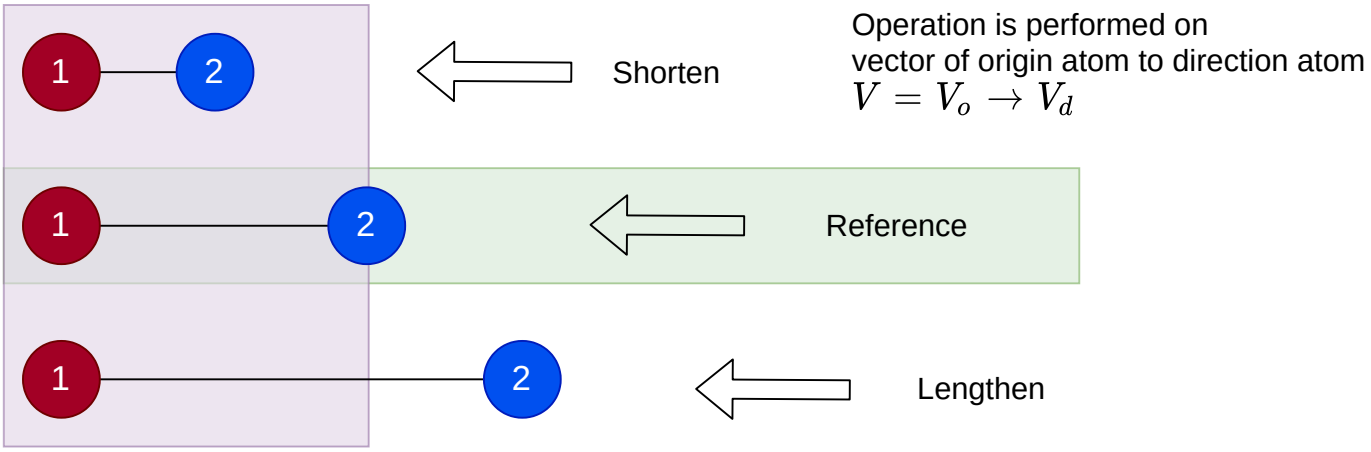
Bonds Variation

More detail info will be written after codes finished.

Let us start at simple cases, for the variation on bonds, it can only have two type of operations, either by shortening or lengthening.

Assume there are two atoms bonded with each other, labelled as 1 & 2, if atom 1 is chosen as the origin;



In the same way, if atom 2 is used as the origin atom, similar operations can be performed on atom 1.

Now, let us define some terms, combinations and permutations, which are actually from mathematics.

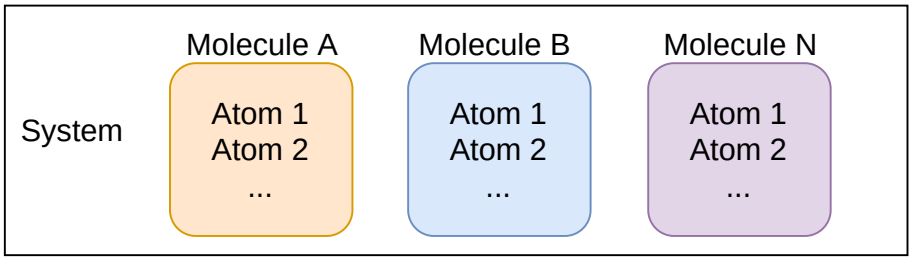
#) Combinations: number of selections on given list, the order of selecting does not matter.
#) Permutations: the order of selection matters.

For example, if there are three items, A, B, and C, two of them will be selected:
*) For Combinations: they can be, A&B, A&C, B&C, in total 3
*) For Permutations: they will be A-B, B-A, A-C, C-A, B-C, C-B, in total 6.

To simplify,
*) Combinations, which is expressed by C_n^i
e.g., $C_3^2 = 3$, where n is the total number of entries, i is the number of selections
*) Permutations, A_n^i
e.g., $A_3^2 = 6$

It is very easy to prove that, $C_n^i = C_n^{n-i}$, because selecting certain number of items is equivalent to its opposite operation, selecting all others.

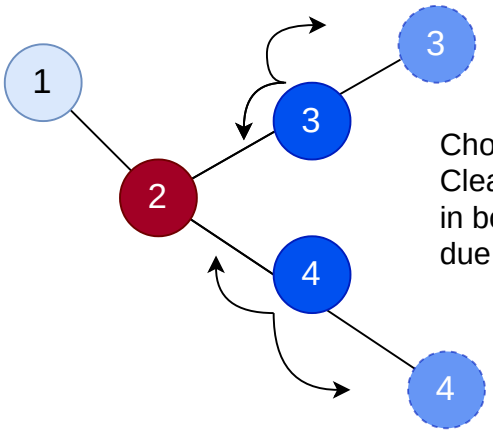
The variation is carried out on system, which can contain many molecules, and each molecules might have many atoms



Most time, the molecule to be sampled will contain many different atoms,
thus every atom will have the different surroundings, termed as in different "environments".

They may have special case, e.g., chemical equivalence, however, it does not matter on bonds variations.

Assume, atom 3 and atom 4 are in chemical equivalence on atom 2.

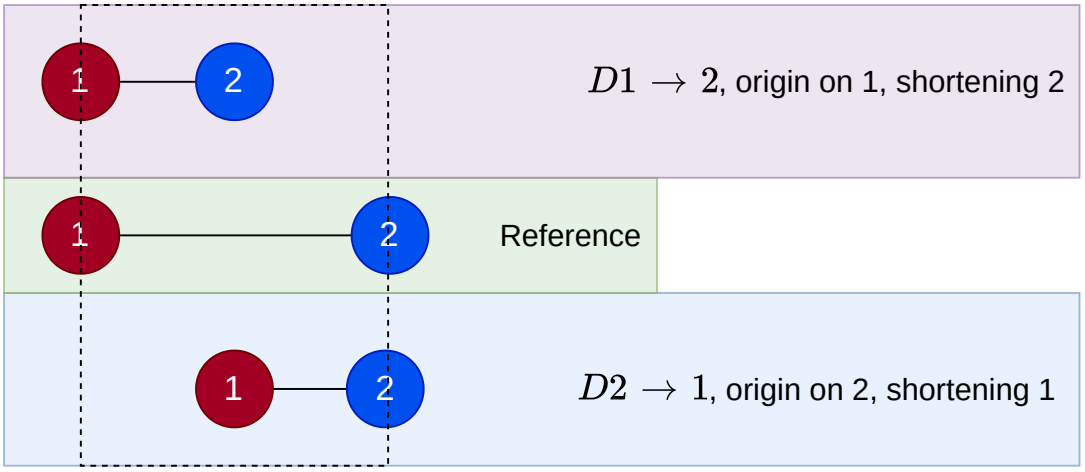


Choose atom 2 as the reference;
Clearly we can see that, varying bonds either on atom 3 and atom 4 in both directions will generate new molecules and without repeats, due to the changes of relative distances with atom 1.

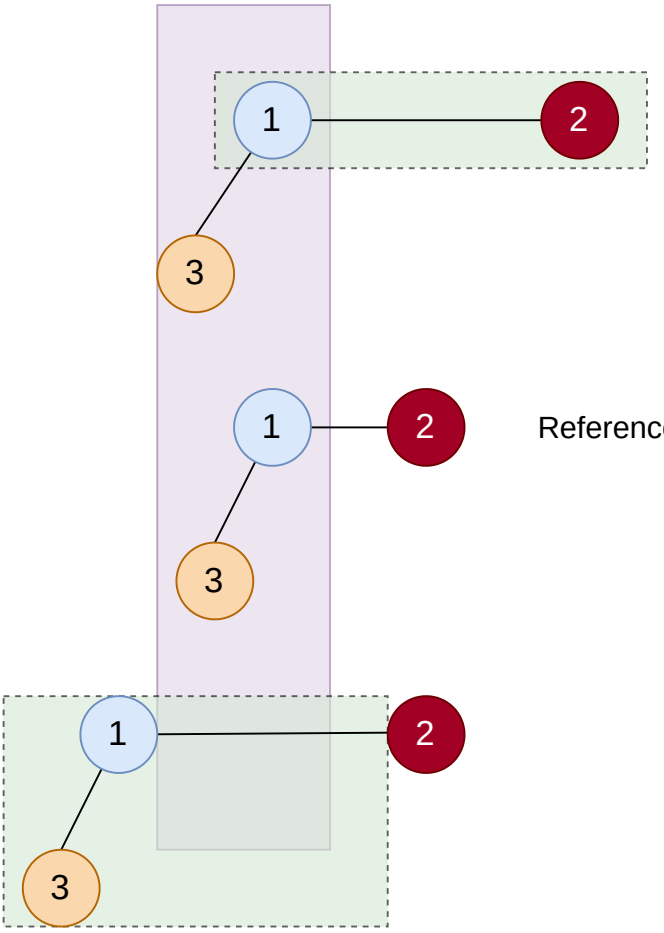
Therefore, it is safely to draw the conclusion that varying on bonds is in **permutations**.

Once the vector is set, we only need to vary bonds on target atoms.
Since the operations will be performed on all target atoms, their order does not matter,
thus, the selection of target atoms are in **combinations**.

For molecule has two atoms, as you can see, it has repeats.



Similarly, for molecule has three atoms;



$D1 \rightarrow 2$, origin on 1, lengthen 2

Reference

$D2 \rightarrow 1$, origin on 2, lengthen 1&3

Still, it has repeats.

It is not hard to understand that lengthening the distance of one items from all others is actually equivalent with lengthening the distance of all others with this atom,
but on the same vector's different directions.

> Those repeats can be proved that, they are only occurring on selecting on one of origin atoms, then varying all others with the opposite direction except direction atoms, in the same distance.

$$C_n^i \rightarrow C_n^{n-i} \quad \& \quad C_n^i \leftarrow C_n^{n-i} = C_n^i \Leftrightarrow C_n^{n-i}$$

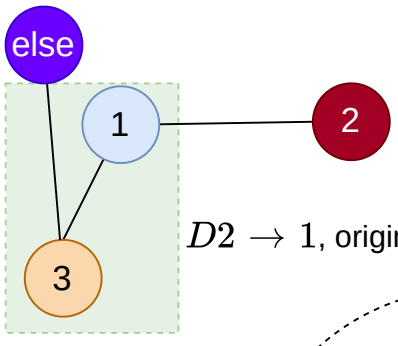
$$Di \rightarrow j, vary(m_j)$$

$$= Dj \rightarrow i, vary \forall (m_k), where\ k \neq j$$

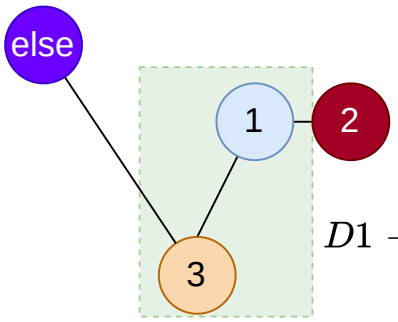
Sign $\forall$ means any/every

How about on other conditions, do they have repeats?

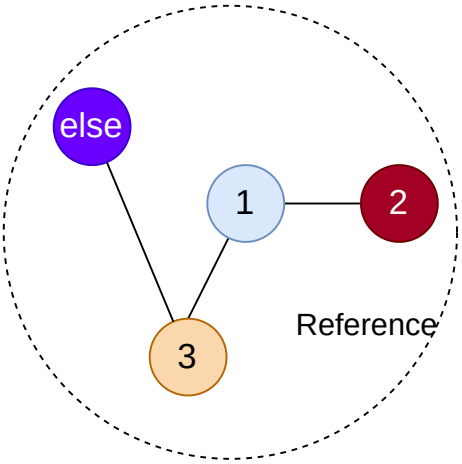
The answer is, no, they do not have. Please see following graphs.



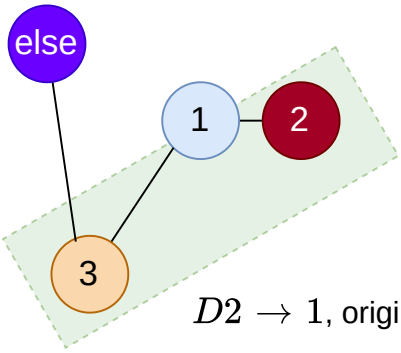
$D2 \rightarrow 1$, origin on 2, vary 1,3



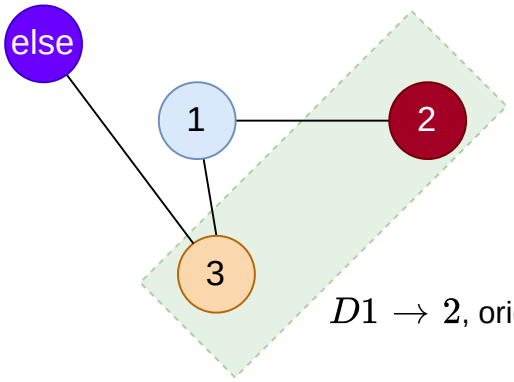
$D1 \rightarrow 2$, origin on 2, vary 1,3



Obviously, all four variations have no repeats at all.

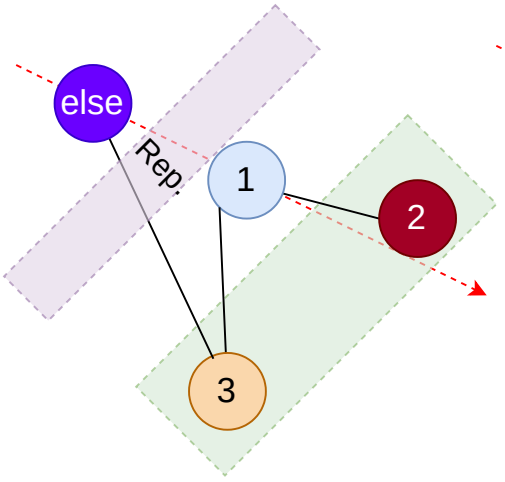


$D2 \rightarrow 1$, origin on 1, vary 2,3

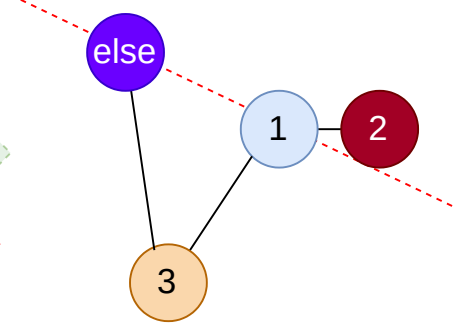


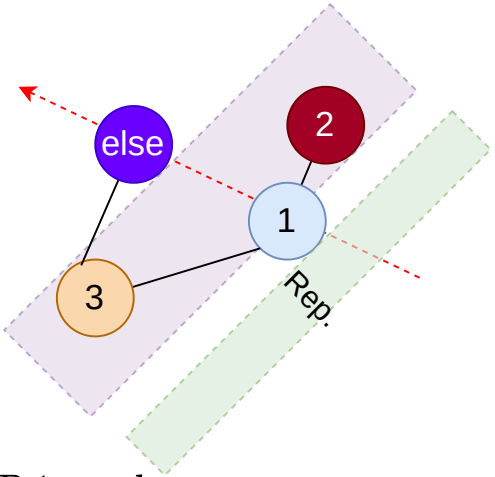
$D1 \rightarrow 2$, origin on 1, vary 2,3

Caution, based on top graph, we naturally assume that the varying of bonds contains the direction atom. Otherwise, they will still have repeats. For example,



$D\ else \rightarrow 1$, origin on else, vary 2,3





$D\ 1 \rightarrow else$, origin on 1, vary 2,3

Thus, for sampling without direction atom, only half of them are considered.
(Facilitated by different background colors)

Based on top discussions, we will have a formula for the total number of bonds variations for n atoms

$$VB(n) = O_2 \cdot A^2_{n\underset{ij}{\rightarrow}} \left[\sum_{\Omega_j=1}^{n-2} C^{\Omega_j}_{n-2} - 1 \right] + A^2_{n\underset{ij}{\rightarrow}} \sum_{\Gamma_{ij}=1}^{n-2} C^{\Gamma_{ij}}_{n-2}$$

$$= A^2_{n\underset{ij}{\rightarrow}} \left[O_2 \sum_{\Omega_j=1}^{n-3} C^{\Omega_j}_{n-2} + \sum_{\Gamma_{ij}=1}^{n-2} C^{\Gamma_{ij}}_{n-2} \right]$$

where, $n \geq 3$, O_2 means operation on dual directions, $Di \rightarrow j$, origin on i

Ω_j means combination always contains j, whereas Γ_{ij} means excludes i and j

Besides, we define $VB(2) = 2$

Then we can calculate the total number of variations for different number of atoms

$VB(3) = 18$
$VB(4) = 108$
$VB(5) = 420$
...
$VB(10) = 68850$

The following are examples, to make you better understand this algorithm...

Take care of their corresponding colors.

For 3 atoms

atoms	origin	direction	dual	single	repeats	nm	
	1	2	2	3	2, 3	1	total number
	2	1	1	3	1, 3	2	6*1*2 + 6*1
	1	3	3	2	3, 2	3	= 18
1, 2, 3	3	1	1	2	1, 2	4	
	2	3	3	1	3, 1	5	
	3	2	2	1	2, 1	6	

For 4 atoms

atoms	origin	direction		dual			single		repeats	nm	
	1	2	2	2, 3	2, 4	3	4	3, 4	2, 3, 4	1	
	2	1	1	1, 3	1, 4	3	4	3, 4	1, 3, 4	2	total number
	1	3	3	3, 2	3, 4	2	4	2, 4	2, 3, 4	3	12*3*2 + 12*3
	3	1	1	1, 2	1, 4	2	4	2, 4	1, 2, 4	4	= 108
	1	4	4	4, 2	4, 3	2	3	2, 3	2, 3, 4	5	
1, 2, 3, 4	4	1	1	1, 2	1, 3	2	3	2, 3	1, 2, 3	6	
	2	3	3	3, 1	3, 4	1	4	1, 4	1, 3, 4	7	
	3	2	2	2, 1	2, 4	1	4	1, 4	1, 2, 4	8	
	2	4	4	4, 1	4, 3	1	3	1, 3	1, 3, 4	9	
	4	2	2	2, 1	2, 3	1	3	1, 3	1, 2, 3	10	
	3	4	4	4, 1	4, 2	1	2	1, 2	1, 2, 4	11	
	4	3	3	3, 1	3, 2	1	2	1, 2	1, 2, 3	12	

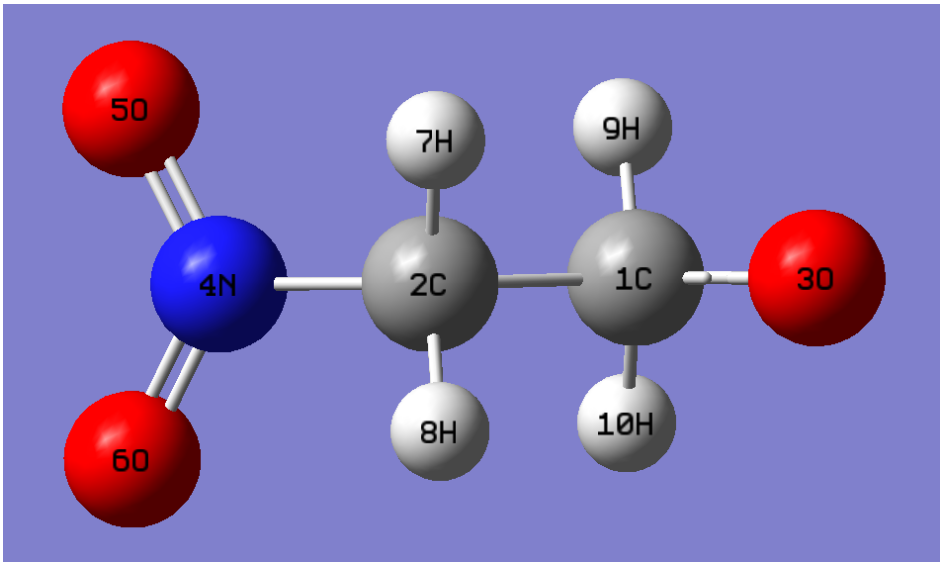
For 5 atoms

atoms	origin	direction				dual							single				repeats
	1	2	2	2, 3	2, 4	2, 5	2, 3, 4	2, 3, 5	2, 4, 5	3	4	5	3, 4	3, 5	4, 5	3, 4, 5	2, 3, 4, 5
	2	1	1	1, 3	1, 4	1, 5	1, 3, 4	1, 3, 5	1, 4, 5	3	4	5	3, 4	3, 5	4, 5	3, 4, 5	1, 3, 4, 5
	1	3	3	3, 2	3, 4	3, 5	3, 2, 4	3, 2, 5	3, 4, 5	2	4	5	2, 4	2, 5	4, 5	2, 4, 5	2, 3, 4, 5
	3	1	1	1, 2	1, 4	1, 5	1, 2, 4	1, 2, 5	1, 4, 5	2	4	5	2, 4	2, 5	4, 5	2, 4, 5	1, 2, 4, 5
	1	4	4	4, 2	4, 3	4, 5	4, 2, 3	4, 2, 5	4, 3, 5	2	3	5	2, 3	2, 5	3, 5	2, 3, 5	2, 3, 4, 5
	4	1	1	1, 2	1, 3	1, 5	1, 2, 3	1, 2, 5	1, 3, 5	2	3	5	2, 3	2, 5	3, 5	2, 3, 5	1, 2, 3, 5
	1	5	5	5, 2	5, 3	5, 4	5, 2, 3	5, 2, 4	5, 3, 4	2	3	4	2, 3	2, 4	3, 4	2, 3, 4	5, 2, 3, 4
	5	1	1	1, 2	1, 3	1, 4	1, 2, 3	1, 2, 4	1, 3, 4	2	3	4	2, 3	2, 4	3, 4	2, 3, 4	1, 2, 3, 4
1, 2, 3, 4, 5	2	3	3	3, 1	3, 4	3, 5	3, 1, 4	3, 1, 5	3, 4, 5	1	4	5	1, 4	1, 5	4, 5	1, 4, 5	1, 3, 4, 5
	3	2	2	2, 1	2, 4	2, 5	2, 1, 4	2, 1, 5	2, 4, 5	1	4	5	1, 4	1, 5	4, 5	1, 4, 5	1, 2, 4, 5
	2	4	4	4, 1	4, 3	4, 5	4, 1, 3	4, 1, 5	4, 3, 5	1	3	5	1, 3	1, 5	3, 5	1, 3, 5	1, 3, 4, 5
	4	2	2	2, 1	2, 3	2, 5	2, 1, 3	2, 1, 5	2, 3, 5	1	3	5	1, 3	1, 5	3, 5	1, 3, 5	1, 2, 3, 5
	2	5	5	5, 1	5, 3	5, 4	5, 1, 3	5, 1, 4	5, 3, 4	1	3	4	1, 3	1, 4	3, 4	1, 3, 4	5, 1, 3, 4
	5	2	2	2, 1	2, 3	2, 4	2, 1, 3	2, 1, 4	2, 3, 4	1	3	4	1, 3	1, 4	3, 4	1, 3, 4	2, 1, 3, 4
	3	4	4	4, 1	4, 2	4, 5	4, 1, 2	4, 1, 5	4, 2, 5	1	2	5	1, 2	1, 5	2, 5	1, 2, 5	1, 2, 4, 5
	4	3	3	3, 1	3, 2	3, 5	3, 1, 2	3, 1, 5	3, 2, 5	1	2	5	1, 2	1, 5	2, 5	1, 2, 5	1, 2, 3, 5
	3	5	5	5, 1	5, 2	5, 4	5, 1, 2	5, 1, 4	5, 2, 4	1	2	4	1, 2	1, 4	2, 4	1, 2, 4	5, 1, 2, 4
	5	3	3	3, 1	3, 2	3, 4	3, 1, 2	3, 1, 4	3, 2, 4	1	2	4	1, 2	1, 4	2, 4	1, 2, 4	3, 1, 2, 4
	4	5	5	5, 1	5, 2	5, 3	5, 1, 2	5, 1, 3	5, 2, 3	1	2	3	1, 2	1, 3	2, 3	1, 2, 3	5, 1, 2, 3
	5	4	4	4, 1	4, 2	4, 3	4, 1, 2	4, 1, 3	4, 2, 3	1	2	3	1, 2	1, 3	2, 3	1, 2, 3	4, 1, 2, 3

total columns are 20

total number
20*7*2 + 20*7
= 420

If we have a reference molecule like,



From the above image, we know its both bonded and non-bonded information.

When the bonds are sampled, it has two basic rules should also be fulfilled.

Rule:
1) Bonds should be kept
2) Nonbonds should be kept

Codes provide manual settings, as well as,

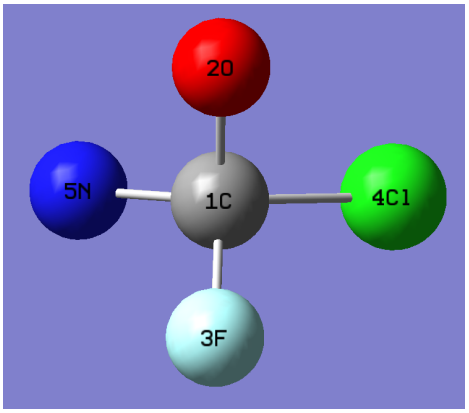
auto-identification, reference used is:

Zhang, Q., et al.
A rule-based algorithm for automatic bond type perception.
J Cheminform 4, 26 (2012). https://doi.org/10.1186/1758-2946-4-26

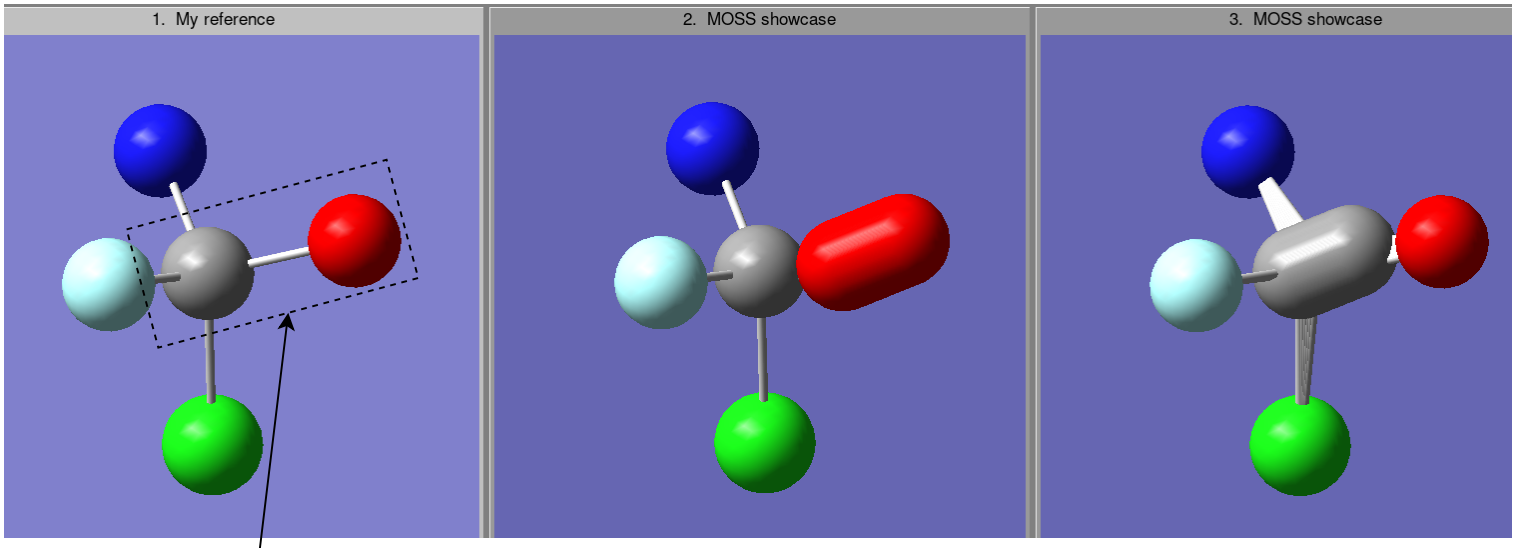
I will not go much detail on bonds perception, we only need to know that codes know the bonded information, and it will do filtrations after sampling is done.

Showcase

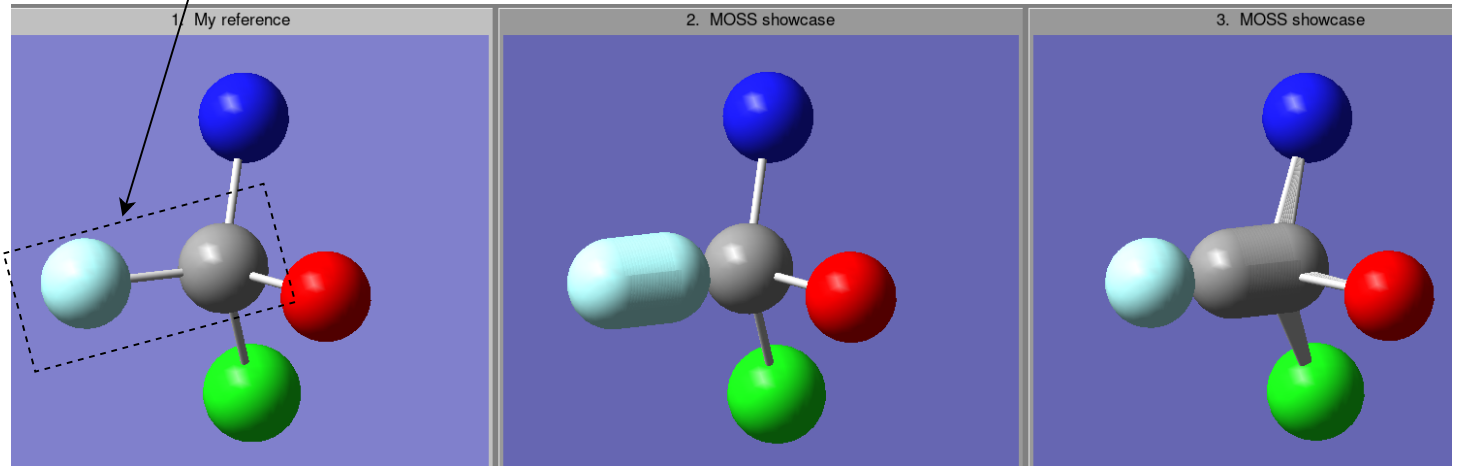
Assume there is a simple molecule like,



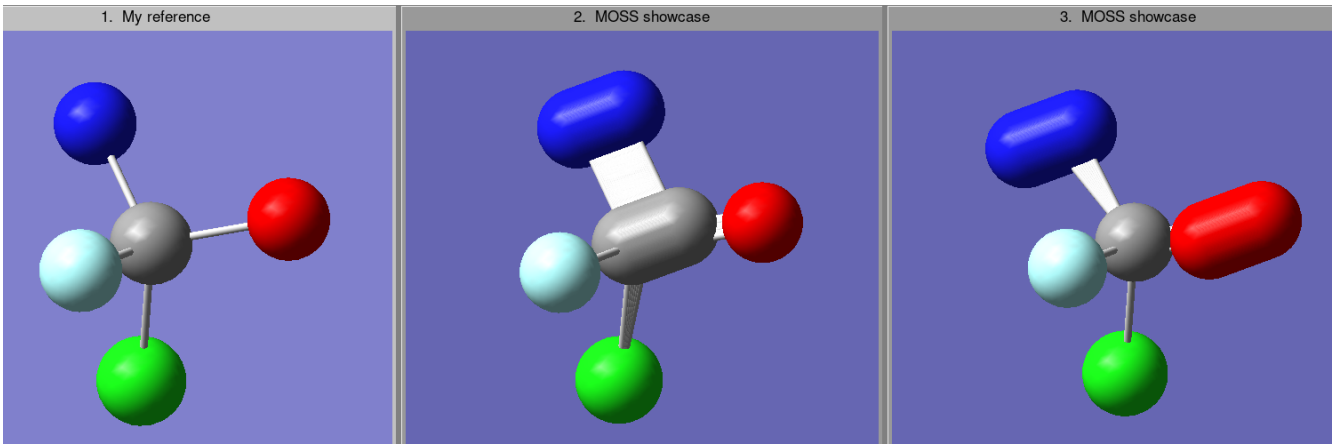
Dual direction, vary single atom



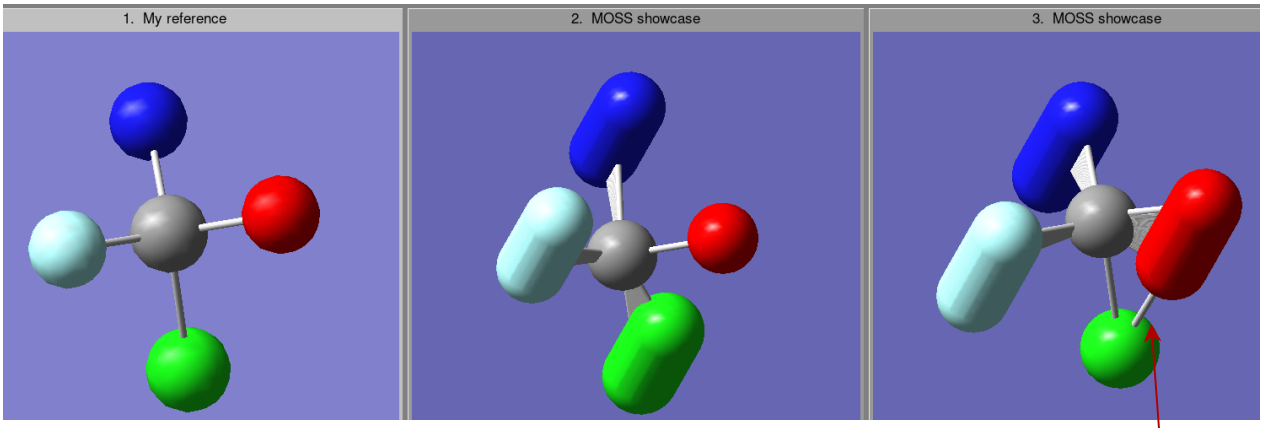
origin atom & direction atom



Dual direction, vary two atoms



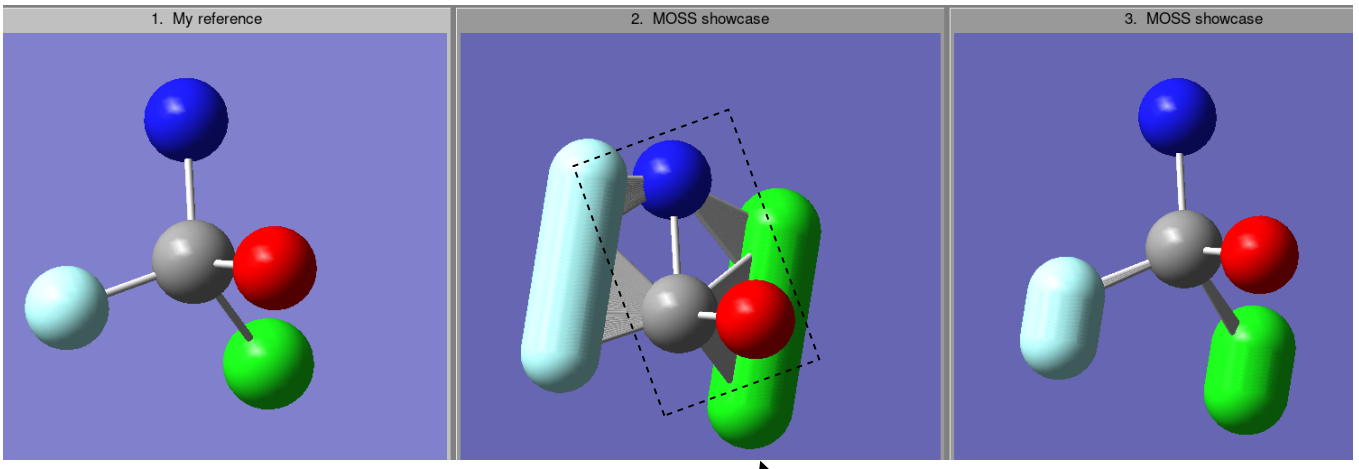
Dual direction, vary three atoms



Because of the different settings in graphical program,
The rendered "bond" does not mean it is the actual bond

From the plots, it again proves that, for dual directions, bonds variation including origin atom in both direction at every increment will generate new/distinct molecules

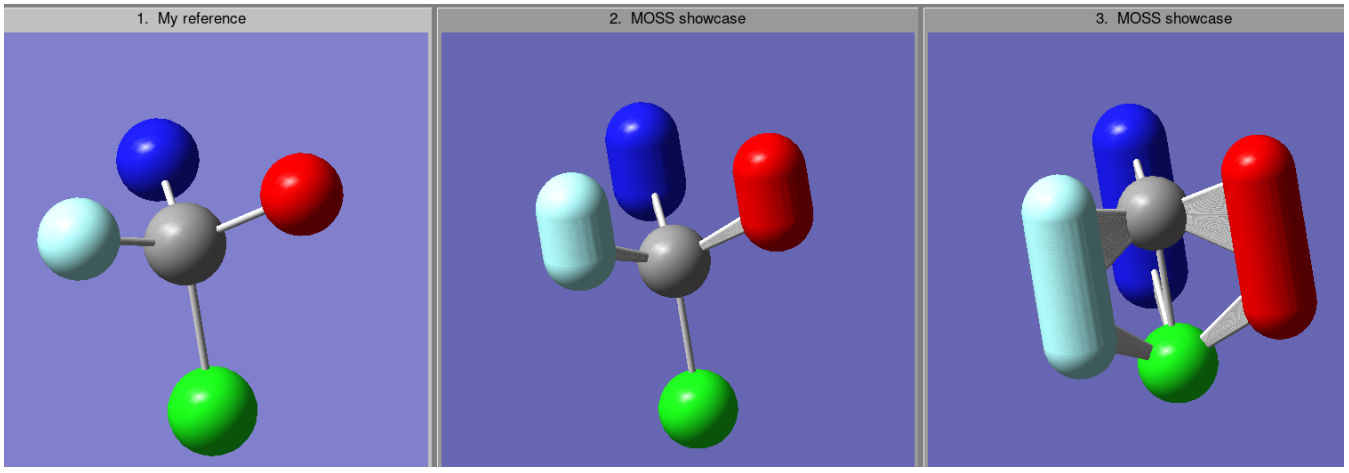
Single direction, vary two atoms



They are actually at the same coordinates with other two's, because the graphical program makes its Center of Coordinates at the center of window, which is why it looks weird.

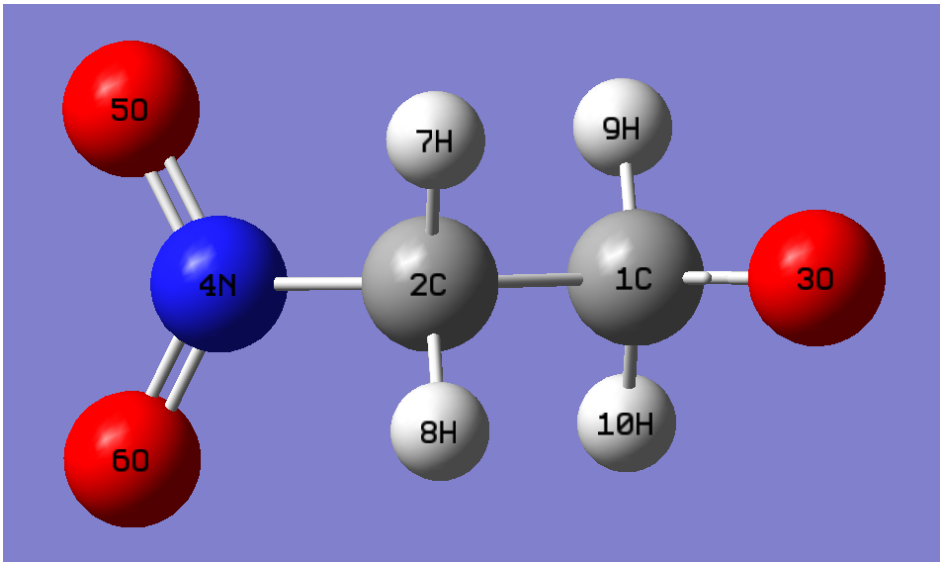
The following plots have the similar problem but with the middle two atoms.

Single direction, vary three atoms



For the molecule that used for Henry reaction that we interest,
it has in total 10 atoms, thus the total number of variations are:

$VB(10)$ = 68850



For the total number of molecule generations,
after double checked/filtered out by bonded and nonbonded settings,
at increment of 0.03 Angstrom,
they are 703249.

The End