ketone              | 33  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| nitro               | 10  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| phenol              | 18  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| thiol               | 16  | 0         | 0.000    | 0         | 0.000    | 0.0     |
| all molecules       | 670 | 3         | 0.004    | —         | —        | 1.1     |

S35
Table S34: Correlation of theoretical spectra against database of experimental spectra in the range 550–1650 cm\(^{-1}\) for GAFF-ESP sorted according to compound classification.

| Class         | N  | Molecules | Pearson \(r\) | Spearman \(\rho\) |
|---------------|----|-----------|---------------|------------------|
|               |    | Matched   | Fraction      | Matched Fraction | Matched Fraction |
| alcohol       | 76 | 0         | 0.000         | 3                | 0.039            |
|               |    | 0         | 0.000         | 0                | 0.000            |
| aldehyde      | 12 | 0         | 0.000         | 0                | 0.000            |
|               |    | 1         | 0.000         | 12               | 0.146            |
| alkane        | 82 | 0         | 0.000         | 19               | 0.109            |
|               |    | 1         | 0.000         | 0                | 0.000            |
| alkene        | 174| 0        | 0.000         | 0                | 0.000            |
| alkylthioamide| 20 | 0         | 0.000         | 0                | 0.000            |
| alkylchloride | 41 | 0         | 0.000         | 2                | 0.049            |
| alkylfluoride | 8  | 0         | 0.000         | 0                | 0.000            |
| alkyne        | 27 | 0         | 0.000         | 0                | 0.000            |
| amide         | 10 | 0         | 0.000         | 0                | 0.000            |
| amine         | 41 | 0         | 0.000         | 0                | 0.000            |
| aromatic      | 172| 0        | 0.000         | 67               | 0.390            |
| arylchloride  | 23 | 0         | 0.000         | 0                | 0.000            |
| arylfluoride  | 13 | 0         | 0.000         | 0                | 0.000            |
| cycloalkane   | 39 | 0         | 0.000         | 0                | 0.000            |
| cycloalkene   | 11 | 0         | 0.000         | 1                | 0.091            |
| halogenated compound | 105| 0 | 0.000 | 13 | 0.124 |
| heterocyclic  | 57 | 0         | 0.000         | 9                | 0.158            |
| inorganic     | 3  | 0         | 0.000         | 0                | 0.000            |
| ketone        | 33 | 0         | 0.000         | 1                | 0.030            |
| nitro         | 10 | 0         | 0.000         | 1                | 0.100            |
| phenol        | 18 | 0         | 0.000         | 0                | 0.000            |
| thiol         | 16 | 0         | 0.000         | 0                | 0.000            |
| all molecules | 670| 1        | 0.001         | —                | 0.2              |

Table S35: Correlation of experimental spectra against database of theoretical spectra in the range 550–1650 cm\(^{-1}\) for GAFF-ESP sorted according to compound classification.

| Class         | N  | Molecules | Pearson \(r\) | Spearman \(\rho\) |
|---------------|----|-----------|---------------|------------------|
|               |    | Matched   | Fraction      | Matched Fraction | Matched Fraction |
| alcohol       | 76 | 0         | 0.000         | 21               | 0.276            |
|               |    | 0         | 0.000         | 0                | 0.000            |
| aldehyde      | 12 | 0         | 0.000         | 0                | 0.000            |
|               |    | 1         | 0.000         | 9                | 0.110            |
| alkane        | 82 | 0         | 0.000         | 20               | 0.115            |
| alkene        | 174| 0        | 0.000         | 6                | 0.300            |
| alkylthioamide| 20 | 0         | 0.000         | 8                | 0.195            |
| alkylchloride | 41 | 0         | 0.000         | 0                | 0.000            |
| alkylfluoride | 8  | 0         | 0.000         | 0                | 0.000            |
| alkyne        | 27 | 0         | 0.000         | 0                | 0.000            |
| amide         | 10 | 0         | 0.000         | 0                | 0.000            |
| amine         | 41 | 0         | 0.000         | 0                | 0.000            |
| aromatic      | 172| 0        | 0.000         | 59               | 0.343            |
| arylchloride  | 23 | 0         | 0.000         | 0                | 0.000            |
| arylfluoride  | 13 | 0         | 0.000         | 0                | 0.000            |
| cycloalkane   | 39 | 0         | 0.000         | 1                | 0.026            |
| cycloalkene   | 11 | 0         | 0.000         | 2                | 0.182            |
| halogenated compound | 105| 0 | 0.000 | 33 | 0.314 |
| heterocyclic  | 57 | 0         | 0.000         | 7                | 0.123            |
| inorganic     | 3  | 0         | 0.000         | 0                | 0.000            |
### Table S36: Correlation of theoretical spectra against database of experimental spectra in the range 550–1650 cm\(^{-1}\) for OPLS sorted according to compound classification.

| Class      | N  | Pearson \(r\) | Molecules | Spearman \(\rho\) | Classes | Molecules | Classes |
|------------|----|---------------|-----------|-------------------|---------|-----------|---------|
|            |    |               | Matched   | Fraction          | Matched | Fraction  | Power   |
| ketone     | 33 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| nitro      | 10 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| phenol     | 18 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| thiol      | 16 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| all molecules | 670 | 0               | 0         | 0.000             | 0       | 0.000     | 0.0     |
| alcohol    | 76 | 0             | 5         | 0.092             | 0       | 0.000     | 0.2     |
| aldehyde   | 12 | 0             | 2         | 0.167             | 0       | 0.000     | 0.0     |
| alkane     | 82 | 0             | 17        | 0.207             | 0       | 0.000     | 1.7     |
| alkene     | 174| 0             | 27        | 0.155             | 0       | 0.000     | 0.0     |
| alkylthromide | 20 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| alkylchloride | 41  | 0         | 1          | 0.024             | 0       | 0.000     | 0.9     |
| alkylthiouride | 8   | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| alkyne     | 27 | 0             | 3         | 0.111             | 0       | 0.000     | 0.5     |
| amide      | 10 | 0             | 1         | 0.100             | 0       | 0.000     | 0.0     |
| amine      | 41 | 0             | 1         | 0.024             | 0       | 0.000     | 0.0     |
| aromatic   | 172| 0             | 41        | 0.238             | 0       | 0.000     | 2.5     |
| arylchloride | 23 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| arylthiouride | 13 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| cycloalkane | 39 | 0             | 1         | 0.026             | 0       | 0.000     | 0.0     |
| cycloalkene | 11 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| halogenated compound | 105 | 0       | 9          | 0.086             | 0       | 0.000     | 0.0     |
| heterocyclic | 57 | 0             | 9         | 0.158             | 1       | 0.018     | 1.7     |
| inorganic  | 3  | 1             | 0.333     | 0.333             | 0       | 0.000     | 0.0     |
| ketone     | 33 | 0             | 1         | 0.030             | 0       | 0.000     | 0.0     |
| nitro      | 10 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| phenol     | 18 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| thiol      | 16 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| all molecules | 670 | 1            | —         | —                 | 0       | 0.006     | 1.7     |

### Table S37: Correlation of experimental spectra against database of theoretical spectra in the range 550–1650 cm\(^{-1}\) for OPLS sorted according to compound classification.

| Class      | N  | Pearson \(r\) | Molecules | Spearman \(\rho\) | Classes | Molecules | Classes |
|------------|----|---------------|-----------|-------------------|---------|-----------|---------|
|            |    |               | Matched   | Fraction          | Matched | Fraction  | Power   |
| alcohol    | 76 | 0             | 5         | 0.066             | 0       | 0.000     | 0.0     |
| aldehyde   | 12 | 0             | 2         | 0.167             | 0       | 0.000     | 1.7     |
| alkane     | 82 | 0             | 36        | 0.439             | 12.0    | 0         | 0.0     |
| alkene     | 174| 1             | 26        | 0.149             | 0       | 0.000     | 2.4     |
| alkylthromide | 20 | 0             | 0         | 0.000             | 0       | 0.000     | 0.0     |
| alkylchloride | 41  | 0           | 1         | 0.024             | 0       | 0.000     | 0.1     |
| alkylthiouride | 8   | 0             | 0         | 0.000             | 0       | 0.000     | 1.0     |
| alkyne     | 27 | 0             | 1         | 0.037             | 0.2     | 0         | 0.5     |

S37
Class | N  | Molecules | Classes | Spearman ρ | Molecules | Classes |
|------|-----|-----------|---------|-----------|-----------|---------|
|      |     | Matched   | Fraction | Matched   | Fraction | Power   | Matched   | Fraction | Matched   | Fraction | Power   |
| amide| 10  | 0         | 0.000    | 0         | 0.000    | 0.000   | 1         | 0.100    | 1         | 0.100    | 0.9     |
| amine| 41  | 0         | 0.000    | 8         | 0.195    | 2.5     | 1         | 0.24     | 9         | 0.22     | 3.2     |
| aromatic| 172 | 0         | 0.000    | 64        | 0.372    | 3.3     | 0         | 0.000    | 57        | 0.33     | 1.8     |
| arylchloride| 23  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 0         | 0.000    | 0.0     |
| arylfluoride| 13  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 0         | 0.000    | 0.0     |
| cycloalkane| 39  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 0         | 0.000    | 0.0     |
| cycloalkene| 11  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 1         | 0.09     | 0.8     |
| halogenated compound| 105 | 0         | 0.000    | 23        | 0.219    | 1.2     | 0         | 0.000    | 18        | 0.17     | 0.4     |
| heterocyclic| 57  | 0         | 0.000    | 5         | 0.088    | 0.3     | 1         | 0.01     | 8         | 0.14     | 1.0     |
| inorganic| 3   | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 0         | 0.000    | 0.0     |
| ketone| 33  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 0         | 0.000    | 0.0     |
| nitro| 10  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 2         | 0.20     | 2.0     |
| phenol| 18  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 0         | 0.000    | 0.0     |
| thiol| 16  | 0         | 0.000    | 0         | 0.000    | 0.0     | 0         | 0.000    | 0         | 0.000    | 0.0     |
| all molecules| 670 | 1         | 0.001    | —         | —        | 0.2     | 2         | 0.003    | —         | —        | 0.6     |

Table S38: Correlation of theoretical spectra against database of experimental spectra in the range 2000–3846 cm⁻¹ for B3LYP/6-31G(2df,p) sorted according to compound classification.
Table S39: Correlation of experimental spectra against database of theoretical spectra in the range 2000–3846 cm$^{-1}$ for B3LYP/6-31G(2df,p) sorted according to compound classification.

| Class          | N   | Pearson $r$ | Spearman $\rho$ |
|----------------|-----|-------------|-----------------|
|                | Molecules | Clases | Matched | Fraction | Matched | Fraction | Power |
| alcohol        | 76  | 1 | 0.013 | 31 | 0.408 | 10.3 |
| aldehyde       | 12  | 0 | 0.000 | 3  | 0.250 | 3.0  |
| alkane         | 82  | 0 | 0.000 | 15 | 0.183 | 1.1  |
| alkenes        | 14  | 6 | 0.034 | 54 | 0.310 | 1.1  |
| alkylthiocyanide | 20  | 0 | 0.000 | 3  | 0.150 | 1.7  |
| alkylhalides   | 41  | 0 | 0.000 | 12 | 0.293 | 5.4  |
| alkyne         | 8   | 0 | 0.000 | 0  | 0.000 | 0.0  |
| amide          | 10  | 0 | 0.000 | 0  | 0.000 | 0.0  |
| amine          | 41  | 3 | 0.073 | 10 | 0.244 | 3.8  |
| aromatic       | 172 | 9 | 0.052 | 114| 0.663 | 28.2 |
| arylchloride   | 23  | 2 | 0.087 | 2  | 0.087 | 0.7  |
| arylfluoride   | 13  | 0 | 0.000 | 1  | 0.077 | 0.6  |
| cycloalkane    | 39  | 0 | 0.000 | 28 | 0.718 | 25.6 |
| cycloalkene    | 11  | 0 | 0.000 | 1  | 0.091 | 0.8  |
| halogenated compound | 105 | 2 | 0.019 | 37 | 0.352 | 6.1  |
| heterocyclic   | 57  | 11 | 0.193 | 20 | 0.351 | 7.7  |
| inorganic      | 3   | 1 | 0.333 | 1  | 0.333 | 1.9  |
| ketone         | 33  | 1 | 0.030 | 2  | 0.061 | 0.3  |
| nitro          | 10  | 0 | 0.000 | 0  | 0.000 | 0.0  |
| phenol         | 18  | 0 | 0.000 | 2  | 0.111 | 1.1  |
| thiol          | 16  | 0 | 0.000 | 1  | 0.062 | 0.5  |
| all molecules  | 670 | 15 | 0.022 | —  | —     | 12.6 |

Table S40: Correlation of theoretical spectra against database of experimental spectra in the range 2000–3846 cm$^{-1}$ for B3LYP/aug-cc-pVTZ sorted according to compound classification.

| Class          | N   | Pearson $r$ | Spearman $\rho$ |
|----------------|-----|-------------|-----------------|
|                | Molecules | Clases | Matched | Fraction | Matched | Fraction | Power |
| alcohol        | 76  | 1 | 0.013 | 31 | 0.408 | 10.3 |
| aldehyde       | 12  | 0 | 0.000 | 3  | 0.250 | 3.0  |
| alkane         | 82  | 0 | 0.000 | 15 | 0.183 | 1.1  |
| alkenes        | 174 | 9 | 0.052 | 70 | 0.402 | 4.5  |
| alkylthiocyanide | 20  | 0 | 0.000 | 1  | 0.050 | 0.3  |
| alkylhalides   | 41  | 0 | 0.000 | 9  | 0.220 | 3.2  |
| alkyne         | 27  | 2 | 0.074 | 5  | 0.185 | 2.4  |
| amide          | 10  | 1 | 0.100 | 2  | 0.200 | 2.0  |
| amine          | 41  | 0 | 0.000 | 7  | 0.171 | 1.9  |
| aromatic       | 172 | 8 | 0.047 | 93 | 0.541 | 14.6 |
| arylchloride   | 23  | 0 | 0.000 | 0  | 0.000 | 0.0  |
| arylfluoride   | 13  | 0 | 0.000 | 1  | 0.077 | 0.6  |
| cycloalkane    | 39  | 0 | 0.000 | 5  | 0.128 | 1.1  |
| cycloalkene    | 11  | 0 | 0.000 | 0  | 0.000 | 0.0  |
| halogenated compound | 105 | 2 | 0.019 | 47 | 0.448 | 11.8 |
| heterocyclic   | 57  | 9 | 0.158 | 21 | 0.368 | 8.5  |
| inorganic      | 3   | 0 | 0.000 | 0  | 0.000 | 0.0  |
| Class          | N | Pearson $r$ | Spearman $\rho$ |
|---------------|---|-------------|-----------------|
|               |   | Molecules   | Clases          | Molecules   | Clases          |                      |
|               |   | Matched     | Fraction        | Matched     | Fraction        |                      |
|               |   | 0           | 0.000           | 1           | 0.030           | 0.1                 |
| ketone        | 33 | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| nitro         | 10 | 0           | 0.000           | 2           | 0.111           | 1.1                 |
| phenol        | 18 | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| thiol         | 16 | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| all molecules | 670| 21          | 0.031           | —           | —               | 20.2                |
|               |   | 0           | 0.000           | 0           | 0.000           | 0.0                 |

Table S41: Correlation of experimental spectra against database of theoretical spectra in the range 2000–3846 cm$^{-1}$ for B3LYP/aug-cc-pVTZ sorted according to compound classification.

| Class          | N | Pearson $r$ | Spearman $\rho$ |
|---------------|---|-------------|-----------------|
|               |   | Molecules   | Clases          | Molecules   | Clases          |                      |
|               |   | Matched     | Fraction        | Matched     | Fraction        |                      |
| alcohol       | 76 | 1           | 0.013           | 37          | 0.487           | 15.1                |
| aldehyde      | 12 | 1           | 0.083           | 3           | 0.250           | 3.0                 |
| alkane        | 82 | 1           | 0.012           | 17          | 0.207           | 1.7                 |
| alkene        | 174| 10          | 0.057           | 68          | 0.391           | 4.0                 |
| aldehyde      | 20 | 0           | 0.000           | 1           | 0.050           | 0.3                 |
| alkylchloride | 41 | 0           | 0.000           | 0           | 0.244           | 3.8                 |
| aldehy        | 8  | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| amide         | 10 | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| amine         | 41 | 0           | 0.000           | 11          | 0.268           | 4.6                 |
| aromatic      | 172| 9           | 0.052           | 113         | 0.657           | 27.4                |
| arylchloride  | 23 | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| arylfluoride  | 13 | 1           | 0.077           | 1           | 0.077           | 0.6                 |
| cycloalkane   | 39 | 0           | 0.000           | 28          | 0.718           | 25.6                |
| cycloalkene   | 11 | 2           | 0.182           | 3           | 0.273           | 3.2                 |
| halogenated compound | 105| 1           | 0.010           | 32          | 0.305           | 4.0                 |
| heterocyclic  | 57 | 8           | 0.140           | 19          | 0.333           | 6.9                 |
| inorganic     | 3  | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| ketone        | 33 | 2           | 0.061           | 2           | 0.061           | 0.3                 |
| nitro         | 10 | 0           | 0.000           | 0           | 0.000           | 0.0                 |
| phenol        | 18 | 1           | 0.056           | 2           | 0.111           | 1.1                 |
| thiol         | 16 | 0           | 0.000           | 1           | 0.062           | 0.5                 |
| all molecules | 670| 21          | 0.031           | —           | —               | 20.2                |

Table S42: Correlation of theoretical spectra against database of experimental spectra in the range 2000–3846 cm$^{-1}$ for CGenFF sorted according to compound classification.
| Class                  | N   | Molecules Matched | Fraction | Classes Matched | Fraction | Power |
|------------------------|-----|-------------------|----------|-----------------|----------|-------|
| amide                  | 10  | 0                 | 0        | 0               | 0        | 0.0   |
| amine                  | 41  | 0                 | 0        | 6               | 0.146    | 1.4   |
| aromatic               | 172 | 4                 | 0.023    | 126             | 0.733    | 38.1  |
| arylchloride           | 23  | 0                 | 0        | 2               | 0.087    | 0.7   |
| arylfluoride           | 13  | 0                 | 0        | 0               | 0.000    | 0.0   |
| cycloalkane            | 39  | 1                 | 0.026    | 11              | 0.282    | 5.0   |
| cycloalkene            | 11  | 0                 | 0        | 2               | 0.182    | 1.9   |
| halogenated compound   | 105 | 1                 | 0.010    | 17              | 0.162    | 0.3   |
| heterocyclic           | 57  | 3                 | 0.053    | 7               | 0.123    | 0.7   |
| inorganic              | 3   | 0                 | 0        | 0               | 0.000    | 0.0   |
| ketone                 | 33  | 0                 | 0        | 0               | 0.000    | 0.0   |
| nitro                  | 10  | 0                 | 0        | 0               | 0.000    | 0.0   |
| phenol                 | 18  | 0                 | 0        | 0               | 0.000    | 0.0   |
| thiol                  | 16  | 0                 | 0        | 0               | 0.000    | 0.0   |
| all molecules          | 670 | 5                 | 0.021    | —               | —        | 11.4  |

Table S43: Correlation of experimental spectra against database of theoretical spectra in the range 2000–3846 cm⁻¹ for CGenFF sorted according to compound classification.
Table S44: Correlation of theoretical spectra against database of experimental spectra in the range 2000–3846 cm\(^{-1}\) for GAFF-BCC sorted according to compound classification.

| Class          | N | Pearson \(r\) Matched Fraction       | Pearson \(r\) Clases Matched Fraction | Spearman \(\rho\) Matched Fraction | Spearman \(\rho\) Clases Matched Fraction |
|----------------|---|-------------------------------------|---------------------------------------|-----------------------------------|------------------------------------------|
| alcohol        | 76| 0 0.000 0 0.000                     | 0 0.000 3 0.039                      | 0 0.000 3 0.039                   | 0.0                                      |
| aldehyde       | 12| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| alkane         | 82| 1 0.012 14 0.171                    | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 3.9                                      |
| alkene         | 174| 0 0.000 31 0.178                    | 1 0.006 143 0.822                    | 53.4                              | 53.4                                     |
| alkylthromide  | 20| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| alkylchloride  | 41| 0 0.000 5 0.122                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| alkylfluoride  | 8 | 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| alkyne         | 27| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| amide          | 10| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| amine          | 41| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| aromatic       | 172| 0 0.000 23 0.134                    | 0 0.000 22 0.128                     | 0.0                               | 0.0                                      |
| arylchloride   | 23| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| arylfluoride   | 13| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| cycloalkane    | 39| 0 0.000 9 0.211                    | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| cycloalkene    | 11| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| halogenated compound | 105| 0 0.000 38 0.362 | 0 0.000 0 0.000 | 0.0 | 0.0 |
| heterocyclic   | 57| 0 0.000 3 0.053                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| inorganic      | 3 | 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| ketone         | 33| 0 0.000 3 0.091                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| nitro          | 10| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| phenol         | 18| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| thiol          | 16| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0 0.000 0 0.000                   | 0.0                                      |
| all molecules  | 670| 1 0.001 — — | 1 0.001 — — | 0.2 | 0.2 |

Table S45: Correlation of experimental spectra against database of theoretical spectra in the range 2000–3846 cm\(^{-1}\) for GAFF-BCC sorted according to compound classification.

| Class          | N | Pearson \(r\) Matched Fraction       | Pearson \(r\) Clases Matched Fraction | Spearman \(\rho\) Matched Fraction | Spearman \(\rho\) Clases Matched Fraction |
|----------------|---|-------------------------------------|---------------------------------------|-----------------------------------|------------------------------------------|
| alcohol        | 76| 0 0.000 4 0.053                     | 0 0.000 38 0.500                      | 16.0                              | 16.0                                     |
| aldehyde       | 12| 0 0.000 5 0.417                     | 0 0.000 3 0.250                      | 3.0                               | 3.0                                      |
| alkane         | 82| 0 0.000 6 0.073                     | 0 0.000 30 0.366                     | 7.9                               | 7.9                                      |
| alkene         | 174| 0 0.000 122 0.701                   | 0 0.000 8 0.046                     | 0.0                               | 0.0                                      |
| alkylthromide  | 20| 0 0.000 1 0.050                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| alkylchloride  | 41| 0 0.000 0 0.000                     | 0 0.000 3 0.073                      | 0.3                               | 0.3                                      |
| alkylfluoride  | 8 | 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| alkyne         | 27| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| amide          | 10| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| amine          | 41| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| aromatic       | 172| 0 0.000 19 0.110                    | 0 0.000 22 0.128                     | 0.0                               | 0.0                                      |
| arylchloride   | 23| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| arylfluoride   | 13| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| cycloalkane    | 39| 0 0.000 2 0.051                     | 0 0.000 0 0.000                      | 0.2                               | 0.2                                      |
| cycloalkene    | 11| 0 0.000 0 0.000                     | 0 0.000 0 0.000                      | 0.0                               | 0.0                                      |
| halogenated compound | 105| 0 0.000 5 0.048 | 0 0.000 0 0.000 | 0.0 | 0.0 |
| heterocyclic   | 57| 0 0.000 3 0.053                     | 1 0.018 11 0.193                     | 2.1                               | 2.1                                      |
| inorganic      | 3 | 0 0.000 0 0.000                     | 0 0.000 1 0.333                     | 1.9                               | 1.9                                      |
Table S46: Correlation of theoretical spectra against database of experimental spectra in the range 2000–3846 cm\(^{-1}\) for GAFF-ESP sorted according to compound classification.

| Class      | N  | Pearson r Molecules |               | Spearman ρ Molecules |               |
|------------|----|---------------------|---------------|----------------------|---------------|
|            |    | Matched Fraction    | Matched Fraction | Power                | Matched Fraction | Matched Fraction | Power                |
| ketone     | 33 | 0 0.000 2 0.061 0.3 |               | 0 0.000 0 0.000 0.0 |               |
| nitro      | 10 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| phenol     | 18 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| thiol      | 16 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| all molecules | 670 | 0 0.000 — — 0.0 |               | 3 0.004 — — 1.1 |               |

Table S47: Correlation of experimental spectra against database of theoretical spectra in the range 2000–3846 cm\(^{-1}\) for GAFF-ESP sorted according to compound classification.

| Class      | N  | Pearson r Molecules |               | Spearman ρ Molecules |               |
|------------|----|---------------------|---------------|----------------------|---------------|
|            |    | Matched Fraction    | Matched Fraction | Power                | Matched Fraction | Matched Fraction | Power                |
| alcohol    | 76 | 0 0.000 6 0.105 0.4 |               | 0 0.000 0 0.000 0.0 |               |
| aldehyde   | 12 | 0 0.000 1 0.003 0.7 |               | 0 0.000 0 0.000 0.0 |               |
| alkane     | 82 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| alkene     | 174| 1 0.006 105 0.603 20.8 |               | 0 0.000 14 0.030 0.1 |               |
| alkylbromide | 20 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| alkylfluoride | 41 | 0 0.000 1 0.024 0.0 |               | 0 0.000 3 0.073 0.3 |               |
| alkyne     | 27 | 0 0.000 4 0.250 3.3 |               | 0 0.000 0 0.000 0.0 |               |
| amide      | 10 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| amine      | 41 | 0 0.000 1 0.024 0.0 |               | 0 0.000 1 0.024 0.0 |               |
| aromatic   | 172| 0 0.000 33 0.192 0.0 |               | 0 0.000 14 0.081 0.0 |               |
| aromatic   | 23 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| arylchloride | 13 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| cycloalkane | 39 | 0 0.000 15 0.385 8.7 |               | 0 0.000 0 0.000 0.0 |               |
| cycloalkene | 11 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| halogenated compound | 105 | 0 0.000 31 0.295 3.6 |               | 0 0.000 0 0.000 0.0 |               |
| heterocyclic | 57 | 0 0.000 6 0.105 0.4 |               | 0 0.000 0 0.000 0.0 |               |
| inorganic  | 3  | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| ketone     | 33 | 0 0.000 2 0.061 0.3 |               | 0 0.000 1 0.030 0.1 |               |
| nitro      | 10 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| phenol     | 18 | 0 0.000 0 0.000 0.0 |               | 0 0.000 0 0.000 0.0 |               |
| thiol      | 16 | 0 0.000 4 0.250 3.3 |               | 0 0.000 0 0.000 0.0 |               |
| all molecules | 670 | 0 0.000 — — 0.0 |               | 0 0.000 — — 0.0 |               |
| Class               | N   | Molecules | Pearson $r$ | Spearman $\rho$ |
|--------------------|-----|-----------|-------------|-----------------|
|                    |     | Matched   | Fraction    | Matched         | Fraction    | Power |
|                    |     | Molecules |             | Classes         | Power       |
|                    |     | Matched   | Fraction    |                  | Matched     | Fraction    | Power |
|                    |     | Molecules |             |                  |             |             |     |
| amide              | 10  | 0         | 0.000       | 1                | 0.100       | 0.9         |     |
| amine              | 41  | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| aromatic           | 172 | 0         | 0.000       | 22               | 0.128       | 0.0         |     |
| arylchloride       | 23  | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| arylfluoride       | 13  | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| cycloalkane        | 39  | 0         | 0.000       | 5                | 0.128       | 1.1         |     |
| cycloalkene        | 11  | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| halogenated compound | 105 | 0         | 0.000       | 10               | 0.095       | 0.0         |     |
| heterocyclic       | 57  | 0         | 0.000       | 6                | 0.105       | 0.4         |     |
| inorganic          | 3   | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| ketone             | 33  | 0         | 0.000       | 1                | 0.030       | 0.1         |     |
| nitro              | 10  | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| phenol             | 18  | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| thiol              | 16  | 0         | 0.000       | 0                | 0.000       | 0.0         |     |
| all molecules      | 670 | 1         | 0.001       | —                | —           | 0.2         |     |

Table S48: Correlation of theoretical spectra against database of experimental spectra in the range 2000–3846 cm$^{-1}$ for OPLS sorted according to compound classification.
Table S49: Correlation of experimental spectra against database of theoretical spectra in the range 2000–3846 cm\(^{-1}\) for OPLS sorted according to compound classification.

| Class                  | N   | Molecules Matched | Fraction | Pearson \(r\) | Power | Molecules Matched | Classes Fraction | Spearman \(\rho\) | Power |
|------------------------|-----|--------------------|----------|----------------|-------|--------------------|-----------------|------------------|-------|
| alcohol                | 76  | 0                  | 0.000    | 0.053          | 0.0   | 1                  | 0.015           | 48               | 0.632 | 26.1 |
| aldehyde               | 12  | 0                  | 0.000    | 0.083          | 0.7   | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| alkane                 | 82  | 0                  | 0.000    | 0.146          | 0.5   | 0                  | 0.000           | 34               | 0.415 | 10.5 |
| aldehyde               | 174 | 1                  | 0.006    | 0.787          | 47.0  | 0                  | 0.000           | 10               | 0.057 | 0.0  |
| alkyl bromide          | 20  | 0                  | 0.000    | 0.100          | 0.9   | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| alkyl fluoride         | 41  | 0                  | 0.000    | 0.024          | 0.0   | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| alkyne                 | 8   | 1                  | 0.125    | 0.125          | 1.0   | 1                  | 0.125           | 3                | 0.375 | 4.0  |
| amide                  | 10  | 0                  | 0.000    | 0              | 0.000 | 0                  | 0.000           | 1                | 0.100 | 0.9  |
| amine                  | 41  | 1                  | 0.024    | 0.024          | 0.0   | 0                  | 0.000           | 3                | 0.073 | 0.3  |
| aromatic               | 172 | 1                  | 0.006    | 0.616          | 22.5  | 0                  | 0.000           | 40               | 0.233 | 0.1  |
| aryliclORIDE           | 23  | 0                  | 0.000    | 0              | 0.000 | 0                  | 0.000           | 1                | 0.043 | 0.3  |
| aryldifluoride         | 13  | 0                  | 0.000    | 0              | 0.000 | 0                  | 0.000           | 1                | 0.077 | 0.6  |
| cycloalkane            | 39  | 0                  | 0.000    | 0.077          | 0.4   | 0                  | 0.000           | 2                | 0.051 | 0.2  |
| cycloalkene            | 11  | 0                  | 0.000    | 0              | 0.000 | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| halogenated compound   | 105 | 1                  | 0.010    | 0.114          | 0.0   | 1                  | 0.010           | 13               | 0.124 | 0.1  |
| heterocyclic           | 57  | 0                  | 0.000    | 0.246          | 3.6   | 0                  | 0.000           | 3                | 0.053 | 0.1  |
| inorganic              | 3   | 1                  | 0.333    | 0.333          | 1.9   | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| ketone                 | 33  | 0                  | 0.000    | 0              | 0.000 | 0                  | 0.000           | 1                | 0.030 | 0.1  |
| nitro                  | 10  | 0                  | 0.000    | 0              | 0.000 | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| phenol                 | 18  | 0                  | 0.000    | 0.278          | 4.0   | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| thiol                  | 16  | 0                  | 0.000    | 0              | 0.000 | 0                  | 0.000           | 0                | 0.000 | 0.0  |
| all molecules          | 670 | 4                  | 0.006    | —              | —     | 1                  | 0.007           | —                | 2.4   |      |

S5  List of Compounds

1,1,1,3,3,3–hexafluoro–2–propanol  
1,1,1–trichloroethane  
1,1,2,2–tetrabromoethane  
1,1,2,2–tetrachloroethane  
1,1,2–trichloroethane  
1,1,3,3–tetramethylurea  
1,1–biphenyl  
1,1–dichloroethane  
1,1–dichloroethene  
1,1–difluoroethane

1,1–difluoroethene  
1,1–dimethoxyethane  
1,1–dimethylethanol  
1,1–dimethylcyclohexane  
1,1–dimethylcyclopropane  
1,2,3,4,5–pentafluorobenzene  
1,2,3,4–tetrafluorobenzene  
1,2,3,4–tetrahydroxynaphthalene  
1,2,3,4–tetramethylbenzene  
1,2,3,6–tetrahydrobenzaldehyde  
1,2,3–propanetriol  
1,2,3–trichlorobenzene
1,2,3-trichloropropane
1,2,3-trimethylbenzene
1,2,3-trimethylcyclopent-1-ene
1,2,4,5-tetrafluorobenzene
1,2,4-trichlorobenzene
1,2,4-trifluorobenzene
1,2,4-trimethylbenzene
1,2-bis-2-methoxyethoxyethane
1,2-dibromobenzene
1,2-dibromoethane
1,2-dibromopropane
1,2-dichlorobenzene
1,2-dichloroethene
1,2-dichloroethane
1,2-dichloropropane
1,2-difluorobenzene
1,2-dimethoxybenzene
1,2-dimethoxyethane
1,2-ethanediamine
1,2-ethanedithiol
1,2-propylene-oxide
1,3-benzothiazole
1,3-dibromobenzene
1,3-dichlorobenzene
1,3-dichloropropane
1,3-difluorobenzene
1,3-dioxane
1,3-dioxolane
1,3-dithiane
1,3-oxazole
1-(3-pyridinyl)ethanone
1,3-thiazole
1,4-butanedithiol
1-(4-chlorophenyl)ethanone
1,4-dichlorobenzene
1,4-dichlorobutane
1,4-dichloro-trans-2-butene
1,4-difluorobenzene
1,5-dimethylcyclopent-1-ene
1,5-pentanediol
1-bromo-2-chloroethane
1-bromo-3-methylbutane
1-bromobutane
1-bromodecane
1-bromoheptane
1-bromohexane
1-bromonaphthalene
1-bromononane
1-bromoocytane
1-bromopentane
1-butanethiol
1-butanol
1-butene
1-butyne
1-chloro-1,1-difluoroethane
1-chloro-2-methylpropane
1-chloro-2-propanol
1-chloro-3-methylbutane
1-chloro-3-nitrobenzene
1-chloro-4-nitrobenzene
1-chloronaphthalene
1-chlorooctane
1-chloropentane
1-chloropropane
1-diethoxyphosphorylethene
1-ethoxy-2-(2-ethoxyethoxy)ethane
(1-ethoxyethyl)phosphoryl-oxethane
1-ethyl-1-methylcyclopentane
1-ethyl-2-methylbenzene
1-ethyl-3-methylbenzene
1-ethyl-4-methylbenzene
1-ethylcyclopent-1-ene
1-ethynaphthalene
1-fluoro-4-nitrobenzene
1-hexyne
1H-indene
1-methyl-2-nitrobenzene
1-methyl-3-nitrobenzene
1-methyl-4-propan-2-ylbenzene
1-methynaphthalene
1-methylpyrrole
1-methylpyrrolidin-2-one
1-nitropropane
1-octanol
1-pentene
1-pentyne
1-phenyl-1-propanol
1-phenylpropan-1-one
1-phenylpropan-2-one
1-propanethiol
1-propanol
1R,2R-1,2-dimethylcyclopropane
1R,2R,3S-1,2,3-trimethylcyclopentane
1R,2S-1,2-dimethylcyclohexane
1R,2S-1,2-dimethylcyclopentane
1R,2S-1,2-dimethylcyclopropane
1R,4R-1,4-dimethylcyclohexane
1S,2R-1-ethyl-2-methylcyclopropane
1S,2S-1,2-dimethylcyclohexane
1S,4S-1,4-dimethylcyclohexane
1Z,3Z-cycloocta-1,3-diene
1Z,5Z-cycloocta-1,5-diene
2,2,2-trichloroethanol
2,2,2-trifluoroethanol
2,2,3,3-tetrafluoro-1-propanol
2,2,3-trimethylbutane
2,2,4-trimethylpent-2-ene
2,2,4-trimethylpentane
2,2-dichloro-1,1,1-trifluoroethane
2,2-dimethylbutane
2,2-dimethylpentane
2,2-dimethylpropanenitrile
2-(2-ethoxyethoxy)ethylacetate
2,3,3-trimethylbut-1-ene
2,3,3-trimethylpentane
2,3,4-trimethylpent-2-ene
2,3,4-trimethylpentane
2,3-dihydro-1H-indene
2,3-dihydro-1H-indole
2,3-dihydrofuran
2,3-dimethylbut-1-ene
2,3-dimethylbutane
2,3-dimethylhex-1-ene
2,3-dimethylhex-2-ene
2,3-dimethylhexane
2,3-dimethylpent-1-ene
2,3-dimethylpent-2-ene
2,3-dimethylpentane
2,4,4-trimethylpent-1-ene
2,4,6-trichlorophenol
2,4-dichlorotoluene
2,4-dimethyl-3-pentanone
2,4-dimethylaniline 2-chloropyridine
2,4-dimethylhexane 2-chlorotoluene
2,4-dimethylpent-1-ene 2E-3,4-dimethylhex-2-ene
2,4-dimethylpent-2-ene 2E-3-methylhex-2-ene
2,4-dimethylpentan-3-ol 2E-4,4-dimethylhex-2-ene
2,4-dimethylpentane 2E-4,4-dimethylpent-2-ene
2,4-pentanediol 2E,4E-hexa-2,4-diene
2,4-pentanedione 2E-4-methylhept-2-ene
2,5-dimethylhexa-2,4-diene 2E-5-methylhex-2-ene
2,5-dimethylhex-1-ene 2E-oct-2-ene
2,5-dimethylhex-2-ene 2-ethoxyethanol
2,5-dimethylhexane 2-ethoxyethylacetate
2,5-norbornadiene 2-ethyl-1-hexanol
2,6-dimethylheptane 2-ethylbut-1-ene
2,6-dimethylpyridine 2-ethynaphthalene
2,6-xylenol 2-ethylphenol
2-acetyloxyethylacetate 2-fluoropyridine
2-aminoethanol 2-hexanone
2-bromobutane 2-hydroxyacetophenone
2-bromopentane 2-mercaptoethanol
2-bromopropene 2-methoxy-2-methylpropane
2-bromopyridine 2-methoxyethanol
2-butanol 2-methoxyphenol
2-butane 2-methyl-13-butadiene
2-butoxyethanol 2-methyl-1-butanol
2-butyne 2-methyl-2-pentanediol
2-chloro-2-methylpropane 2-methyl-2-butanol
2-chloroaniline 2-methyl-2-propanethiol
2-chlorobutanoic acid 2-methyl-2-propanol
2-chloroethanol 2-methyl-3-methylidenepentane
2-chloronaphthalene 2-methylaniline
2-chloropropane 2-methylbut-1-ene

S48
2-methylbut-2-ene  2Z-3-methylhex-2-ene
2-methyldecane  2Z-4,4-dimethylpent-2-ene
2-methylfuran  2Z-4-methylhex-2-ene
2-methylenept-2-ene  2Z-5,5-dimethylhex-2-ene
2-methylheptane  2Z-5-methylhex-2-ene
2-methylhex-1-ene  3,3-dimethyl-1-butene
2-methylhex-2-ene  3,3-dimethylbut-1-ene
2-methylmethoxybenzene  3,3-dimethylbutan-2-one
2-methylnonane  3,3-dimethylhex-1-ene
2-methyl-N-phenylmethylpropan-2-amine  3,3-dimethylhexane
2-methylpent-1-ene  3,3-dimethylpent-1-ene
2-methylpenta-1,4-diene  3,3-dimethylpentane
2-methylpentane  3,4-dichloro-1-butene
2-methylphenol  3,4-dihydro-2H-pyran
2-methylprop-2-enenitrile  3,4-dimethylhex-1-ene
2-methylpropanenitrile  3,4-dimethylhexane
2-methylpropanoic-acid  3,4-dimethylpent-1-ene
2-methylpropyl-cyclopentane  3,5-dichloroaniline
2-methylpyridine  3-amino-1-propanol
2-methylquinoline  3-bromopentane
2-nitropropane  3-bromopropene
2-pentanone  3-chloro-1,2-propanediol
2-phenylethanol  3-chlorophenol
2-propan-2-ylphenol  3-chloropropan-1-ol
2-propanethiol  3-chloropropene
2-propanol  3-chloropyridine
2-propenenitrile  3-chlorotoluene
2-propyn-1-ol  3E-2,2-dimethylhex-3-ene
2R-1-bromo-2-methylbutane  3E-2,5-dimethylhex-3-ene
2S,5R-2,5-dimethyltetrahydrofuran  3E-2-methylhex-3-ene
2-thiophene-carbonitrile  3E-2-methylpenta-1,3-diene
2Z-3,4,4-trimethylpent-2-ene  3E-3-methyl-1,3-pentadiene
2Z-3,4-dimethylpent-2-ene  3E-5-methylhept-3-ene
S49
3E-hept-3-ene
3E-oct-3-ene
3-ethyl-2-methylpent-1-ene
3-ethyl-2-methylpent-2-ene
3-ethyl-2-methylpentane
3-ethyl-3-methylpent-1-ene
3-ethyl-4-methylpent-1-ene
3-ethylcyclopent-1-ene
3-ethylhex-3-ene
3-ethylhexane
3-ethylpent-1-ene
3-ethylpentane
3-ethylpentan-3-ol
3-ethylpyridine
3-fluoropyridine
3-fluorotoluene
3-heptanol
3-heptanone
3-hexanol
3-hydroquinone
3-hydroxymethyl-pyridine
3-methoxypropanenitrile
3-methyl-1-butaneliol
3-methyl-1-butanol
3-methyl-1-butene
3-methyl-1-butylacetate
3-methyl-1-butyne
3-methyl-1-pentyne
3-methyl-2-butanone
3-methyl-2-cyclopenten-1-one
3-methyl-4-methylidenehexane
3-methylaniline
3-methylbenzenethiol
3-methylbutanoic-acid
3-methylbutyraldehyde
3-methylcyclopentanone
3-methylcyclopentanone
3-methylcyclopentene
3-methylheptane
3-methylhex-1-ene
3-methylhexane
3-methylideneheptane
3-methylpent-1-ene
3-methylpentane
3-methylpentane
3-methylphenol
3-methylpyridine
3-methylsulfolane
3-methylthiophene
3-pentanol
3-pentanone
3-phenylpropan-1-ol
3-pyridinecarboxaldehyde
3Z-2,5-dimethylhex-3-ene
3Z-2-methylhept-3-ene
3Z-2-methylhex-3-ene
3Z-3-methylhex-3-ene
3Z-3-methyl-penta-13-diene
4-1H-1,3-benzodiazol-2-yl-1,3-thiazole
4,4-dimethylhex-1-ene
4,4-dimethylpent-2-yne
4,5-dimethylhex-1-ene
4-bromoaniline
4-chlorobenzenemethanethiol
4-chlorobutanoic-acid
4-chlorophenol
4-chlorotoluene
4E-hexa-1,4-diene
4E-oct-4-ene
4-ethenylcyclohex-1-ene
4-ethylhex-1-ene
4-fluoroaniline
4-methoxybenzaldehyde
4-methyl-1-pentyne
4-methylaniline
4-methylheptane
4-methylhex-1-ene
4-methylhex-2-yne
4-methylmethoxybenzene
4-methyloctane
4-methylpent-1-ene
4-methylpenta-1,3-diene
4-methylpentan-2-one
4-methylphenol
4-methylpyridine
4R-4-methylhept-1-ene
4Z-oct-4-ene
5-methyl-2(3H)-furanone
5-methylhex-1-yne
6-methylhept-1-ene
7-methyloct-3-yne
8-hydroxyquinoline
9-chloroanthracene
9H-fluorene
acenaphthene
acetic anhydride
acetonitrile
acetophenone
acridine
allyl-acrylate
allylamine
ammonia
aniline
anisole
anthracene-9,10-dione
anthracene
azulene
benzaldehyde
benzamide
benzene
benzenethiol
benzonitrile
benzoyl-chloride
bis-2-chloroethyl-ether
bromochloromethane
bromoethane
bromoethene
but-3-en-2-ol
buta-1,3-diene
butanal
butane-1,2-diol
butane-2,3-diol
butane
butanoic-acid
butylbenzene
butylcyclopentane
butyl-ethyl-ether
butyl-formate
butyl-vinyl-ether
caprolactam
chloroacetyl-chloride
chlorocyclohexane-axial
chloroethane
dichloromethane
cyclohepta-1,3,5-triene
dicyclopenty1methanone
cyclohexa-1,3-diene
diethyl-carbonate
cyclohexa-1,4-diene
diethylene-glycol
cyclohexane
diethylene-glycol-monoethyl-ether
cyclohexanethiol
diethylene-glycol-monomethyl-ether
cyclohexanol
diethyl-malonate
cyclohexanone
diethyl-oxalate
cyclohexene
diethyl-sulfide
cyclohexylamine
diisopropyl-ether
cycloheptylamine
dimethoxymethane
cyclooctane
dimethylethanolamine
cyclooctatetraene
dimethylamine
cyclooctene
dimethyl-carbonate
cyclopent-2-en-1-one
dimethyl-disulfide
cyclopentanol
dimethylsulfide
cyclopentanone
dimethylsulfone
cyclopentene
dimethylether
cyclopropylamine
dimethyl-sulfate
cyclopropylmethylketone
dimethyl-sulfoxide
decan-1-ol
dipenty1-ether
decane
diphenyl-ether
dodecane
diphenylmethanone
dodecanethiol
diacetone-alcohol
dodecane
dibromomethane
dodecanetwo
dibutylamine
e-2-butenene
dibutyl-ether
e-3-methylpent-2-ene
dichloroacetic-acid
e-4-methylpent-2-ene
e-2-pent-2-ene
ethane
e-thanethiol
ethanol
ethoxymethoxyethane  
ethyl−2−chloroacetate  
ethyl−2−cyanoacetate  
ethyl−2−hydroxybenzoate  
ethyl−3−oxobutanoate  
ethyl−acetate  
ethyl−acrylate  
ethylamine  
ethylbenzene  
ethyl−benzoate  
ethyl−butanoate  
ethylcyclobutane  
ethylcyclohexane  
ethylcyclopropane  
ethyleneglycol  
ethyl−formate  
ethyl−propanoate  
ethyl−propiolate  
ethyl−trans−cinnamate  
fluoranthene  
fluorobenzene  
fluoroethene  
formamide  
formic−acid  
furfural  
furfuryl−alcohol  
hept−1−yne  
heptan−1−ol  
heptan−1−one  
heptane−1−thiol  
heptane  
heptanoic−acid  
hex−1−ene  

hex−3−yne  
hexa−1,5−diyne  
hexadecane  
hexene−1,6−dithiol  
hexene−1−thiol  
hexane  
hexanoic−acid  
indole  
isobutane  
isobutyl−acrylate  
isobutyl−formate  
isopentane  
isopropylamine  
isopropylbenzene  
isopropylcyclopropane  
isquinoline  
methacrylic−acid  
methane  
methyl−acetate  
methyl−acetoacetate  
methyl−acrylate  
methyl−benzoate  
methyl−butanoate  
methyl−chloroacetate  
methylcyclobutane  
methylcyclohexane  
methylcyclopentane  
methylcyclopentene  
methylenecyclohexane  
methyl−formate  
methyl−hexanoate  
methyl−isocyanate  
methyl−isothiocyanate
methyl-methacrylate
methyl-octanoate
methylpropene
methyl-propionate
methyl-vinyl-ether
morpholine
m-phenylenediamine
m-tolualdehyde
n-aminoethyl-ethanolamine
naphthalen-1-amine
naphthalen-2-amine
naphthalene-1-carbaldehyde
naphthalene-2-carbaldehyde
naphthalene
N-buty lacetamide
N-butyl-N-phenylacetamide
neopentane
N-ethylacetamide
N-ethyl aniline
N-ethylmorpholine
nitrobenzene
nitroethane
nitromethane
N-methyl-1-phenylmethanamine
N-methylaniline
N-methylformamide
N-methyl-N-phenyl-formamide
N,N-diethylacetamide
N,N-diethylaniline
N,N-diethylformamide
N,N-dimethylacetamide
N,N-dimethylaniline
non-1-yn e
non-2-yn e
non-4-yn e
nonane
nonanoic-acid
N-pentylamine
N-prop-2-enylprop-2-en-1-amine
n-propyl-propionate
oct-1-ene
oct-1-yn e
oct-2-yn e
oct-3-yn e
oct-4-yn e
octane
octanoic-acid
o-ethylaniline
oxetane
oxolan-2-one
paraldehyde
pent-2-yn e
pentachloroethane
pentadecane
pentane
pentanoic-acid
pentlyl-formate
p-ethylphenol
phenanthrene
phenazine
phenetole
phenol
phenyl-acetate
phenylacetylene
phenylmethanethiol
phenylmethanol
phenylmethoxymethylbenzene  
phthalan  
p−hydroquinone  
p−hydroxybenzoic−acid  
piperidine−axial  
prop−1−en−2−ylbenzene  
prop−2−en−1−ol  
prop−2−en−1−ylbenzene  
propane−1,2−diol  
propane−1,3−diol  
propane  
propanenitrile  
propanoic−acid  
propenal  
propene  
propyl−acetate  
propylamine  
propylbenzene  
propylene−glycol−monomethyl−ether  
propyl−formate  
pyrene  
pyrimididine  
pyrrole  
pyrrolidine−axial  
quinoline  
salicylaldehyde  
sec−butyl−acetate  
sec−hexyl−acetate  
spiropentane  
succinic−acid  
succinimide  
tert−butylamine  
tert−butylbenzene  
tetrahydrofuran  
tetrahydrofurfuryl−alcohol  
tetrahydropyran  
thiane  
thiobis−2−ethanol  
trans−1,2−dichloroethene  
trans−1,3−pentadiene  
trans−1−propenylbenzene  
trans−3−methylcyclohexanol  
trans−decalin  
trichloromethyl−benzene  
trifluorooacetic−acid  
trifluoromethyl−benzene  
undecane  
undecanenitrile  
vinyl−acetate  
Z−3−methylpent−2−ene  
Z−4−methylpent−2−ene  
Z−hex−2−ene  
Z−hex−3−ene  
Z−pent−2−ene