Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/2/11/e1600954/DC1

Supplementary Materials for

Materials design by evolutionary optimization of functional groups in metal-organic frameworks

Sean P. Collins, Thomas D. Daff, Sarah S. Piotrkowski, Tom K. Woo

Published 23 November 2016, *Sci. Adv.* **2**, e1600954 (2016)
DOI: 10.1126/sciadv.1600954

This PDF file includes:

- table S3. Values used to fit transformation function (eq. S2) of performance properties.
- table S4. Fitted values used in eq. S2 for each performance properties. R^2 values are calculated using table S3 values.
- table S5. Sterically viable structures for training MOFs.
- table S6. Parameters used by MOFF-GA that were optimized.
- table S7. Terms used in parasitic energy with a brief description.
- table S8. Functionalized MOFs with CO_2 uptake greater than 3 mmol/g with the corresponding functional groups.
- table S9. Details of functional group codes and their associated structure.
- References (*33*–*39*)

Supplementary Materials

1. Details of the GA

Genetic Representation:

Genetic Algorithms (GAs) are built upon the genetic representation, or chromosome, of the system. For MOFF-GA we were interested in optimizing the functional groups on the organic linker, also known as a secondary building unit (SBU). The hydrogen atoms of the SBU are labelled in order to identify each unique functionalization site. Sites which would heavily increase synthetic difficulty if functionalized, such as those on the nitrogen of aniline, were ignored. The functionalized SBU could then be represented by the functional groups attached to each labelled site which we call the Functional Group Code (FGC). The example in fig. S1 shows the labelled sites H1, H2, H3 on the SBU and the FGC F@H1.Cl@H3 which indicates that site H1 is functionalized with a fluoride group and H2 is functionalized with a chloride group. For unfunctionalized sites the label is omitted from the FGC. The combination of parent MOF and FGC uniquely identify each functionalized MOF.

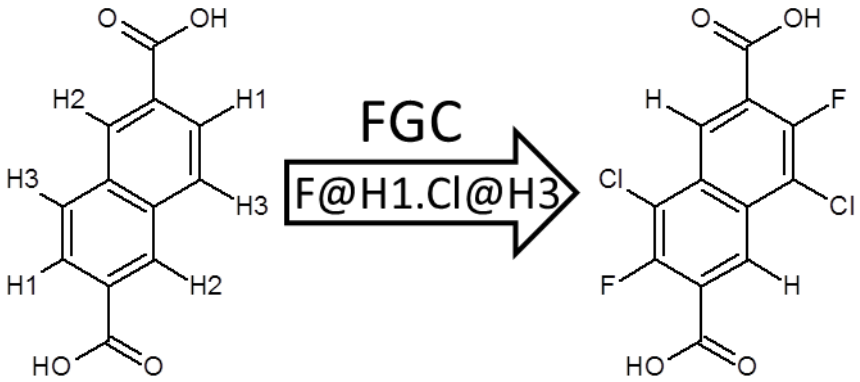


fig. S1. Example of the application of a functional group code to the unfunctionalized SBU of the parent MOF.

General Procedure:

Our GA follows most of the same procedures as other GAs. An initial set of individuals are randomly created, the number of members of the set is known as the *Population*. A set of individuals at a given time is known as a generation. All individuals in the generation are evaluated for their fitness, such as CO_2 uptake. The next generation is constructed form the previous one with mating and mutation mechanisms. Our GA employs elitism which carries forward a fraction of the top performing individuals from one generation into the next generation with no modification. The fraction of top performers carried forward is a known as the *Elite*. The top performers are monitored until they converge on a result. Several parameters, (described in Section 2) are used to tune the performance of the whole procedure.

Mating Scheme:

Most of the individuals in a generation are created by a mating mechanism. Mating is an important part of how a GA works as it ensures the new generation inherits favourable traits of the parents.

Choosing Parents:

In order to create the new generation with higher performing individuals the top performers from the previous generation need to be selected. We use a single metric, *x*, that measures the performance of the material, such as the CO_2 uptake or parasitic energy (P_E). We then define a scaling function, *s(x)*, which favours the higher performing individuals. The scaling functions will differ based on the optimizing property, as shown in table S1. P_E, for example, used a scaling function that favoured smaller energies as we want to minimize the property. The CO_2 uptake scaling function is used to place more weight on higher uptake materials. This is done as the range of CO_2 uptake within a generation can be limited. An exponential function will give the higher performing individuals a higher weight during parent selection.

table S1. Scaling functions used for fitness.

Property	Scaling Function
CO_2 Uptake	$s(CO_2Uptake) = e^{CO_2Uptake}$
Gravimetric Surface Area	$s(SA) = SA$
Parasitic Energy	$s(P_E) = \frac{1}{P_E}$

In eq. 1 the i^{th} individual of the population has its scaled performance, $s(x_i)$, normalized to the entire population. For the entire population this creates a set which sums to 1 with no individual going below 0. These values are able to be used in a selection process known as fitness proportionate selection, or more commonly, roulette wheel. The roulette wheel technique allows any member of the generation to be selected at random based on its weight. The higher the weight the more likely it will be chosen

$$\boldsymbol{Weight(x_i) = \frac{s(x_i)}{\sum_{n=1}^{pop_n} s(x_i)}} \quad (1)$$

The roulette wheel works by randomly selecting a random number between 0 and 1. The weight of each individual are then added, in descending order, until the cumulative weight is greater than the random number. The individual that caused the cumulative weight to go beyond the random number is selected as a parent. This process is repeated for a second parent. If both parents are the same individual both parents are reselected. This allows the new generation to come up with new, untested individuals to test. Once both parents are chosen they are mated by either a 1 or 2-cut mating scheme. There is a random choice for selecting between the 1 and 2-cut schemes which is known *Single Cut Rate*. The larger the *Single Cut Rate* the more likely the 1-cut mating scheme will occur.

1-cut mating scheme:

In the 1-cut mating scheme, the chromosomes of both parents are cut at a single, randomly selected position and complementary pieces from the two parents are combined. This process is shown in fig. S2. The selection of the first and second portion is also randomly selected with an equal chance.

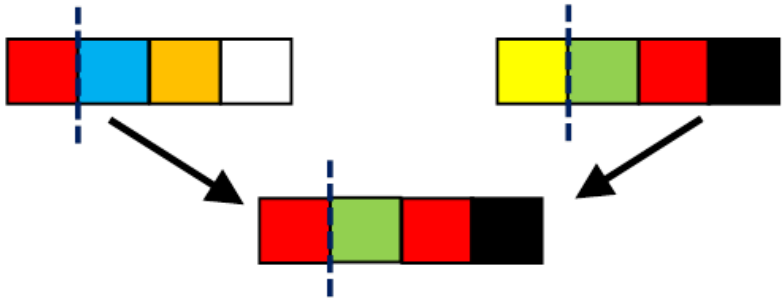


fig. S2. Schematic of the one-cut mating process.

2-cut mating scheme:

The 2 cut mating procedure is similar to that of the 1-cut mating scheme however two unique locations are chosen. The two locations are chosen at random and must not be the same. There is an equal chance for each parent to provide the middle slice or the outer slices. These sections are then joined together to form the new child, as seen in fig. S3.

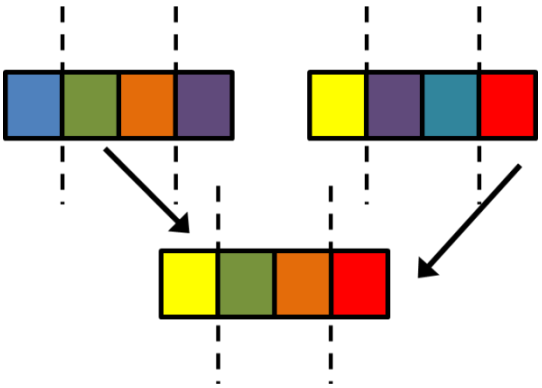


fig. S3. Schematic of the two-cut mating process.

Mutations:

After mating occurs, the new child undergoes mutation. MOFF-GA used two distinct mutations, Type 1 and Type 2 Mutations (described below). The number of mutations a single child can have is equal to the number of functional positions plus 1. Type 1 Mutation can occur once during the mutation process. Type 2 Mutation can occur at every functional position during the mutation process. The rate that mutations occur is known as the *Mutation Rate*. This single value controls how often all mutations occur in MOFF-GA. We have also included a biasing scheme that prefers selection of functional groups that appear more often in high performing members (Described later).

Type 1 Mutation:

The first mutation is known as a swapping mutation. This mutation will swap the functional groups of two randomly chosen functionalization sites. If the two functional groups are identical, then the mutation completes even though there is no effective change in the chromosome. The mutation is schematically shown in fig. S4.

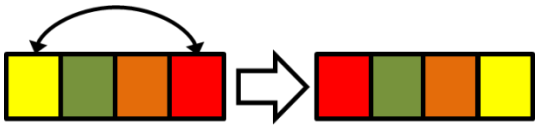


fig. S4. Schematic of swapping mutation.

Type 2 Mutation:

The second mutation is a replacement mutation which changes the functional group at a single functional position. The mutation can occur at every functional group in the chromosome, with a probability given by the *Mutation Rate*, potentially creating a fully random chromosome from the parents. If a functional group position is selected for mutation by the *Mutation Rate* a subsequent choice is made of whether to replace it with a chemically similar or dissimilar

functional group. The choice of similar or dissimilar is chosen randomly according to the *Similarity Probability* parameter. The higher the *Similarity Probability* is, the more likely a chemically similar functional group will be chosen.

Chemical similarity was determined by 3 properties, the Electrostatic Potential (ESP), the van der Waals Potential (VdWP), and steric hindrance. For all functional groups, the groups were aligned as if attached to a benzene ring, and all of the properties were calculated on identical 3D grids which were larger than the largest functional groups. ESPs were calculated using charge equilibration (QEq) atomic charges on the functional group with a point charge probe. The VdWPs were calculated using a Lennard-Jones 6-12 potential with universal force field (UFF) parameters (*18*) and a carbon probe. Steric hindrance was decided with a binary output using the VdWP. If the VdWP at a grid point was 0 or greater it was set as sterically unavailable and assigned a value of 1. If the VdWP was below 0 it was sterically available and assigned a value of 0.

To calculate the similarity between two functional groups we used a continuous Tanimoto coefficient. The Tanimoto coefficient is a pairwise similarity measure shown in eq. 2. In Tanimoto calculations, two functional groups (A and B) have paired components, *i*, compared to calculate the overall similarity. For our chemical similarity the components used were the value of each property at each grid point. The similarity is a normalized value that ranges 0 (maximum dissimilarity) to 1 (the same)

$$\boldsymbol{Tanimoto}(\boldsymbol{A},\boldsymbol{B}) = \frac{\sum_i \boldsymbol{A_i B_i}}{\sum_i \boldsymbol{A_i A_i} + \sum_i \boldsymbol{B_i B_i} - \sum_i \boldsymbol{A_i B_i}} \quad (2)$$

Using the Tanimoto coefficients two unique sets are created for every functional group, the chemically similar and dissimilar sets. These are created by assigning a *Similarity Threshold Value*. This single value is used to discriminate between chemically similar and dissimilar functional groups for each combination. If the Tanimoto coefficient for two functional groups is less than the *Similarity Threshold Value* they are classified as dissimilar while if they higher they are similar.

Biased Functional Group Selection:

When MOFF-GA is initialized all chemically similar (or chemically dissimilar) functional groups have an equal probability of being selected during Type 2 Mutation. The Biased Functional Group (BFG) function makes functional groups that appear more often in high performing members have a greater chance of being selected during Type 2 Mutation. Similarly it will make functional groups which continually appear in low performing individuals and have a lower chance of selection. This process is controlled by 3 unique parameters known as the *Weighting Cut*, the *Weighting Cut-Off* and the *Weighting Change*. The *Weighting Cut* is the fraction of top (bottom) individuals that are considered as the top (bottom) performers in the function. The *Weighting Cut-Off* will determine how often a single functional group needs to appear in the top (bottom) performers for its weighting to be changed. Finally the *Weighting Change* will determine how much to add (subtract) from the weighting of the functional group. The initial weighting for every functional group is set to 50 and is limited to never drop below 1. During selection of functional groups the weighting of a functional group describes the probability of it being selected. For example if two functional groups, A and B, have weightings of 50 and 1 respectively then functional group A will be selected 50 times more often than functional group B.

Stagnation:

Once the top performing individual has remained constant for a set number of generations, determined by *Stagnation*, MOFF-GA enters a stagnation phase. During the stagnation phase MOFF-GA uses 3 methods to create new individuals: mutating the best; random creation; and normal mating. When mutating the best, individuals are created which differ from the best performer by one functional group. All combinations of these individuals are created randomly over stagnant generations and tested for their performance. A fraction of the population each generation, determined by the *Best Mutated* parameter, is reserved for these individuals. Random creation, during the stagnation adds completely randomly made individuals each generation of the stagnation phase. The amount of randomly created individuals each generation is set by the *Random Mutated* parameter. The remaining population are created using the normal mating scheme.

Convergence:

The endpoint of MOFF-GA is based on convergence criteria since the GA cannot know when and if it has found the best individual. Once the criteria are met, the GA will finish. For MOFF-GA there are two convergence criteria. The first is the top performing individual must stay the same for a set amount of generations known as the *Convergence*. The second is that all individuals which differ by only one functional group from the top performer must have been tested. Once these two criteria have been met MOFF-GA is considered complete.

2. GA parameters

The GA has 13 unique parameters which can be modified. All parameters are mentioned in the GA detail (Section 1). Table S2 lists all parameters and their effects on MOFF-GA.

table S2. Description of the MOFF-GA optimization parameters.

Property	Description
Population	Number of individuals within a single generation
Elite	Fraction of best performing individuals carried over to next generation
Single Cut Rate	Probability of performing a 1-cut vs 2-cut mating during mating process
Mutation Rate	Probability of a Type 1 Mutation or Type 2 Mutation at each functional position occurring
Similarity Threshold Value	Similarity threshold value for determining chemically similar and chemically dissimilar functional groups
Similarity Probability	Probability of muting with a chemically similar functional group vs chemically dissimilar functional group
Weighting Cut	Fraction of top and bottom individuals used during weighting change
Weighting Cut-off	Number of ties a single functional group needs to be in the Weighting Cut fraction for a Weighting Change to occur
Weighting Change	Value of weight change for a functional group if 'Weighting Cut-Off' is achieved (Initial for all functional groups is weight of 50)
Stagnation	Number of generations of the same top performer before stagnation phase begins
Best Mutated	Fraction of individuals during stagnation that are similar to the top performer
Random Mutated	Fraction of individuals during stagnation that are randomly created (Not from mating)
Convergence	Minimum number of generations of the same top performer before convergence is achieved

3. Parameter optimization

Genetic Algorithm Performance Index (GAPI):

There are many ways to rank the performance of a GA, such as how often it finds the top performer, or how many individuals are tested. We developed a term known as the genetic algorithm performance index (GAPI) to rank the performance of MOFF-GA by a single number. GAPI (eq. 3) combines three unique MOFF-GA performance properties: 1) how often the top performer is found (Best Find Rate), 2) the number of the top 50 performing individuals found, and 3) how many unique individuals are tested. These were selected with the idea of wanting MOFF-GA to find the top performer, many good performers, and to do so with testing as few

unique individuals as possible (reduce the computations). These are built on having a high best find rate, a high amount of the top 50 MOFs found, as well as few individuals tested as possible

$$\boldsymbol{GAPI = S_{BFR} + S_{Top\,50} + S_{Unique}} \quad (3)$$

S_{BFR}, $S_{Top\ 50}$ and S_{Unique} are transformation functions which convert the best find rate, amount of top 50 found and amount of unique MOFs tested respectively. As we felt no single performance property was more important we constrained each function to lie between 0 and 1. This allowed every performance property an equal opportunity to contribute to the overall GAPI. We decided not to use simple weighting or scaling functions as each performance property was seen as non-linear. For example if we consider two unique cases of MOFF-GA, one where 0 of the top 50 performers are recovered and one where 35 are recovered. Increasing from 0 to 10 of the top 50 performers should have a larger effect on GAPI than increasing from 35 to 45. This is reasonable as it is more important to improve the first amount of performance parameters rather than to fully maximise a property. By applying a more complex transformation function, such as the sigmoidal function shown in eq. 4, we could both scale the properties appropriately and capture the non-linear effects

$$\boldsymbol{S_i(x) = \frac{K}{1 + A * e^{-r*x}} + C} \quad (4)$$

Sigmoidal functions can be easily fit to scaled, non-linear data due to high flexibility of parameters. First we needed to define scale the values of *x*. Using the absolute value for BFR or amount of unique used would cause the bottom term to nearly disappear. We scaled all performance properties to lie between 0 and 1. For best find rate and top 50 recovered this was done by normalizing values to the maximum (100 and 50 respectively). The scaling for the amount unique MOFs tested was found by first subtracting the amount of chemically similar (differing from the top performer by one functional group) and then taking the inverse (eq. 5). The amount of chemically similar MOFs is removed because these need to be tested by the convergence criteria. This would make the amount of unique MOFs less the chemically similar ones the absolute minimum that could be tested

$$Scaling\ Unique\ MOFs = \frac{1}{Unique\ MOFs - Chemically\ Similar} \quad (5)$$

By setting eq. S4, to 0 and 1 at the lowest and highest possible values for *x* respectively, we could rearrange for *K* and *C* in terms of *A* and *r*. This constrained the function to go between 0 and 1 over the range for all possible values of *x* regardless of the values of *A* and *r*. Each transformation function could then be fit by using only *A* and *r* and would still remain within the 0 to 1 range.

The final thing necessary before the fitting could actually occur was to define the remaining data points. These were objectively selected values chosen from *a priori* knowledge of how MOFF-GA worked. The values, seen in table S3, were selected based on a scale of performance. Low Function Value (0.2) would be known as a 'very bad' performance, while a high Function Value, such as 0.9, would be a 'very good' performance. The Function Values of 0 and 1 correspond to the worst and best performances respectively.

table S3. Values used to fit transformation function (eq. S2) of performance properties.

Function Value	BFR (%)	Top 50 Recovered	Unique MOFs	Unique MOFs (2-sites)
0	0	0	Infinite	730
0.2	10	3	3000	350
0.4	30	8	2200	250
0.6	50	14	1600	180
0.8	80	24	1100	90
0.9	95	37	500	30
1	100	50	1	1

Using the sigmoidal function we respected both range from 0 to 1 of the Function Values as well as the non-linearity of the MOFF-GA's performance. By having the constrained range all GAPI's would lie between 0 (worst) and 3 (best). This allowed a quick understanding of how MOFF-GA performed during those trials. The non-linearity of the sigmoidal functions allowed each property to be treated uniquely as previously mentioned. This is most evidently seen during parameter optimization if one performance property reached a 'very good' (0.9) performance while the others were at 'bad' (0.4) levels. It would be more beneficial, and potentially easier, to

improve the two ‘bad’ properties than to try maximise the one already at ‘very good’.

Table S3 shows two sets of Unique MOFs values, one for 2-site MOFs and one for larger site (3+ site) MOFs. The amount needed for 2-site MOFs is significantly smaller than the large search space MOFs. This is best seen that for a large site MOF a ‘very good’ performance was set to 500 unique MOFs tested. For a 2 site MOF 500 individuals would be almost all possible MOFs (784) and would make MOFF-GA unnecessary. This did not allow a good range for the amount of unique MOFs used, and therefore a second transformation function for the amount of unique MOFs tested is used when 2-site MOFs are considered.

All equations were fit using the SciPy package in python. As mentioned only the values of *A* and *r* were fitted. Table S4 shows values used for each transformation function. R^2 is also shown for each transformation function. The 2-site Unique MOFs transformation function had the smallest R^2 at 0.942. Figures S5 to S8 show he transformation functions for each performance property with the values they were fitted against.

table S4. Fitted values used in eq. S2 for each performance properties. R^2 values are calculated using table S3 values.

Property	X	A	r	K	C	R^2
BFR (%)	BFR (%) / 100	-0.992	0.00376	-0.0226	3.015	0.986
Top 50 Recovered	Top 50 Recovered / 50	-1.732	-1.340	0.841	1.150	0.999
Unique MOFs	1 / (Unique – Chemically Similar[a])	8.352	4174.663	1.120	-0.120	0.984
Unique MOFs (2-Site)	1 / (Unique – Chemically Similar[a])	3.239	308.689	1.471	-0.4712	0.942

[a]Chemically Similar refers to the number of MOFs which differ from the top performer by 1 functional group.

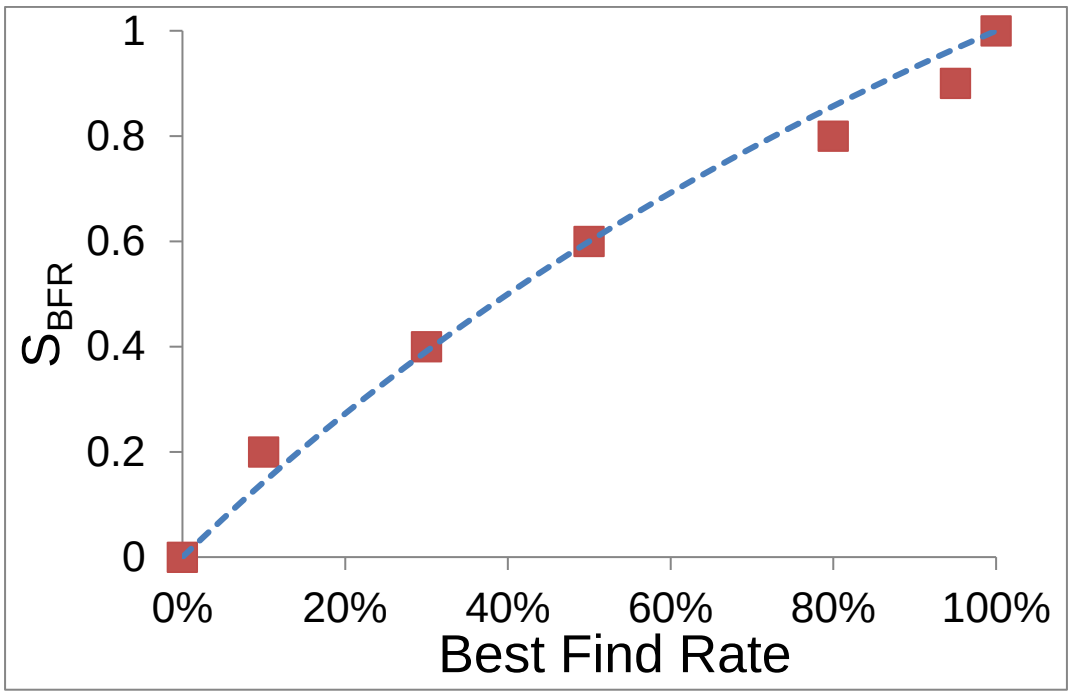


fig. S5. Fitted transformation function (blue dotted line) used in GAPI (genetic algorithm performance indicator) for the best-find rate.

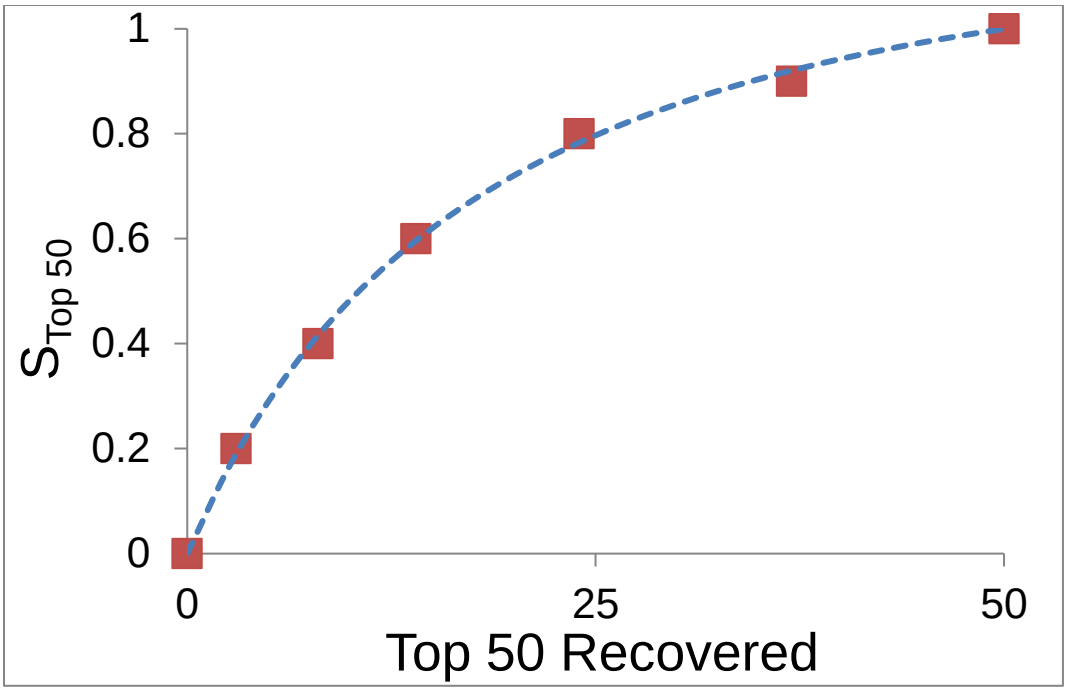


fig. S6. Fitted transformation function (blue dotted line) used in GAPI for the top 50 performers recovered.

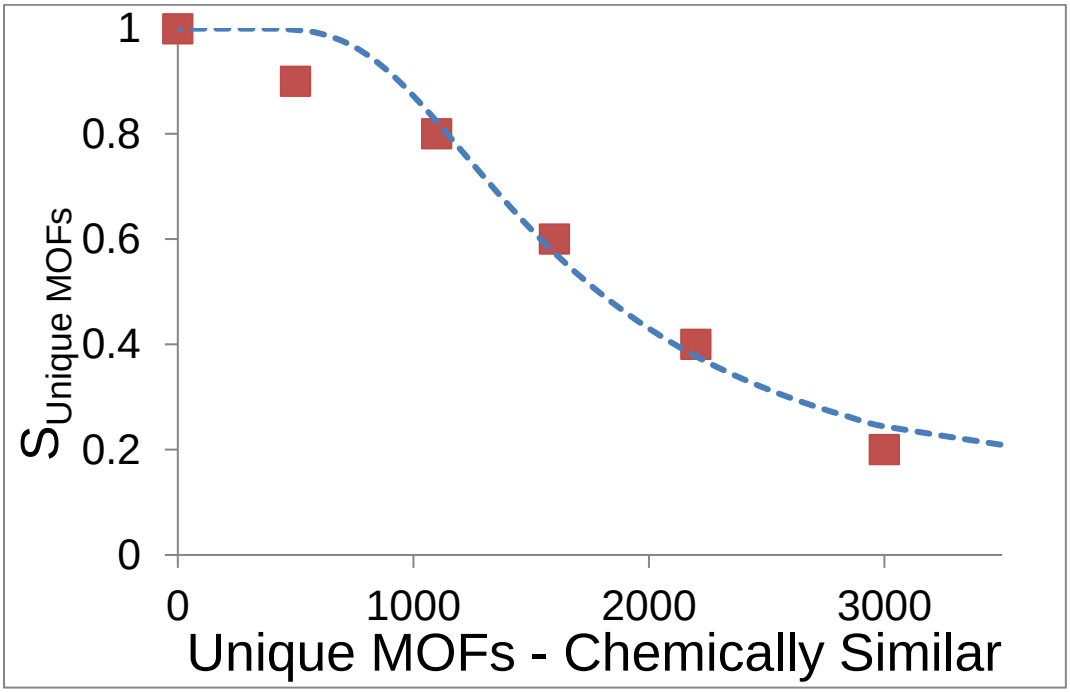


fig. S7. Fitted transformation function (blue dotted line) used in GAPI for unique MOFs tested for large search space (three or more sites) MOFs.

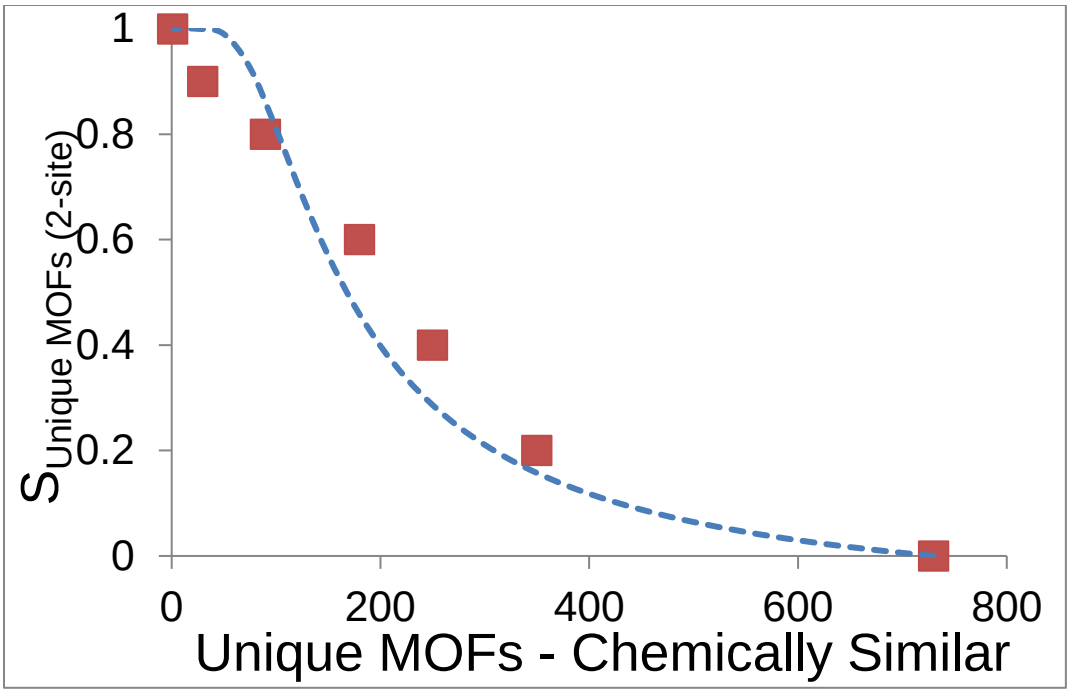


fig. S8. Fitted transformation function (blue dotted line) used in GAPI for unique MOFs tested for two-site MOFs.

Optimization Sets:

To find the 13 optimal MOFF-GA parameters (Section 2), a test set of 7 different MOFs was used: OCIHIS (*33*), bio-MOF-11 (*34*), IRMOF-6 (*35*), MITSEO (*36*), IRMOF-16 (*35*), UTEXAT (*37*) and Zn_2(1-4-benzenedicarboxylate)$_2$(pyrazine), ZBP. For each of these MOFs, the complete search space was evaluated for the three properties: CO_2 uptake at 0.15 atm and 298 K, gravimetric surface area, and the parasitic energy, P_E (see Section 7). Table S5, details the size of the search space for each of the seven MOFs used to optimize the GA parameters. In this work, three sets of MOFF-GA parameters were developed, one for large search spaces (4+-site parameters), one for small search spaces (3-site parameters) and one for very small search spaces, which we call the 2-site GA parameters. To find the optimal parameters for large search spaces, the MOFs ZBP and UTEXAT, which have 4 or more functionalization sites, were used. To find a general and robust set of MOFF-GA parameters, the aggregate GAPI was optimized for all 3 properties, for these 2 MOFs simultaneously. For small search spaces, the same was performed with the 2 MOFS: MITSEO, IRMOF-16. The GA parameters were also optimized for very small search spaces with the 3 MOFs: OCIHIS, Bio-MOF-11 and IRMOF-6.

table S5. Sterically viable structures for training MOFs.

Functional Positions	Total Possible Structures	MOF	Viable Structures
2	784	OCIHIS	621
		Bio-MOF-11	629
		IRMOF-6	644
3	21,952	MITSEO	14,293
		IRMOF-16	17,514
4	614,656	ZBP	96,156
5	> 17 million	UTEXAT	36,501

MOFF-GA Parameter Optimization:

A variety of methods could have been used to optimize the 13 GA parameters; however, we opted to use a GA. As not to confuse it with the MOFF-GA, we will term the GA used to optimize the MOFF-GA parameters, pGA. The structure of pGA is similar to MOFF-GA, where the 13 parameters are their values are used to make up the pGA chromosome (parameter set). Each generation of pGA consisted of 25 unique parameter sets. The convergence criteria of pGA was set to the top performing parameter set remaining constant for 10 generations.
To optimize the parameters, the fitness function used was the sum of the GAPIs for all MOFs in each set (4+-site, 3-site and 2-site), for all three properties (CO_2 uptake, surface area, and P_E). To evaluate the GAPI MOFF-GA was ran 100 times for each MOF, for each property. From those trials the best find rate, the average of the top 50, and number of unique individuals sampled were determined. To ensure high performing parameter sets pGA was run a total of 5 times on each MOF set.

Parameter Set Performance:

Each set of GA parameters were optimized on a subset of the MOFs with complete search

spaces. A subset was selected for parameter optimization due to the computational expenses. There were a total of 50 MOFs (25 2-site, 20 3-site and 5 4+-site) which had a complete area scan completed for the 3 properties. As we only optimized the parameters on the subset of all available MOFs, we tested each of top 5 parameter sets from the parameter optimization.

The total of 10 parameter sets found from small and large search space optimization were tested on both small and large search spaces. It was thought for small and large space MOFs there could be a single robust parameter set that would perform well on all of them. This was thought because values between the small and large space parameter sets were relatively close and could potentially be transferable. The values for the very small search spaces were significantly different than the other parameter sets and were treated differently. We then tested the 5 parameter sets determined from very small search space optimization only on the very small search space MOFs. The parameter sets were tested by performing 100 runs of MOFF-GA, for each property, for each MOF. For each property the 100 trials were used to calculate a GAPI. The sum of all GAPIs was used as the determining factor for performance.

It was found the 2nd best parameter set from the large search space optimization worked best on all MOFs with 3 or more sites, for all properties. We chose this parameter set as default parameters for MOFF-GA. For MOFs with 2-sites it was found the best parameter set from the optimization was the top performing parameter set on very small search spaces. This parameter set was then referred to as 2-site parameters. The exact values for each parameter set and their ranges during optimization are shown in table S6.

4. MOFF-GA parameter values

table S6. Parameters used by MOFF-GA that were optimized.

Property	Min.	Max.	Default (3+ Sites) Parameters	2-Site Parameters
Population	10	400	113	27
Elite	0	0.5	0.272	0.478
Single Cut Rate	0	1	0.958	0.415
Mutation Rate	0	1	0.446	0.298
Similarity Threshold Value	0	0.9	0.312	0.538
Similarity Probability	0	1	0.305	0.417
Weighting Cut	0	0.5	0.419	0.201
Weighting Cut-off	0	20	2	1
Weighting Change	0	50	24	18
Stagnation	0	5	3	1
Best Mutated	0	1 – Elite	0.038	0.256
Random Mutated	0	1 – (Elite + Best Mutated)	0.064	0.036
Convergence	Stagnation + 1	Stagnation + 5	5	2

5. Structure preparation and construction

MOFs were found from the Cambridge Structural Database (CSD). MOFs were cleaned in Materials Studio. Cleaning involved removing guest/solvent molecules, removing disorder from the framework and assigning symmetry. Symmetry was assigned by Materials Studio and used as a basis for the number of functional sites on the SBUs.

For some MOFs the symmetry of the SBUs was greater than the MOF's overall symmetry. An example of this idea is seen in fig. S6. In this example the SBU on the left would be the symmetry determined by Materials Studio. By modifying the names of the hydrogen atoms within the CIF an artificial symmetry was imposed on the SBU. This does not change the symmetry within the CIF but does affect how the MOF is functionalized. If two hydrogen atoms have the same name they were seen as symmetrical by our functionalization program.

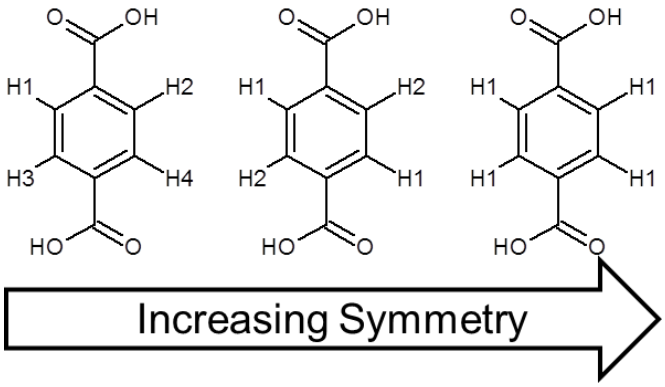


fig. S9. Linker symmetry.

All functionalized SBUs found at the higher symmetry can be found at the lower symmetry. These artificial symmetries were placed on the SBUs to increase synthetic feasibility. At the lower symmetry SBUs, if all synthetic positions contain different functional groups it could be difficult to synthesis both the SBU and the MOF. Additionally the higher symmetry limits the search space as the size if defined by the number of symmetrical positions raised to the number of functional groups available. Although the low symmetry SBU could reach the same functionalizations as the highly symmetrized SBU it is not always guaranteed. For these reasons we have included the same MOFs at different levels of SBU symmetry.

Functionalization of MOFs was carried out using an in-house program, *Fapswitch*. *Fapswitch* works by first identifying symmetrical atoms in the MOF. Using the FGC (fig. S1) functional groups are placed sequentially into the MOF. *Fapswitch* ensures that there are no steric collisions from the inserted functional groups by doing a simple conformational search. Combinations of functional groups are rejected if atoms fall within a factor of the atom's Van der Waals radius. A factor of $2^{(1/6)}$ of the VdW radius was used as this ensures that the VdW potential of any inserted atom is 0 or lower in a Lennard-Jones 12-6 potential. For each site, the functional group is inserted, aligned with the structure using the minimum energy configuration when attached to a benzene ring. All inserted atoms are tested for overlap. If there is steric overlap the group is rotated about the bond to the structure incrementally until there is no overlap. If a complete rotation is completed without finding a configuration with no overlap, that FGC is rejected. The procedure is repeated for all sites in the MOF and all codes in the FGC.
To relax the induced stresses, all MOFs had their geometries optimized with UFF as

implemented in the General Utility Lattice Program (GULP), version 4.0 (*38*). Bonding information was included in the generation of the structures and passed to the optimizer.

6. Molecular simulations

Gas adsorption calculations were performed using an in-house Grand Canonical Monte Carlo (GCMC) code based on DL_POLY 2 molecular dynamics package (*28*). Non-bonding interactions were calculated with a Lennard-Jones potential utilizing parameters for the framework atoms taken directly from the UFF (*18*) with Lorentz-Berthelot mixing rules for cross-terms. Electrostatics were based on partial atomic charges calculated by charge equilibration using the MEPO-QEq parameters (*19*), which were fit to reproduce the electrostatic potential obtained from REPEAT atomic partial charges (*29*). The CO_2 molecules were modeled using the force field developed by García-Sánchez et al. (*30*) and the N_2 molecules were modelled using the TraPPE force field parameters (*31*).

All GCMC simulations consisted of 30000 cycles of equilibration and 30000 cycles of production. One cycle consists of a N of trial moves where N is equal to the number of guest molecules in the system at that time. All simulations included random insertion, deletion, and translation moves of molecules with equal probabilities. Atoms in the framework were held fixed at their crystallographic positions. A LJ cut-off distance of 12.5 Å was used for all simulations and a supercell is constructed for each structure that satisfies the minimum image criterion. The Ideal gas law was assumed when computing the chemical potential in the grand canonical ensemble.

Geometric properties were calculated with Zeo++ (*32*) using helium probe of 1 Å to determine the solvent accessible surface areas and pore sizes.

7. Parasitic energy

The parasitic energy (P_E) is a term to describe the energy needed to remove CO_2 from a solid sorbent. P_E is a combinatorial term which has information from adsorption conditions and desorption conditions. For adsorption conditions we used flue gas conditions (298 K with 0.15

bar CO_2 and 0.75 bar N_2), while desorption condition were at increased temperature and decreased pressure (413 K with 0.70 bad CO_2 and 0.01 bar N_2).

P_E is broken into two terms, the thermal contribution (Q) and the work of compression (W_{Comp}). Q (eq. 5) contains the energy necessary to raise the temperature of the system and disrupt the host-guest interactions. It is seen that Q contains the heat capacity of the adsorbent being used. Determining the heat capacity of a material can be a computational expensive calculation to run. In this study we used a constant heat capacity for all materials of 1 kJ/kgK. This was chosen as it was average value over a range of similar solid sorbents from a previous study (*39*). W_{Comp} (eq. 6) is the energy necessary to change the pressure during desorption process as well as to compress it for transport conditions (313 K and a total of 150 bar)

$$\boldsymbol{Q} = \frac{\Delta \boldsymbol{T}\left(\boldsymbol{C_p} + \sum_i^{gas} \boldsymbol{C_i q_i^a}\right)}{\Delta \boldsymbol{q_{CO_2}}} + \frac{\sum_i^{gas} \Delta \boldsymbol{h_i^a q_i^a} - \Delta \boldsymbol{h_i^d q_i^d}}{\Delta \boldsymbol{q_{CO_2}}} \quad (5)$$

$$\boldsymbol{W_{Comp}} = \boldsymbol{R}\left\{\boldsymbol{T_{comp}}\left|\boldsymbol{ln}\left(\frac{\boldsymbol{p_c}}{\boldsymbol{p_d}}\right)\right| + \boldsymbol{T_{de}}\left|\boldsymbol{ln}\left(\frac{\boldsymbol{p_d}}{\boldsymbol{p_a}}\right)\right|\right\}\sum_{\boldsymbol{i=1}}^{\boldsymbol{gas}} \frac{\Delta \boldsymbol{q_i}}{\Delta \boldsymbol{q_{CO_2}}} \quad (6)$$

P_E (eq. 7) adds the thermal contribution and work of compression, while taking into account energy recovery from the desorption process. The terms that make up Q, W_{Comp}, and P_E are described in table S7

$$\boldsymbol{P_E} = \boldsymbol{0.75\,\eta_{T_{final}} Q + W_{comp}} \quad (7)$$

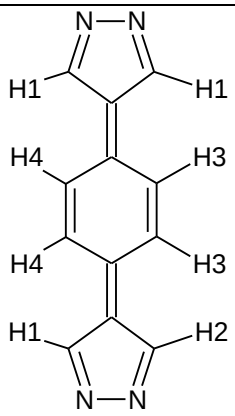
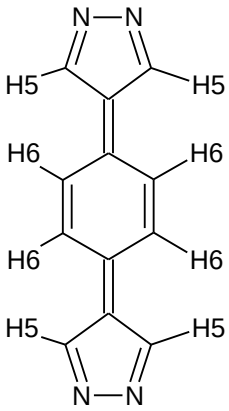
table S7. Terms used in parasitic energy with a brief description.

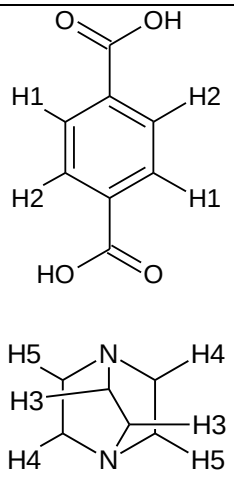
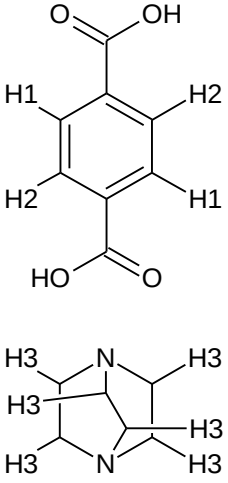
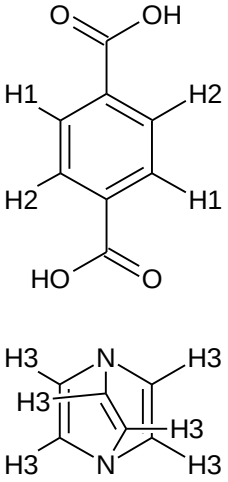
Term	Description
ΔT	Change in temperature from adsorption to desorption conditions
C_P	Heat capacity of adsorbent (1 kJ/kgK)
C_i	Heat capacity of the gas *i*
$q_i^{a(d)}$	Amount of gas *i* adsorbed at adsorption (desorption) condition
$\Delta h_i^{a(d)}$	Heat of adsorption of gas *i* as adsorption (desorption) condition
Δq_i	Working capacity of a gas *i*
$T_{C(d)}$	Temperature at compression (desorption) condition
$P_{c(d)(a)}$	Pressure at compression (desorption) (adsorption) condition
η_{TFinal}	Carnot efficiency of steam generator (0.18)

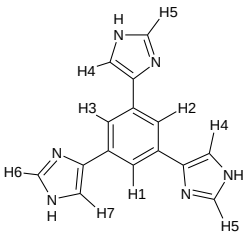
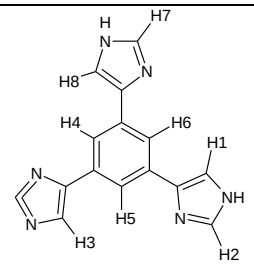
8. Top performing structures

Provided in table S8 are a list of all MOFs found CO_2 uptake greater than 3 mmol/g at flue gas conditions. Parent MOFs are the unfunctionalized base structure with their SBUs given. For each high performing functionalization the CO_2 uptake and the FGC are given. Figure S1 shows how the FGC works by changing the parent MOF's SBUs into the functionalized SBUs. For simplicity if a functional position is a hydrogen atom than it is not omitted in the FGC. The details of the functional groups, such as name and structure, are given in table S9.

table S8. Functionalized MOFs with CO_2 uptake greater than 3 mmol/g with the corresponding functional groups. A blank Functional Group Code means the unfunctionalized Parent MOF.

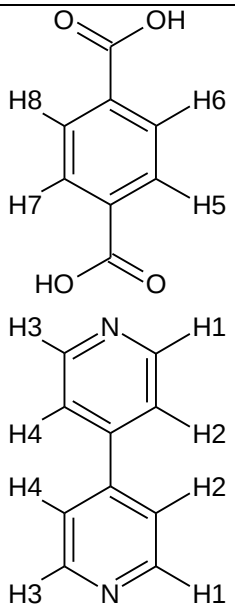
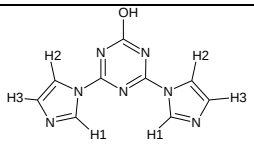
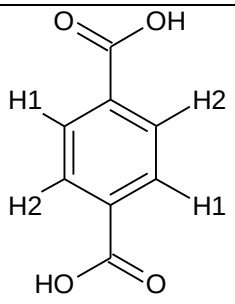
Parent MOF	Secondary Building Unit	Functional Group Code	CO_2 Uptake (g/mmol)
AFOYOK		Me@H6	4.090
		Me@H4.Cl@H6	3.790
		HCO@H4	3.783
		HCO@H4.F@H6	3.744
		HCO@H4.OH@H6	3.619
		F@H1.OH@H3.Me@H6	3.595
		Me@H4.CN@H6	3.593
		HCO@H4.OH@H5	3.582
		HCO@H6	3.542
		Me@H4	3.496
		OH@H6	3.467
		Cl@H6	3.427
		CHNH@H1.HCO@H2.Me@H5.NO2@H6	3.410
		Me@H4.OH@H5	3.395
		NH2@H1.HCO@H6	3.237
		CN@H6	3.233
		Me@H3.OH@H6	3.222
		OH@H3.Cl@H6	3.185
		Et@H6	3.181
			3.163
		F@H2.Me@H4	3.160
		F@H1.OH@H3	3.153
		F@H6	3.136
		HCO@H4.NH2@H5	3.135
		OH@H1.OH@H2.HCO@H6	3.127
		OH@H1.HCO@H6	3.090
		Cl@H4	3.060
		NH2@H4.HCO@H6	3.043

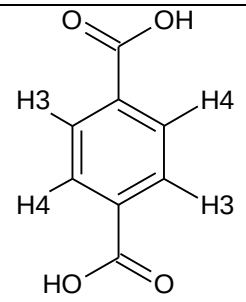
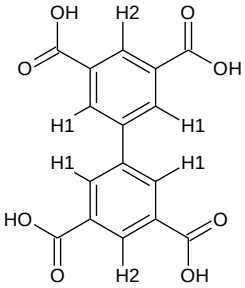
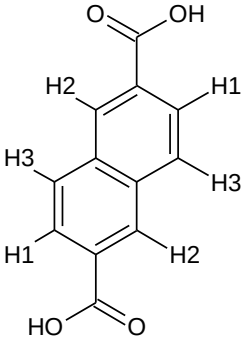
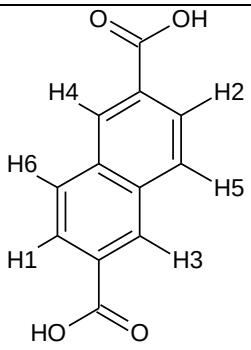
AJORAT		MeNH2@H1.OH@H2.HCO@H5	4.104
		OMe@H1.OH@H2.HCO@H5	3.695
		MeNH2@H1.F@H2.HCO@H5	3.543
		Et@H1.OH@H2.HCO@H5	3.376
		SO3H@H1.OH@H2.HCO@H5	3.364
		Me@H1.HCO@H5	3.356
		COOH@H3.HCO@H4	3.355
		NO2@H3.HCO@H4	3.229
		Me@H1.CN@H5	3.151
		Me@H1.OH@H2.HCO@H5	3.095
		Me@H1.F@H2.HCO@H5	3.053
		HCO@H3.HCO@H4	3.048
		OEt@H1.OH@H3	3.045
		NO2@H1	3.042
		OMe@H1.NH2@H5	3.028
AJORAT_3		NO2@H1	3.073
		OEt@H1	3.072
AJORATR		CHCH2@H2.CHNH@H3.CHNH@H5	3.682
		OH@H1.Et@H3.OH@H4.OH@H5	3.417
		CHNH@H1.Me@H3.OH@H4.HCO@H5	3.341
		Me@H3.HCO@H4.HCO@H5	3.326
		OMe@H1.OH@H2.Et@H3	3.307
		MeNH2@H1.HCO@H5	3.214
		SO3H@H1.F@H2.OH@H3	3.166
		OEt@H1.F@H3.OH@H5	3.102
		OMe@H1.Et@H3.NH2@H5	3.091
		CHNH@H1.Me@H3.OH@H4.OH@H5	3.061
		OMe@H1.OH@H2.NHMe@H3	3.055
		CHCH2@H2.CHNH@H3.CONH2@H5	3.054
		CONH2@H1.OMe@H3.OH@H5	3.043
		OEte@H1.OMe@H3	3.042

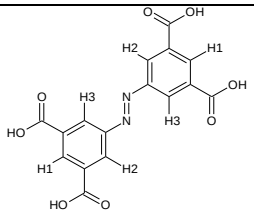
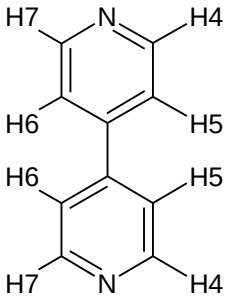
		Me@H1.HCO@H5	3.031
		CHNH@H1.HCO@H3.HCO@H5	3.031
		CHNH@H1.CHCH2@H2.CHNH@H3.CHNH@H5	3.007
		OMe@H1.MeNH2@H3.OH@H5	3.000
CoHLDMFH2O_desolvated		OPre@H1.HCO@H2.CHCH2@H3.CN@H5.F@H6	3.190
			3.188
		F@H1.COOH@H2.CHNH@H5	3.120
3.7CoHLH2O2_desolvated		Me@H1.Me@H2	3.517
		Et@H1.HCO@H3	3.455
		Me@H1.NH2@H2	3.427
		CHCH2@H1.HCO@H8	3.411
		HCO@H1.COOH@H3.HCO@H6	3.383
		Et@H1.OH@H3	3.379
		OMe@H1.NH2@H2	3.376
		HCO@H1.COOH@H3.F@H4.Me@H6	3.371
		Et@H1.NO2@H3	3.355
		OMe@H1.HCO@H8	3.351
		CHCH2@H1.Me@H2	3.346
		Me@H1	3.342
		Et@H1.HCO@H8	3.340
		Et@H1.Me@H2	3.337
		Me@H1.OH@H2	3.332
		Me@H1.HCO@H8	3.319
		CHCH2@H1.HCO@H3	3.298
		NHMe@H1.HCO@H8	3.285
		MeNH2@H1.HCO@H3	3.279
		HCO@H1.COOH@H3.OH@H4.HCO@H6	3.249
		Me@H2.HCO@H6	3.240
		MeNH2@H1.Me@H2	3.236
		CHCH2@H1	3.230
		NHMe@H1.HCO@H3	3.228
		Me@H1.Cl@H2	3.223
		OMe@H1.CHNH@H2	3.222
		Me@H1.HCO@H3	3.206
		Et@H1.NH2@H2	3.202
		HCO@H1.F@H2.COOH@H3.HCO@H6	3.199
		CHCH2@H1.NH2@H2	3.195

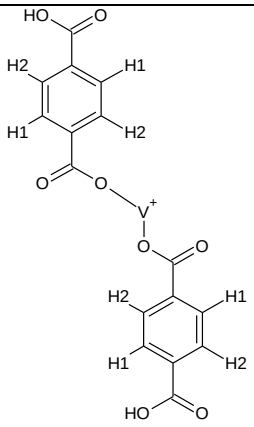
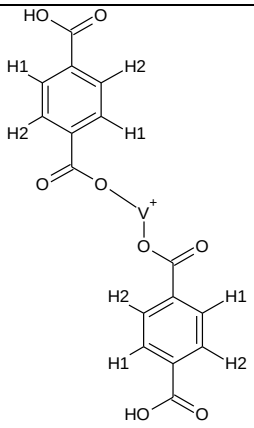
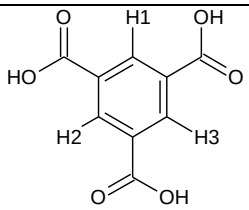
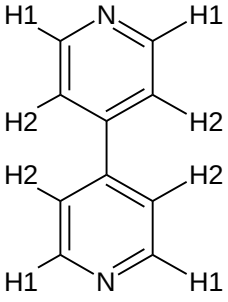
		Et@H1	3.194
		Me@H1.HCO@H2	3.193
		CCH@H1.Me@H2	3.193
		NHMe@H1.Me@H2	3.190
		HCO@H1.NH2@H2	3.171
		Me@H1.F@H2	3.166
		OMe@H1	3.162
		OH@H4.HCO@H8	3.161
		OMe@H1.F@H6	3.160
		OMe@H1.HCO@H3	3.158
		CHNH@H1.Me@H2	3.152
		HCO@H1.HCO@H3	3.151
		CONH2@H1.Me@H2	3.144
		HCO@H1.COOH@H3.F@H4.HCO@H6	3.142
		OMe@H1.Me@H2	3.140
		HCO@H1.Me@H2.F@H3.OH@H6	3.138
		HCO@H1.F@H2.COOH@H3	3.137
		MeNH2@H1.OH@H3	3.131
		CHCH2@H1.OH@H3	3.125
		NHMe@H1.OH@H3	3.123
		Et@H1.HCO@H7	3.123
		OMe@H1.NO2@H7	3.121
		OMe@H1.COOH@H7	3.115
		CHCH2@H1.HCO@H2	3.113
		CHCH2@H1.CHNH@H2	3.112
		MeNH2@H1.NH2@H2	3.111
		OMe@H1.HCO@H7	3.110
		COOH@H1.Me@H2	3.110
		CHCH2@H1.OH@H8	3.109
		MeNH2@H1.HCO@H2	3.105
		COOH@H3.F@H6	3.104
		Et@H1.NH2@H3	3.103
		OMe@H1.OH@H8	3.100
		Et@H1.HCO@H2	3.099
		OMe@H1.OH@H4	3.096
		NHMe@H1.NH2@H2	3.095
		NH2@H1.Me@H2	3.093
		OMe@H1.OH@H3	3.092
		HCO@H1.Me@H2.F@H3.F@H6	3.089
		HCO@H1.Me@H2	3.088
		HCO@H1.CN@H2.COOH@H3.OH@H4.HCO@H6	3.087

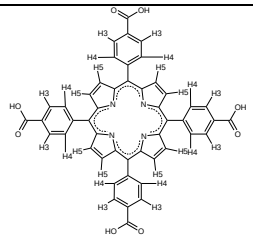
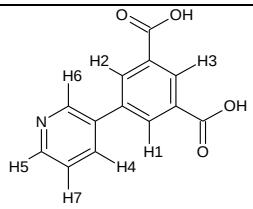
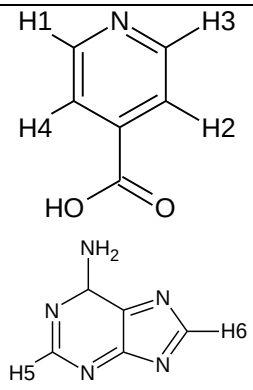
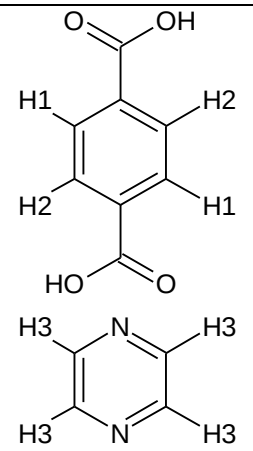
		NH2@H1.HCO@H8	3.085
		Et@H1.OH@H2	3.076
		CCH@H1.NH2@H2	3.074
		OH@H1.Me@H2	3.074
		Et@H1.NH2@H7	3.073
		OMe@H1.F@H2	3.072
		Et@H1.Me@H7	3.065
		CHNH@H1.HCO@H8	3.064
		Et@H1.OH@H4	3.063
		Et@H1.OH@H8	3.062
		Me@H1.CHNH@H2	3.062
		NHMe@H1.CHNH@H2	3.056
		CHCH2@H1.F@H6	3.053
		CCH@H1.HCO@H8	3.049
		CONH2@H1.NH2@H2	3.047
		NH2@H2	3.045
		HCO@H1.Me@H2.HCO@H4.HCO@H6.OH@H8	3.044
		Et@H1.CHNH@H2	3.041
		COOH@H1.NH2@H2	3.039
		OMe@H1.Me@H7	3.038
		Et@H1.F@H6	3.037
		CCH@H1.CHNH@H2	3.033
		CHNH@H1.HCO@H3	3.033
		NO2@H1.OH@H4	3.033
		CHCH2@H1.OH@H2	3.029
		Me@H1.Me@H2.NH2@H3.OH@H4.CHNH@H6	3.023
		Et@H1.Cl@H3	3.022
		Me@H2	3.022
		CHCH2@H1.F@H2	3.021
		Me@H1.F@H6	3.018
		Et@H1.NO2@H7	3.014
		CHNH@H1.NH2@H2	3.013
		OMe@H1.HCO@H2	3.013
		Et@H1.Cl@H2	3.011
		NHMe@H1	3.009
		NHMe@H1.HCO@H2	3.007
		Me@H1.OH@H3	3.006
		CHCH2@H1.Me@H8	3.004
		OH@H1	3.004
		NHMe@H1.OH@H8	3.004

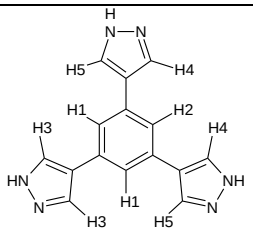
CUHPUR		CHCH2@H1.OH@H2.HCO@H3.F@H7	3.170
		CHCH2@H1.OH@H2.CHCH2@H3.F@H5.Me@H7	3.097
HECQUB		NH2@H1.CONH2@H2.MeNH2@H3	3.765
		NH2@H1.HCO@H2.MeNH2@H3	3.512
		NH2@H1.OEt@H2.HCO@H3	3.462
		NH2@H1.OEt@H2.MeNH2@H3	3.461
		NH2@H1.COOH@H2.MeNH2@H3	3.458
		NH2@H1.CHNH@H2.MeNH2@H3	3.332
		OEt@H2.HCO@H3	3.293
		NH2@H1.OEt@H2.NH2@H3	3.287
		Br@H1.CONH2@H2.MeNH2@H3	3.176
		Cl@H1.COOH@H2.MeNH2@H3	3.148
		Cl@H1.HCO@H2.MeNH2@H3	3.142
		NH2@H1.Me@H2.MeNH2@H3	3.138
		HCO@H1.CONH2@H2.CCH@H3	3.132
		HCO@H1.HCO@H2.CHCH2@H3	3.044
		NH2@H1.CONH2@H2.Me@H3	3.043
		HCO@H1.CONH2@H2.NH2@H3	3.023
HIFTOG		HCO@H1.MeNH2@H2.HCO@H3.COOH@H4	3.114

			
In_MOF		F@H2	3.087
		OH@H2	3.087
			3.069
InMOFKorea		Me@H1	3.345
		OMe@H1.Me@H3	3.103
		F@H1.CHNH@H3	3.027
InMOFKorea_6		NH2@H1.CHCH2@H2	3.627
		NH2@H2.OH@H3.HCO@H4.SO3H@H5.NH2@H6	3.423
		Me@H1	3.345
		Me@H1.HCO@H2.F@H3	3.338
		Me@H2.OH@H3.HCO@H4.SO3H@H5.NH2@H6	3.334
		OH@H1.F@H2.CN@H3.NH2@H4.OPre@H5.NO2@H6	3.302
		Me@H1.NO2@H2.NH2@H3.HCO@H4.SO3H@H5.HCO@H6	3.288
		NH2@H1.OPre@H3.OH@H4.HCO@H5.NH2@H6	3.259
		NH2@H2.Me@H3.HCO@H4.Me@H5.NH2@H6	3.245
		F@H1.OH@H2.OPre@H3.HCO@H5.NH2@H6	3.232
		F@H1.OPre@H3.HCO@H4.HCO@H5.OH@H6	3.227
		NH2@H1.OH@H2.OPre@H3.F@H4.OH@H5.Me@H6	3.225
		NH2@H1.Me@H2.NH2@H4.F@H5	3.193
		Me@H1.Me@H2.Me@H3.F@H4.OPre@H5.SO3H@	3.170

		H6	
		NH2@H1.F@H2.OPre@H3.OH@H4.HCO@H5.NH2@H6	3.170
		NH2@H1.Me@H2.OPre@H3.HCO@H4.HCO@H5.OH@H6	3.129
		NH2@H2.NH2@H4.NH2@H5.NH2@H6	3.117
		OMe@H1.Me@H3	3.103
		Me@H1.Me@H3.OH@H4.SO3H@H5	3.082
		Me@H1.OH@H2.OPre@H3.NH2@H5.Me@H6	3.062
		NO2@H2.OPre@H3.OH@H4.HCO@H5.Cl@H6	3.062
		OH@H1.Me@H2.NH2@H3.OH@H5.NH2@H6	3.038
		F@H1.CHNH@H3	3.027
		NH2@H1.OH@H2.OH@H3.Me@H4.HCO@H5.SO3H@H6	3.026
		NO2@H1.NO2@H2.NH2@H3.HCO@H4.SO3H@H5.HCO@H6	3.015
		COOH@H2.OH@H3.HCO@H4.SO3H@H5.NH2@H6	3.014
		OH@H1.Me@H2.OPre@H3.NH2@H6	3.009
		Me@H2.CONH2@H3.HCO@H4.SO3H@H5.NH2@H6	3.007
ISOHEE	 	CCH@H2.HCO@H3.Me@H4	3.550
		Me@H2.HCO@H3.Et@H4.NO2@H7	3.399
		Me@H2.CCH@H3.Et@H4	3.330
		CCH@H2.NO2@H3.Et@H4	3.300
		OMe@H2.Me@H3.HCO@H4	3.296
		HCO@H2.CCH@H3.Et@H4	3.142
		HCO@H2.HCO@H3.HCO@H4	3.129
		Me@H1.Me@H2.HCO@H3.Et@H4.NO2@H7	3.103
		CCH@H2.NH2@H3.MeNH2@H4	3.101
		CCH@H2.Me@H3.Me@H4	3.077
		HCO@H2.HCO@H3.Et@H4	3.059
		NO2@H2.Me@H3.Me@H4	3.042
MIL-47A		HCO@H1	3.568
		HCO@H2	3.520
		CHCH2@H2	3.368
		CHCH2@H1	3.297
		CONH2@H1	3.082
		CONH2@H2	3.078
		CHNH@H2	3.024
		CHNH@H1	3.023

			
MIL-47B		CHNH@H1	3.946
		COOH@H1	3.862
		HCO@H1	3.560
		CONH2@H2	3.086
MITSUE		NH2@H1.OEt@H2.NO2@H3	3.555
		Me@H1.Pr@H2.NH2@H3	3.266
		HCO@H1.OEt@H2	3.252
		NH2@H1.CCH@H2.CCH@H3	3.243
		Me@H1.Pr@H2.Me@H3	3.200
		CCH@H2.CCH@H3	3.106
		Me@H1.Pr@H2.HCO@H3	3.068
		NH2@H1.Pr@H2.Me@H3	3.007
MOYZIK		OPre@H2.Ph@H4.HCO@H5	3.001

			
NJU-Bai7			3.065
TIF-A1-desolvated-6		MeNH2@H1.OH@H2.NH2@H3.OH@H4.OPre@H6	3.132
TONXIE		SO3H@H1.OMe@H2	3.168
		COOH@H1.OMe@H2.CHNH@H3	3.044
		NH2@H1.OMe@H2.HCO@H3	3.042
UTEXAT		HCO@H2.Me@H3	4.360
		Me@H2.Me@H3	4.360
		CN@H2.Me@H3	4.270

		COOH@H2.Me@H3	4.252
		NO2@H2	4.241
		CCH@H2.Me@H3	4.229
		NO2@H2.Me@H3	4.213
		OH@H2.Me@H3	4.206
		Me@H3.OH@H4	4.177
		NH2@H2.Me@H3	4.144
		Cl@H2.Me@H3	4.102
		Me@H2.Me@H3.OH@H4	4.094
		CONH2@H2.Me@H3	4.080
		HCO@H2	4.039
		CHNH@H2.Me@H3	4.032
		OH@H3.OH@H4	4.022
		CN@H2.NH2@H3	4.009
		HCO@H2.CHNH@H3	4.005
		CHNH@H2.Me@H3.OH@H4	3.993
		NH2@H2.Me@H3.OH@H4	3.940
		CCH@H2.Me@H3.OH@H4	3.917
		Me@H2.HCO@H3	3.910
		HCO@H2.Me@H3.OH@H4	3.907
		CHCH2@H2.Me@H3	3.903
		COOH@H2.CHNH@H3	3.884
		CONH2@H2	3.869
		COOH@H2	3.868
		OH@H2.Me@H3.OH@H4	3.860
		OH@H2.OH@H3	3.842
		OMe@H2.Me@H3	3.840
		COOH@H2.OH@H3	3.837
		MeNH2@H2.Me@H3	3.833
		HCO@H2.NH2@H3	3.833
		MeNH2@H2.OH@H4	3.826
		NO2@H2.OH@H3	3.826
		Cl@H2.Me@H3.OH@H4	3.813
		CONH2@H2.NH2@H3.OH@H4	3.811
		OH@H2.HCO@H3	3.801
		Br@H2.Me@H3	3.800
		NH2@H2.NH2@H3	3.798
		Me@H3.F@H4	3.796
		F@H2.Me@H3.OH@H4	3.788
		Me@H2.NH2@H3	3.779
		NH2@H2.OH@H3	3.766

		MeNH2@H2.NH2@H3	3.766
		HCO@H2.HCO@H3	3.764
		COOH@H2.HCO@H3	3.759
		CCH@H2.NH2@H3	3.754
		OH@H4	3.738
		CONH2@H2.HCO@H3	3.713
		OMe@H2.NH2@H3	3.708
		COOH@H2.Me@H3.OH@H4	3.704
		COOH@H2.HCO@H3.F@H5	3.700
		Me@H2.HCO@H3.F@H5	3.698
		OH@H2.HCO@H4	3.687
		OH@H2	3.684
		COOH@H2.Cl@H3	3.680
		NHMe@H2.HCO@H3	3.677
		NO2@H2.Cl@H3	3.674
		HCO@H2.NH2@H3.OH@H4	3.670
		Me@H3	3.664
		Me@H2.OH@H3	3.658
		HCO@H2.Me@H3.F@H4	3.656
		CCH@H2.Me@H3.F@H4	3.655
		Cl@H2.HCO@H3	3.645
		Me@H2.Me@H3.OH@H4.F@H5	3.636
		OH@H2.OH@H4	3.634
		MeNH2@H2.HCO@H3	3.631
		NH2@H2.HCO@H3	3.631
		HCO@H2.HCO@H4	3.628
		NH2@H2.Me@H3.F@H4	3.621
		NO2@H4	3.620
		CN@H2.Cl@H3	3.616
		CONH2@H2.Me@H3.OH@H4	3.614
		HCO@H2.OH@H4	3.604
		OH@H2.NO2@H4	3.582
		NO2@H2.NH2@H3	3.564
		COOH@H2.NH2@H3.F@H4	3.564
		COOH@H2.NH2@H3	3.558
		CHCH2@H2.HCO@H3	3.558
		CN@H2.Me@H3.F@H4	3.557
		OH@H3.NH2@H4	3.556
		Me@H3.NH2@H4	3.555
		Br@H2.NH2@H3	3.546
		Cl@H2.NH2@H3	3.540

		NH2@H2.OH@H4	3.539
		HCO@H2.Cl@H3	3.539
		Me@H2.Cl@H3	3.538
		F@H2.HCO@H4	3.530
		CCH@H2.Cl@H3	3.525
		CONH2@H2.NH2@H3	3.521
		CCH@H2.Me@H3.OH@H4.F@H5	3.520
		COOH@H2.NH2@H3.F@H5	3.517
		OMe@H2.HCO@H3.F@H5	3.515
		NO2@H2.Me@H3.OH@H4	3.513
		HCO@H2.Me@H3.HCO@H4	3.510
		COOH@H2.Me@H3.F@H4	3.504
		HCO@H4	3.501
		F@H2.Me@H3.F@H4	3.496
		Et@H2.Me@H3	3.492
		CN@H2.CHNH@H3	3.486
		Me@H2.HCO@H3.OH@H4	3.478
		Cl@H2.CHNH@H3	3.475
		CHCH2@H2.NH2@H3	3.470
		CHNH@H2.NH2@H3.OH@H4	3.469
		CCH@H2.OH@H4	3.465
		CN@H2.Me@H3.OH@H4	3.462
		OH@H2.HCO@H3.F@H5	3.461
		OEt@H2	3.457
		NH2@H2.HCO@H3.F@H5	3.455
		Pr@H2.Me@H3	3.449
		COOH@H2.OH@H5	3.447
		Et@H2.HCO@H3	3.445
		MeNH2@H3	3.435
		F@H2.NH2@H3	3.433
		Me@H3.HCO@H4	3.432
		CCH@H2.HCO@H3.F@H5	3.430
		MeNH2@H2	3.430
		HCO@H3.F@H5	3.427
		OMe@H2	3.425
		OH@H2.Me@H3.F@H4	3.424
		NO2@H2.Me@H3.F@H4	3.417
		OEte@H2.Me@H3	3.416
		OMe@H2.Cl@H3	3.415
		CN@H2.OH@H3	3.413
		NH2@H2.CHNH@H3	3.410

		CHCH2@H2.NH2@H4	3.405
		CHNH@H2.HCO@H3	3.404
		MeNH2@H2.Cl@H3	3.402
		NH2@H2.Cl@H3	3.400
		NHMe@H2.Cl@H3	3.398
		Me@H2.NH2@H3.OH@H4	3.397
		NO2@H2.OH@H5	3.396
		F@H2.OH@H4	3.391
		CN@H2	3.381
		CCH@H2	3.381
		OEt@H2.Me@H3	3.379
		Br@H2.HCO@H3	3.367
		Me@H2.OH@H4.Me@H5	3.365
		NH2@H3.HCO@H4	3.363
		Me@H3.OH@H4.F@H5	3.359
		CHNH@H2.Cl@H3	3.357
		CHNH@H2.HCO@H3.OH@H4	3.351
		CHNH@H2.OH@H4	3.350
		F@H2.OH@H3.OH@H4	3.349
		F@H2.Me@H3	3.349
		NH2@H2.HCO@H4	3.344
		CN@H2.Me@H3.OH@H4.F@H5	3.343
		COOH@H2.OH@H4	3.342
		OH@H2.Me@H3.HCO@H4	3.340
		Et@H2.Cl@H3	3.337
		CCH@H2.Me@H3.F@H5	3.337
		Me@H2.OH@H3.OH@H5	3.337
		CCH@H2.Cl@H3.OH@H4	3.336
		Me@H2.CHNH@H3	3.327
		NH2@H4	3.326
		CHNH@H2.NH2@H3.NH2@H4	3.324
		Cl@H2.Cl@H3	3.323
		NH2@H2.Me@H3.NH2@H4	3.322
		Me@H4	3.322
		CHNH@H2.Me@H3.F@H4	3.320
		Et@H2.OH@H3	3.320
		NH2@H2.Me@H3.F@H5	3.319
		OH@H3.HCO@H4	3.318
		OEt@H2.OH@H4	3.312
		NO2@H2.Me@H3.F@H5	3.312
		Cl@H2.Me@H3.F@H4	3.312

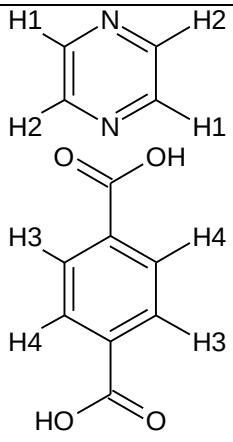
		HCO@H2.NH2@H4	3.309
		Me@H2.Me@H3.F@H5	3.308
		COOH@H2.Me@H3.OH@H4.F@H5	3.308
		OEt@H2.NH2@H3	3.306
		NO2@H2.HCO@H3	3.306
		MeNH2@H2.HCO@H3.F@H5	3.305
		CHNH@H2.HCO@H4	3.304
		HCO@H3	3.303
		CCH@H2.NH2@H3.OH@H4	3.303
		HCO@H3.OH@H4	3.302
		OH@H2.Me@H3.NH2@H4	3.300
		Br@H2.Me@H3.OH@H4	3.300
		F@H2.HCO@H3	3.296
		Et@H2.NH2@H3	3.292
		OH@H2.Cl@H3	3.290
		NH2@H2.HCO@H3.OH@H4	3.287
		OH@H2.Me@H4	3.283
		NH2@H2.Cl@H3.OH@H4	3.276
		COOH@H2.Me@H3.HCO@H4	3.274
		HCO@H2.OH@H5	3.271
		NH2@H2.NH2@H3.F@H5	3.269
		Pr@H2	3.268
		CHCH2@H2	3.267
		OEt@H2.HCO@H3	3.266
		OH@H2.NH2@H4	3.262
		OMe@H3.OH@H4	3.261
		HCO@H2.Me@H3.F@H5	3.259
		HCO@H2.CHNH@H3.OH@H4	3.258
		COOH@H2.Me@H3.F@H5	3.257
		NH2@H2.Me@H3.HCO@H4.OH@H5	3.256
		I@H2.Me@H3	3.256
		CCH@H3.OH@H4	3.252
		HCO@H4.NH2@H5	3.252
		Cl@H2.Me@H3.NH2@H4	3.249
		Cl@H2.OH@H4	3.245
		CCH@H2.Me@H3.NH2@H4	3.242
		CN@H2.OH@H4	3.241
		NO2@H2.NO2@H3	3.238
		CN@H2.Me@H3.F@H5	3.235
		NH2@H2.OH@H5	3.234
		CF3@H2.Me@H3	3.234

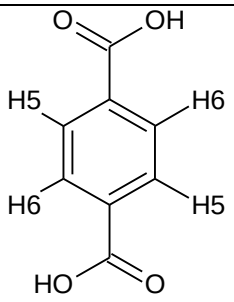
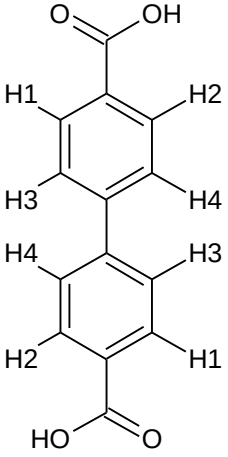
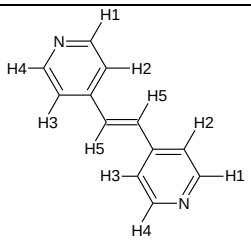
		NO2@H2.F@H5	3.233
		Cl@H2.OH@H3	3.233
		CHNH@H2.NH2@H3	3.228
		COOH@H2.NH2@H3.OH@H4	3.225
		HCO@H2.Me@H3.OH@H4.F@H5	3.224
		MeNH2@H2.OH@H3	3.222
		Me@H2.Cl@H3.OH@H4	3.222
		OH@H4.MeNH2@H5	3.221
		CCH@H3	3.217
		HCO@H2.Cl@H3.OH@H4	3.217
		SO3H@H2.Me@H3	3.216
		Me@H2.NH2@H3.F@H4	3.215
		OH@H3	3.213
		Cl@H2.HCO@H3.F@H5	3.213
		NH2@H2.Me@H3.OH@H4.F@H5	3.211
		OH@H3.F@H4	3.207
		CCH@H2.CHNH@H3.OH@H4	3.204
		OMe@H2.Me@H3.F@H5	3.204
		Pr@H2.OH@H3	3.204
		COOH@H2.Et@H3.OH@H4.OH@H5	3.204
		CN@H2.NH2@H3.F@H5	3.204
		Me@H2.Me@H3.HCO@H4.F@H5	3.203
		Me@H2	3.202
		OH@H1.Cl@H2.CCH@H3.Me@H5	3.197
		COOH@H2.NH2@H4	3.197
		OH@H2.HCO@H4.CN@H5	3.194
		CONH2@H2.Me@H3.F@H4	3.190
		OEte@H2	3.190
		NO2@H2.OH@H4	3.189
		NH2@H3	3.186
		Me@H2.Me@H3.HCO@H4	3.186
		NH2@H2.HCO@H3.NH2@H4	3.185
		F@H2.Me@H3.NH2@H4	3.184
		OH@H3.NO2@H4	3.183
		CHNH@H2.CHNH@H3	3.183
		NH2@H3.F@H4	3.183
		NHMe@H2.Me@H3	3.182
		OH@H2.NH2@H3.F@H5	3.180
		NO2@H2.CCH@H3	3.179
		Et@H2	3.176
			3.176

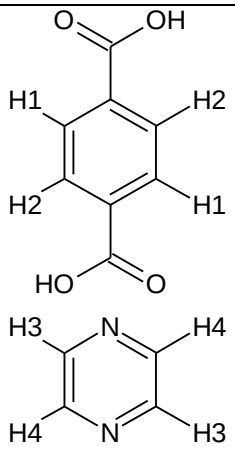
		CONH2@H2.Me@H3.NH2@H4	3.176
		NHMe@H2.HCO@H3.F@H5	3.175
		CHCH2@H2.CHNH@H3	3.170
		OH@H2.CCH@H3.OH@H4	3.169
		OEte@H2.HCO@H3	3.168
		CONH2@H2.CHNH@H3	3.167
		Me@H2.OH@H4.CN@H5	3.164
		CCH@H2.Me@H3.OH@H4.OH@H5	3.162
		HCO@H2.Me@H3.NH2@H4	3.162
		CONH2@H2.Me@H3.OH@H4.F@H5	3.162
		Me@H2.OH@H3.OH@H4	3.161
		OH@H2.NO2@H4.OH@H5	3.159
		Cl@H2.Et@H3.OH@H4	3.159
		NH2@H2.Me@H3.HCO@H4	3.158
		CN@H2.Me@H3.NH2@H4	3.156
		CCH@H2.CHNH@H3	3.155
		COOH@H2.Me@H3.NH2@H4	3.154
		F@H2.OMe@H3.OH@H4	3.154
		CONH2@H2.HCO@H3.F@H5	3.154
		CCH@H2.OH@H4.NH2@H5	3.152
		Pr@H2.NH2@H3	3.152
		NH2@H2.HCO@H3.OH@H5	3.148
		NH2@H2.OH@H3.OH@H4	3.147
		CHCH2@H2.Cl@H3	3.147
		NH2@H2.MeNH2@H3	3.146
		Br@H2.Cl@H3	3.146
		I@H2.HCO@H3	3.145
		NHMe@H2	3.144
		HCO@H2.F@H4	3.142
		CN@H2.HCO@H3.OH@H4.F@H5	3.141
		HCO@H2.F@H5	3.139
		OH@H2.CCH@H3	3.136
		CHNH@H2.Me@H3.OH@H4.NH2@H5	3.134
		COOH@H2.HCO@H3.NH2@H4	3.132
		OMe@H2.CHNH@H3	3.132
		OH@H2.OH@H3.NH2@H4	3.129
		CHNH@H3.OH@H4	3.128
		CCH@H2.HCO@H3	3.128
		OMe@H2.HCO@H3	3.127
		Et@H2.HCO@H3.F@H5	3.126
		SO3H@H2.NH2@H3	3.125

		Cl@H2.Me@H3.OH@H4.F@H5	3.125
		HCO@H2.OH@H3.NH2@H4	3.125
		Me@H2.OH@H3.F@H5	3.122
		Me@H2.NO2@H3	3.118
		CCH@H2.Me@H3.Me@H4	3.114
		OEte@H2.NH2@H3	3.113
		Et@H2.OH@H3.OH@H4.Et@H5	3.113
		CHNH@H2.Me@H3.NH2@H4	3.111
		CHNH@H3	3.110
		NO2@H2.NH2@H4	3.109
		OMe@H2.OH@H3	3.109
		CN@H2.HCO@H3.F@H5	3.105
		Me@H2.HCO@H4	3.105
		CN@H2.NH2@H3.OH@H4	3.104
		HCO@H2.OH@H3.F@H5	3.103
		CONH2@H2.HCO@H3.OH@H4.F@H5	3.101
		CONH2@H2.OH@H5	3.100
		CN@H2.OH@H3.NH2@H4	3.100
		CF3@H2.HCO@H3	3.098
		Me@H2.Me@H3.OH@H4.OH@H5	3.098
		Me@H2.NH2@H4	3.096
		MeNH2@H2.NO2@H3	3.095
		Me@H2.OH@H4.OH@H5	3.095
		CONH2@H1.NH2@H2.NH2@H5	3.093
		HCO@H2.NO2@H3	3.090
		NO2@H2.NH2@H5	3.089
		NO2@H2.OH@H3.NH2@H4	3.088
		CCH@H2.NH2@H3.F@H4	3.087
		OH@H2.Me@H3.F@H5	3.086
		Me@H2.OH@H5	3.085
		COOH@H2.HCO@H3.F@H4	3.085
		Et@H2.NH2@H3.OH@H5	3.085
		OH@H3.OH@H4.NH2@H5	3.084
		MeNH2@H2.CHNH@H3	3.084
		COOH@H2.CHNH@H3.OH@H4	3.084
		OH@H2.Cl@H3.OH@H4	3.074
		NH2@H3.MeNH2@H4	3.073
		NH2@H2.MeNH2@H3.OH@H4	3.072
		CONH2@H2.Cl@H3.OH@H4	3.071
		F@H2.CCH@H3.OH@H4	3.069
		NH2@H2.NH2@H4	3.066

		Cl@H2.NH2@H3.OH@H4	3.064
		OH@H2.MeNH2@H3.OH@H4	3.064
		CONH2@H2.Me@H3.F@H5	3.064
		Me@H2.HCO@H3.OH@H4.F@H5	3.064
		Me@H2.Me@H3.OH@H4.CONH2@H5	3.063
		OH@H1.OH@H2.OH@H3.Me@H5	3.063
		Cl@H2	3.062
		CCH@H2.NH2@H3.HCO@H4	3.061
		HCO@H2.OH@H3.OH@H4	3.060
		F@H2.Cl@H3	3.060
		Me@H4.OH@H5	3.059
		COOH@H2.F@H4	3.058
		Cl@H3	3.057
		NHMe@H2.OH@H4	3.057
		Me@H2.OH@H4	3.056
		HCO@H2.CCH@H3	3.056
		CCH@H2.HCO@H3.OH@H4	3.055
		OH@H2.Me@H3.OH@H4.F@H5	3.055
		NH2@H2.HCO@H4.CCH@H5	3.053
		OH@H1.CN@H3.NH2@H5	3.053
		OH@H2.MeNH2@H3	3.050
		MeNH2@H2.OH@H5	3.047
		Cl@H2.OH@H4.CN@H5	3.047
		OH@H2.HCO@H3.OH@H4	3.045
		F@H2.OH@H3	3.045
		OH@H2.NH2@H3.OH@H4	3.043
		OH@H2.Cl@H4	3.043
		CN@H2.NH2@H4	3.043
		OEt@H2.F@H4	3.041
		F@H2.Me@H3.OH@H4.F@H5	3.038
		COOH@H2.OH@H3.NH2@H4	3.037
		F@H2.HCO@H3.OH@H4	3.031
		CHCH2@H2.OH@H4	3.031
		HCO@H2.CCH@H3.OH@H4	3.028
		NHMe@H2.Me@H3.F@H5	3.026
		CONH2@H2.OH@H3.NH2@H4	3.024
		CHNH@H2.Cl@H3.OH@H4	3.024
		OEt@H2.OH@H3	3.023
		Br@H2.HCO@H3.F@H5	3.021
		HCO@H2.NH2@H3.OH@H4.F@H5	3.020
		CHNH@H2.OH@H3	3.020

		CN@H2.CHNH@H3.OH@H4	3.020
		Et@H2.CCH@H3.OH@H4	3.019
		NO2@H2.Cl@H3.F@H4	3.019
		F@H2.NH2@H3.OH@H4	3.017
		Et@H2.CHNH@H3	3.016
		Cl@H2.CHNH@H3.OH@H4	3.016
		NHMe@H2.OH@H5	3.016
		HCO@H2.NH2@H3.HCO@H4	3.016
		OH@H1.OMe@H5	3.015
		Me@H2.CHNH@H3.OH@H4	3.014
		NO2@H2.Me@H5	3.012
		HCO@H2.OH@H3.HCO@H4	3.011
		Me@H2.Me@H3.NH2@H4	3.010
		HCO@H2.HCO@H3.NH2@H4	3.009
		CHNH@H2.OH@H3.F@H5	3.007
		NH2@H2.HCO@H4.NH2@H5	3.006
		Cl@H2.HCO@H3.OH@H4.F@H5	3.006
		NO2@H2.Me@H3.OH@H4.F@H5	3.005
		Me@H2.Me@H3.NHMe@H5	3.005
		COOH@H2.OH@H3.F@H5	3.005
		HCO@H3.NH2@H4	3.005
		HCO@H2.OH@H4.OH@H5	3.004
		F@H2.CHNH@H3	3.003
		CHNH@H2.HCO@H3.F@H5	3.002
		COOH@H2.NH2@H3.OH@H4.F@H5	3.001
		NH2@H2.NH2@H3.HCO@H4	3.001
		CONH2@H2.NH2@H3.HCO@H4	3.001
		NHMe@H2.CHNH@H3	3.001
		Me@H2.HCO@H3.F@H4	3.000
WAFKEU02		NH2@H1.CHCH2@H2.NO2@H4.HCO@H5	3.504
		OH@H2.SO3H@H4.Pr@H5.HCO@H6	3.497
		NH2@H1.CHCH2@H2.NO2@H4.OMe@H5	3.338
		NH2@H1.CHCH2@H2.CONH2@H4.HCO@H5.F@H6	3.160
		NHMe@H1.OH@H3.SO3H@H4.CCH@H5.HCO@H6	3.102
		CCH@H1.HCO@H2.HCO@H3.HCO@H5.CHCH2@H6	3.075
		Me@H1.CHCH2@H2.NO2@H4.HCO@H6	3.069
		OMe@H1.OH@H3.SO3H@H4.CCH@H5.HCO@H6	3.026

			
WUJFOX		Me@H4	3.005
XACYAB		HCO@H3.CHNH@H5	3.607
		NO2@H3.CHNH@H5	3.533
		CHNH@H1.NO2@H3.F@H5	3.484
		OH@H2.CONH2@H5	3.475
		OH@H2.F@H3.CCH@H5	3.231
		NO2@H3.NO2@H5	3.210
		OEte@H2.CN@H3	3.206
		HCO@H3.CCH@H5	3.181
		F@H1.OH@H2	3.175
		NH2@H1.HCO@H3.CHNH@H5	3.174
		HCO@H3.NO2@H5	3.123
		NO2@H3.NH2@H5	3.122
		OEte@H2.NO2@H5	3.118
		OH@H1	3.092
		OEte@H2.Cl@H3.HCO@H5	3.050
		OMe@H2.HCO@H5	3.036
		NO2@H1.NH2@H3.CHNH@H5	3.031
		OH@H1.NO2@H3	3.009

ZBP		HCO@H1.HCO@H3.CHCH2@H4	4.238
		NO2@H1.Me@H3.HCO@H4	4.213
		Me@H2.HCO@H3.NO2@H4	4.188
		NO2@H1.HCO@H3.NHMe@H4	4.025
		HCO@H1.CHCH2@H3.HCO@H4	4.000
		Me@H2.HCO@H3.HCO@H4	3.961
		NO2@H1.Me@H4	3.934
		NO2@H1.NH2@H3.Me@H4	3.923
		NO2@H1.Me@H3.NH2@H4	3.921
		NO2@H1.Me@H3.MeNH2@H4	3.918
		NO2@H1.OH@H3.HCO@H4	3.914
		NO2@H1.HCO@H3.CHCH2@H4	3.900
		Me@H2.CONH2@H3.HCO@H4	3.859
		Me@H2.NO2@H3.NO2@H4	3.842
		NO2@H1.Me@H3.CHNH@H4	3.829
		HCO@H1.NH2@H3.Me@H4	3.820
		NO2@H1.NH2@H3.NHMe@H4	3.797
		NH2@H2.OMe@H3.OMe@H4	3.775
		NO2@H1.HCO@H3.HCO@H4	3.718
		NO2@H1.CN@H3.CHCH2@H4	3.710
		Me@H2.NHMe@H3.HCO@H4	3.699
		NO2@H1.HCO@H3.NH2@H4	3.691
		Me@H2.HCO@H3.CONH2@H4	3.675
		NO2@H1.HCO@H3	3.669
		NO2@H1.HCO@H3.OMe@H4	3.669
		Pr@H4	3.641
		CCH@H2.HCO@H3.OMe@H4	3.634
		HCO@H1.HCO@H3.HCO@H4	3.633
		CHCH2@H1.NO2@H3.CHCH2@H4	3.619
		NO2@H1.NHMe@H3.OMe@H4	3.618
		Me@H2.CONH2@H3.NO2@H4	3.617
		NHMe@H1.OH@H2.MeNH2@H3.CN@H4	3.609
		NO2@H1.NH2@H3.CHCH2@H4	3.604
		MeNH2@H1.HCO@H3.Me@H4	3.603
		NO2@H1.CHCH2@H3.CHCH2@H4	3.591
		CHCH2@H1.OH@H2.NO2@H3.COOH@H4	3.588
		HCO@H1.NO2@H3.HCO@H4	3.584
		SO3H@H3.NO2@H4	3.582
		OH@H1.SO3H@H3.HCO@H4	3.579
		NH2@H1.OEte@H2.CN@H3.OH@H4	3.576
		NH2@H1.OH@H2.Et@H3.HCO@H4	3.560

		HCO@H1.COOH@H3.HCO@H4	3.554
		HCO@H1.Me@H3.HCO@H4	3.550
		CHCH2@H1.CHNH@H3.HCO@H4	3.549
		NHMe@H1.OH@H2.CCH@H3.Me@H4	3.548
		NO2@H1.HCO@H3.MeNH2@H4	3.544
		HCO@H1.HCO@H3.CHNH@H4	3.544
		NO2@H1.Cl@H3.CHCH2@H4	3.539
		HCO@H1.HCO@H3.OMe@H4	3.528
		HCO@H3.CHCH2@H4	3.519
		NO2@H3.SO3H@H4	3.518
		Me@H2.COOH@H3.OMe@H4	3.518
		NO2@H1.CCH@H3.OH@H4	3.502
		NO2@H1.MeNH2@H3.OH@H4	3.500
		CHCH2@H2.SO3H@H3.MeNH2@H4	3.499
		HCO@H1.Me@H4	3.495
		CCH@H2.OMe@H3.HCO@H4	3.493
		OH@H1.HCO@H3.SO3H@H4	3.492
		HCO@H1.NH2@H3.HCO@H4	3.491
		NO2@H1.OH@H3.NH2@H4	3.491
		NO2@H1.CN@H3	3.484
		NO2@H1.Me@H3.Cl@H4	3.478
		HCO@H1.Me@H3.CN@H4	3.477
		NO2@H1.OMe@H3.HCO@H4	3.477
		NO2@H1.NH2@H3.CHNH@H4	3.477
		Me@H2.OMe@H3.OMe@H4	3.476
		HCO@H1.CHCH2@H3.NO2@H4	3.475
		NHMe@H1.OH@H2.Cl@H3.Cl@H4	3.473
		HCO@H1.OH@H3.OMe@H4	3.462
		CHCH2@H2.MeNH2@H3.SO3H@H4	3.461
		NO2@H1.Cl@H3.OMe@H4	3.456
		HCO@H1.CHNH@H3.HCO@H4	3.455
		Me@H2.NO2@H3.CHCH2@H4	3.436
		CHCH2@H2.SO3H@H3.OH@H4	3.436
		NHMe@H1.OH@H2.Me@H3.Cl@H4	3.431
		NO2@H1.HCO@H3.OH@H4	3.430
		OMe@H1.HCO@H3.HCO@H4	3.428
		HCO@H1.Me@H3.Cl@H4	3.427
		HCO@H1.OH@H3.HCO@H4	3.411
		NH2@H2.COOH@H3.OMe@H4	3.411
		CHCH2@H3.HCO@H4	3.409
		HCO@H1.OMe@H2.CONH2@H3.HCO@H4	3.409

		HCO@H1.Et@H3.HCO@H4	3.408
		OMe@H1.CHNH@H3.CHNH@H4	3.407
		CHCH2@H2.NH2@H3.SO3H@H4	3.397
		CHCH2@H2.SO3H@H3.CHCH2@H4	3.396
		NO2@H1.F@H3.Me@H4	3.394
		NO2@H1.NH2@H2.CONH2@H3.NH2@H4	3.393
		NO2@H1.Me@H2.NHMe@H3.Cl@H4	3.388
		NO2@H1.NH2@H2.CONH2@H3.Me@H4	3.385
		CHCH2@H2.SO3H@H3.NH2@H4	3.379
		HCO@H1.Pr@H3.F@H4	3.376
		OH@H2.OEt@H3.OMe@H4	3.372
		NO2@H1.CCH@H3.CHNH@H4	3.363
		NO2@H1.CCH@H2.HCO@H3	3.356
		NO2@H1.CHCH2@H3.Me@H4	3.353
		MeNH2@H1.CHNH@H3.Me@H4	3.350
		NO2@H1.NH2@H2.Me@H3.CONH2@H4	3.349
		NO2@H1.F@H3.CHCH2@H4	3.345
		MeNH2@H1.OH@H2.Cl@H3.Me@H4	3.344
		HCO@H3.SO3H@H4	3.339
		NO2@H1.NH2@H2.Et@H3.Me@H4	3.338
		HCO@H1.CHNH@H3.CHNH@H4	3.334
		Me@H2.NO2@H3.Cl@H4	3.331
		CHCH2@H2.OH@H3.SO3H@H4	3.328
		NO2@H1.Me@H3.CN@H4	3.327
		HCO@H1.MeNH2@H3.HCO@H4	3.325
		OH@H2.OMe@H3.OEt@H4	3.323
		NO2@H1.OMe@H3.CHCH2@H4	3.322
		NHMe@H2.SO3H@H4	3.321
		CHCH2@H1.CHNH@H3.NO2@H4	3.319
		Me@H2.NH2@H3.NO2@H4	3.318
		SO3H@H3.HCO@H4	3.317
		MeNH2@H1.NHMe@H3.HCO@H4	3.316
		CHCH2@H1.SO3H@H3.SO3H@H4	3.314
		NO2@H1.CHCH2@H3.HCO@H4	3.309
		HCO@H1.OH@H2.NHMe@H3.HCO@H4	3.304
		HCO@H1.CHNH@H3.Cl@H4	3.297
		NO2@H1.CCH@H3.NH2@H4	3.297
		MeNH2@H1.OH@H2.Et@H3.OH@H4	3.294
		OMe@H1.OH@H2.CHNH@H3.CHNH@H4	3.294
		HCO@H1.OEte@H3.OH@H4	3.291
		CHCH2@H1.COOH@H2.HCO@H4	3.284

		NO2@H1.CHCH2@H3.NH2@H4	3.282
		NHMe@H1.OH@H2.Et@H3.NH2@H4	3.281
		NO2@H1.Et@H3.CHCH2@H4	3.280
		SO3H@H3.COOH@H4	3.277
		HCO@H1.CN@H3.HCO@H4	3.276
		NO2@H1.CHCH2@H3.OH@H4	3.276
		OMe@H1.OH@H2.HCO@H3.CHNH@H4	3.275
		Me@H2.HCO@H3.CHNH@H4	3.274
		OMe@H1.HCO@H3.COOH@H4	3.273
		HCO@H1.OH@H2.Et@H3.HCO@H4	3.271
		NO2@H1.CCH@H3	3.270
		HCO@H1.Me@H3.NHMe@H4	3.268
		NH2@H1.OEte@H2.CN@H3.Cl@H4	3.266
		MeNH2@H1.OH@H2.CONH2@H3.CHNH@H4	3.262
		HCO@H3.Et@H4	3.261
		HCO@H1.F@H3.Me@H4	3.261
		OH@H1.OEt@H2.HCO@H4	3.258
		NH2@H1.OH@H2.CN@H3.Et@H4	3.254
		NO2@H1.OH@H3.CHNH@H4	3.251
		NO2@H2.NO2@H3.CHCH2@H4	3.251
		CHCH2@H2.SO3H@H3.NHMe@H4	3.250
		NHMe@H2.HCO@H3.HCO@H4	3.246
		CHCH2@H1.HCO@H3.CHNH@H4	3.246
		NO2@H1.NH2@H2.NHMe@H3.Me@H4	3.242
		MeNH2@H1.CN@H2.CHNH@H3.CHNH@H4	3.239
		HCO@H1.NH2@H3.CHNH@H4	3.232
		HCO@H1.NO2@H3.NHMe@H4	3.230
		MeNH2@H1.OH@H2.Me@H3.HCO@H4	3.229
		OH@H1.OMe@H3.NO2@H4	3.225
		CHCH2@H2.NO2@H3.SO3H@H4	3.223
		NH2@H1.Et@H3.HCO@H4	3.223
		CHCH2@H2.CHCH2@H3.SO3H@H4	3.218
		HCO@H1.OH@H2.COOH@H3.COOH@H4	3.217
		HCO@H1.Cl@H3.HCO@H4	3.215
		NO2@H1.CN@H3.Me@H4	3.213
		NH2@H2.HCO@H3.HCO@H4	3.213
		HCO@H1.OH@H2.OMe@H3	3.212
		NO2@H1.CCH@H3.Me@H4	3.209
		NO2@H1.Me@H3.CCH@H4	3.209
		NHMe@H2.Et@H3.HCO@H4	3.207
		HCO@H1.CHNH@H3.NH2@H4	3.206

		HCO@H1.CCH@H2.HCO@H3.NH2@H4	3.206
		HCO@H1.HCO@H3.NO2@H4	3.205
		CHCH2@H1.OH@H2.CHNH@H3.CHCH2@H4	3.202
		CHCH2@H2.SO3H@H3.Cl@H4	3.202
		HCO@H1.Et@H3.CHNH@H4	3.197
		NO2@H1.CCH@H3.F@H4	3.197
		MeNH2@H1.HCO@H3.CHCH2@H4	3.197
		NO2@H1.Me@H3.Me@H4	3.196
		NO2@H1.CHCH2@H3.NHMe@H4	3.193
		CHCH2@H1.NO2@H4	3.192
		HCO@H1.HCO@H3.NH2@H4	3.188
		Me@H2.NH2@H3.HCO@H4	3.182
		MeNH2@H1.OH@H2.COOH@H3.CHNH@H4	3.181
		NO2@H1.OEte@H4	3.180
		NO2@H1.CN@H3.F@H4	3.179
		SO3H@H3.CN@H4	3.179
		NHMe@H1.OH@H2.Et@H3.CN@H4	3.177
		OEte@H2.Pr@H4	3.176
		NO2@H1.Me@H2.OMe@H3.Cl@H4	3.176
		NO2@H1.OH@H2.Me@H3.Me@H4	3.174
		NH2@H2.NO2@H3.HCO@H4	3.174
		MeNH2@H1.OH@H2.Me@H3.Cl@H4	3.174
		HCO@H1.CCH@H2.HCO@H3.HCO@H4	3.173
		MeNH2@H1.NH2@H2.CONH2@H3.CHNH@H4	3.170
		CHCH2@H2.HCO@H3.HCO@H4	3.169
		OMe@H1.NO2@H3.CHNH@H4	3.168
		NO2@H1.HCO@H3.Cl@H4	3.168
		NO2@H2.NH2@H3.CHNH@H4	3.164
		CHCH2@H1.OH@H2.HCO@H3.CHNH@H4	3.163
		MeNH2@H1.NH2@H2.CHNH@H3.CHNH@H4	3.159
		NHMe@H2.COOH@H3.Et@H4	3.154
		HCO@H1.OH@H2.CCH@H3	3.153
		MeNH2@H1.Me@H3.CHCH2@H4	3.152
		NO2@H1.OH@H3.OH@H4	3.150
		OMe@H1.NHMe@H3.HCO@H4	3.150
		OH@H1.HCO@H3.CHCH2@H4	3.148
		Me@H2.NO2@H3.OMe@H4	3.147
		Me@H2.OH@H3.HCO@H4	3.146
		MeNH2@H1.OH@H2.HCO@H3.CN@H4	3.145
		CHCH2@H2.NO2@H3.MeNH2@H4	3.143
		HCO@H1.OMe@H3.HCO@H4	3.142

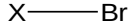
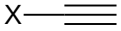
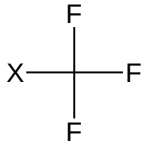
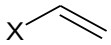
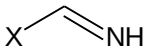
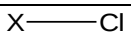
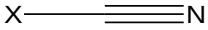
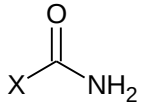
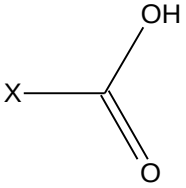
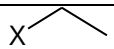
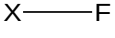
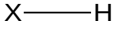
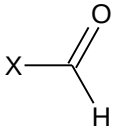
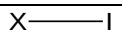
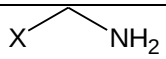
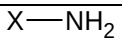
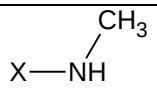
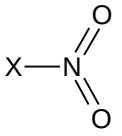
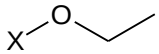
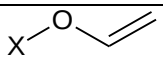
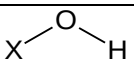
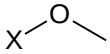
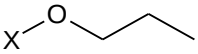
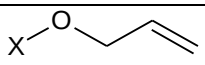
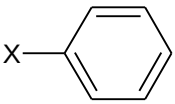
		NH2@H2.CONH2@H3.NO2@H4	3.141
		CHCH2@H2.NHMe@H3.SO3H@H4	3.141
		Me@H2.CONH2@H3.COOH@H4	3.140
		Me@H3.HCO@H4	3.139
		HCO@H1.HCO@H3.NHMe@H4	3.138
		NO2@H2.NH2@H3.CONH2@H4	3.137
		HCO@H1.SO3H@H3.F@H4	3.135
		NO2@H1.CN@H3.NH2@H4	3.131
		NO2@H2.CONH2@H3.NH2@H4	3.131
		NO2@H1.OH@H2.SO3H@H3.Me@H4	3.130
		CHCH2@H1.OH@H2.HCO@H3.CCH@H4	3.127
		MeNH2@H1.NH2@H2.HCO@H3.CHNH@H4	3.126
		NHMe@H1.OH@H2.MeNH2@H3.Cl@H4	3.125
		NO2@H1.Me@H3.Br@H4	3.124
		NO2@H1.OH@H4	3.123
		Me@H2.OMe@H3.NO2@H4	3.123
		OH@H2.SO3H@H3.CCH@H4	3.122
		HCO@H1.CN@H3.OMe@H4	3.120
		HCO@H1.Cl@H3.Me@H4	3.118
		NH2@H2.OMe@H3.OPr@H4	3.112
		NO2@H1.NH2@H2.CHNH@H3.CHCH2@H4	3.111
		NO2@H1.COOH@H3	3.109
		NHMe@H1.OH@H2.Et@H3.Cl@H4	3.109
		NHMe@H1.HCO@H3.NHMe@H4	3.109
		NH2@H2.OMe@H3.COOH@H4	3.104
		OMe@H1.CHNH@H3.Et@H4	3.104
		NH2@H1.OEt@H2.COOH@H3.F@H4	3.104
		NO2@H1.SO3H@H3.HCO@H4	3.104
		NO2@H2.HCO@H3.CHCH2@H4	3.104
		HCO@H1.CCH@H3.OMe@H4	3.101
		HCO@H1.F@H2.CHCH2@H3.COOH@H4	3.100
		HCO@H1.CCH@H2.NHMe@H3	3.100
		CHCH2@H2.SO3H@H4	3.098
		OMe@H1.CN@H2.CHNH@H3.HCO@H4	3.097
		HCO@H1.Et@H3.CHCH2@H4	3.097
		HCO@H1.Br@H3.Me@H4	3.095
		CHCH2@H2.SO3H@H3.OEt@H4	3.095
		F@H1.Me@H2.HCO@H3.HCO@H4	3.095
		NO2@H1.MeNH2@H3.HCO@H4	3.094
		CHCH2@H2.SO3H@H3	3.093
		NO2@H3.HCO@H4	3.093

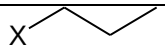
		NO2@H1.NHMe@H4	3.092
		NO2@H2.NO2@H3.NH2@H4	3.091
		Me@H2.CHNH@H3.COOH@H4	3.091
		CHCH2@H2.OMe@H3.SO3H@H4	3.090
		Me@H2.COOH@H3.CONH2@H4	3.089
		NH2@H1.CN@H2.OMe@H3.HCO@H4	3.089
		NO2@H2.HCO@H3.HCO@H4	3.087
		NHMe@H1.SO3H@H3.NHMe@H4	3.087
		MeNH2@H1.HCO@H3.CHNH@H4	3.086
		NO2@H1.CCH@H3.CHCH2@H4	3.086
		OPre@H3.OMe@H4	3.084
		Me@H2.NO2@H3.NH2@H4	3.081
		NO2@H1.F@H3.NH2@H4	3.079
		HCO@H1.CCH@H3.HCO@H4	3.078
		NO2@H1.COOH@H3.CHCH2@H4	3.078
		NO2@H1.OMe@H3.OH@H4	3.078
		NHMe@H1.HCO@H3.CHNH@H4	3.077
		NO2@H1.Me@H3.OEte@H4	3.077
		NO2@H1.NH2@H3.OEte@H4	3.077
		NH2@H1.OMe@H2.CHCH2@H3.SO3H@H4	3.075
		OMe@H1.NO2@H3.COOH@H4	3.074
		NH2@H1.OEte@H2.CONH2@H3.COOH@H4	3.074
		NHMe@H1.OH@H2.NO2@H3.HCO@H4	3.074
		MeNH2@H1.NH2@H2.Cl@H3.CHNH@H4	3.073
		NO2@H1.NH2@H3.OH@H4	3.072
		HCO@H1.HCO@H3.SO3H@H4	3.071
		CHCH2@H1.OH@H2.HCO@H3.CHCH2@H4	3.071
		MeNH2@H1.OH@H2.NO2@H3.CHNH@H4	3.069
		CHCH2@H1.NO2@H2.NH2@H4	3.067
		OMe@H1.CN@H2.NH2@H3.HCO@H4	3.067
		Me@H2.HCO@H3.NH2@H4	3.066
		CHCH2@H1.CHNH@H3.COOH@H4	3.065
		CCH@H2.OMe@H3.NO2@H4	3.064
		Me@H2.NO2@H3.NHMe@H4	3.064
		NH2@H1.OMe@H2.COOH@H3.NO2@H4	3.062
		NO2@H1.NHMe@H3.NHMe@H4	3.061
		OH@H2.CCH@H3.SO3H@H4	3.059
		NO2@H1.COOH@H3.NHMe@H4	3.058
		NHMe@H1.OH@H2.CN@H3.HCO@H4	3.057
		NO2@H1.Br@H3.OH@H4	3.057
		NH2@H2.OMe@H3.Et@H4	3.055

		HCO@H1.Me@H3.CHNH@H4	3.054
		OMe@H1.HCO@H3.CHCH2@H4	3.054
		CN@H2.NO2@H3.CONH2@H4	3.051
		OH@H1.OEt@H2.Me@H4	3.050
		NHMe@H1.OH@H2.CHCH2@H3.CONH2@H4	3.050
		NO2@H2.CHNH@H3.NH2@H4	3.049
		CN@H2.HCO@H3.OMe@H4	3.049
		OMe@H1.HCO@H3.SO3H@H4	3.049
		NH2@H2.OMe@H3.HCO@H4	3.049
		Cl@H2.SO3H@H3.HCO@H4	3.049
		NH2@H2.Et@H3.OMe@H4	3.047
		OMe@H1.NO2@H3.CONH2@H4	3.045
		NH2@H2.CONH2@H3.OMe@H4	3.045
		MeNH2@H1.CCH@H2.CHNH@H3.NH2@H4	3.043
		MeNH2@H1.HCO@H3.HCO@H4	3.042
		HCO@H1.F@H2.Cl@H3.OH@H4	3.042
		CCH@H2.HCO@H3.HCO@H4	3.040
		HCO@H3.NHMe@H4	3.039
		NO2@H1.Cl@H3.Me@H4	3.039
		NH2@H1.NHMe@H2.COOH@H3.CONH2@H4	3.037
		CHCH2@H2.CONH2@H3.SO3H@H4	3.035
		HCO@H1.F@H2.F@H3.COOH@H4	3.035
		CHCH2@H2.NO2@H3.OMe@H4	3.034
		CHCH2@H2.SO3H@H3.OMe@H4	3.033
		CHCH2@H1.SO3H@H3.MeNH2@H4	3.033
		HCO@H1.NO2@H3.OMe@H4	3.033
		MeNH2@H1.NHMe@H3.NO2@H4	3.031
		NHMe@H1.SO3H@H3.SO3H@H4	3.030
		NO2@H1.NO2@H3.Me@H4	3.029
		OMe@H1.NO2@H3.HCO@H4	3.026
		NH2@H1.CHCH2@H2.CHCH2@H3.SO3H@H4	3.026
		NO2@H1.OMe@H4	3.025
		MeNH2@H1.OH@H2.Me@H3.NH2@H4	3.025
		HCO@H1.CN@H2.CONH2@H3.HCO@H4	3.023
		MeNH2@H1.F@H2.CHNH@H3.CN@H4	3.023
		NHMe@H1.NH2@H2.NO2@H3.HCO@H4	3.022
		HCO@H1.CONH2@H3.SO3H@H4	3.021
		HCO@H1.CN@H3.CHCH2@H4	3.021
		HCO@H1.F@H2.CONH2@H3.HCO@H4	3.019
		NO2@H1.CN@H3.HCO@H4	3.018
		NO2@H1.F@H3.CHNH@H4	3.017

		HCO@H1.OH@H3.CHNH@H4	3.017
		OH@H2.Et@H3.Me@H4	3.016
		CHCH2@H1.OH@H2.CN@H3.HCO@H4	3.015
		NO2@H1.Me@H2.CHCH2@H3.Cl@H4	3.015
		NO2@H1.HCO@H3.CCH@H4	3.014
		Me@H2.NO2@H3.HCO@H4	3.012
		NH2@H1.CHCH2@H2.SO3H@H3.CHCH2@H4	3.012
		NH2@H2.NO2@H3.COOH@H4	3.012
		OH@H1.NH2@H3.SO3H@H4	3.010
		CHCH2@H1.OH@H2.Me@H3.MeNH2@H4	3.008
		NO2@H2.OH@H3.CONH2@H4	3.007
		CHCH2@H1.Et@H3.HCO@H4	3.005
		CONH2@H3.HCO@H4	3.005
		HCO@H1.CHNH@H3.CCH@H4	3.005
		NO2@H1.HCO@H4	3.005
		Me@H2.NO2@H3	3.004
		HCO@H1.OMe@H3.NO2@H4	3.004
		OMe@H1.NH2@H3.HCO@H4	3.003
		CHCH2@H1.NO2@H3.NHMe@H4	3.003
		NO2@H1.CCH@H3.HCO@H4	3.002
		NH2@H1.CCH@H2.NO2@H3.CONH2@H4	3.002
		OPre@H2.OH@H3	3.001
		CHCH2@H1.CONH2@H3.HCO@H4	3.000

table S9. Details of functional group codes and their associated structure.

Functional Group Code	Name	Structure	Functional Group Code	Name	Structure
Br	Bromide		CCH	Ethyne	
CF_3	Triflouromethyl		$CHCH_2$	Ethene	
CHNH	Primary Aldimine		Cl	Cloride	
CN	Cyano		$CONH_2$	Acetamide	
COOH	Carboxylic Acid		Et	Ethyl	
F	Fluoride		H	-	
HCO	Aldehyde		I	Iodide	
Me	Methyl		$MeNH_2$	Pendent-Methylamine	
NH_2	Amine		NHMe	Methylamine	
NO_2	Nitro		OEt	Ethoxy	
OEte	Ethene Ether		OH	Hydroxyl	
OMe	Methoxy		OPr	Propoxy	
OPre	Propene Ether		Ph	Phenyl	

Pr	Propyl		SO_3H	Sulfonic Acid