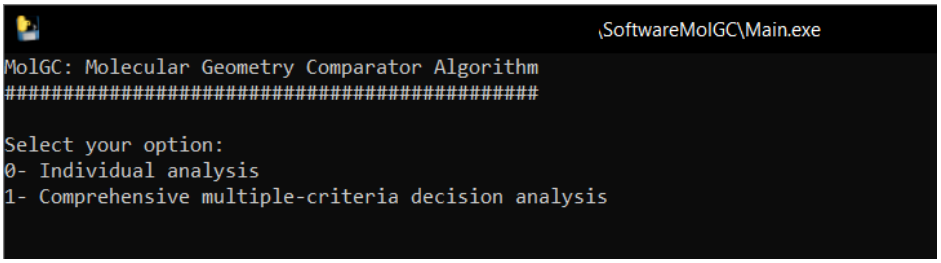
MolGC: Molecular Geometry Comparator Algorithm for Bond Length Mean Absolute Error Computation on Molecules

Instructions for Setting Up and Running MolGC

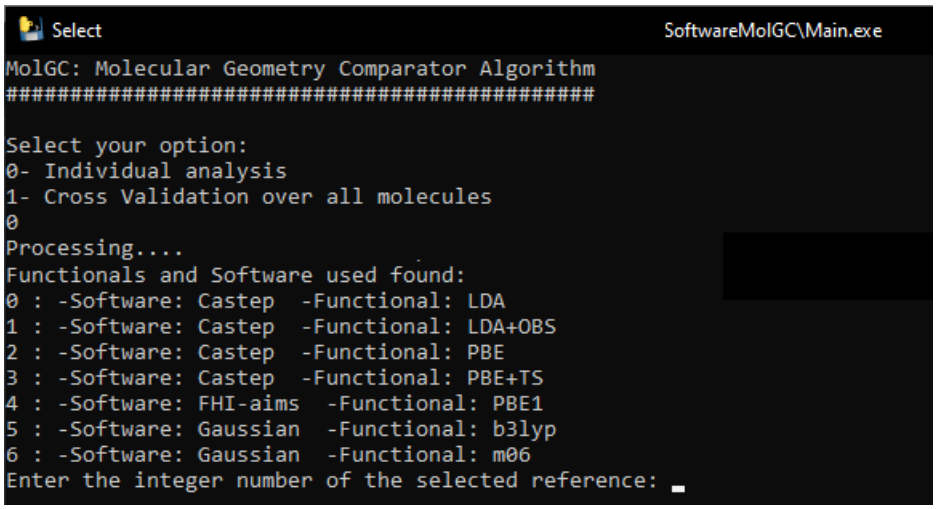
1. Download the files included in the repository https://github.com/AbimaelGP/MolGC/tree/main
2. Extract the downloaded files to your preferred directory.
   The files and folders correspond to molecules from **three different families**.
3. The executable file of MolGC can be downloaded from https://drive.google.com/file/d/1LIWz2i9R6OR2iskV27tD4KlR1kloa0h3/view?usp=sharing
4. Move the downloaded executable file into the ***SoftwareMolGC*** folder within the extracted directory.
5. Execute the Main.exe program.
6. In the first dialog window, select the option you want to run.



Option 0. Individual analysis - Allows the user to specify a reference functional for the analysis.

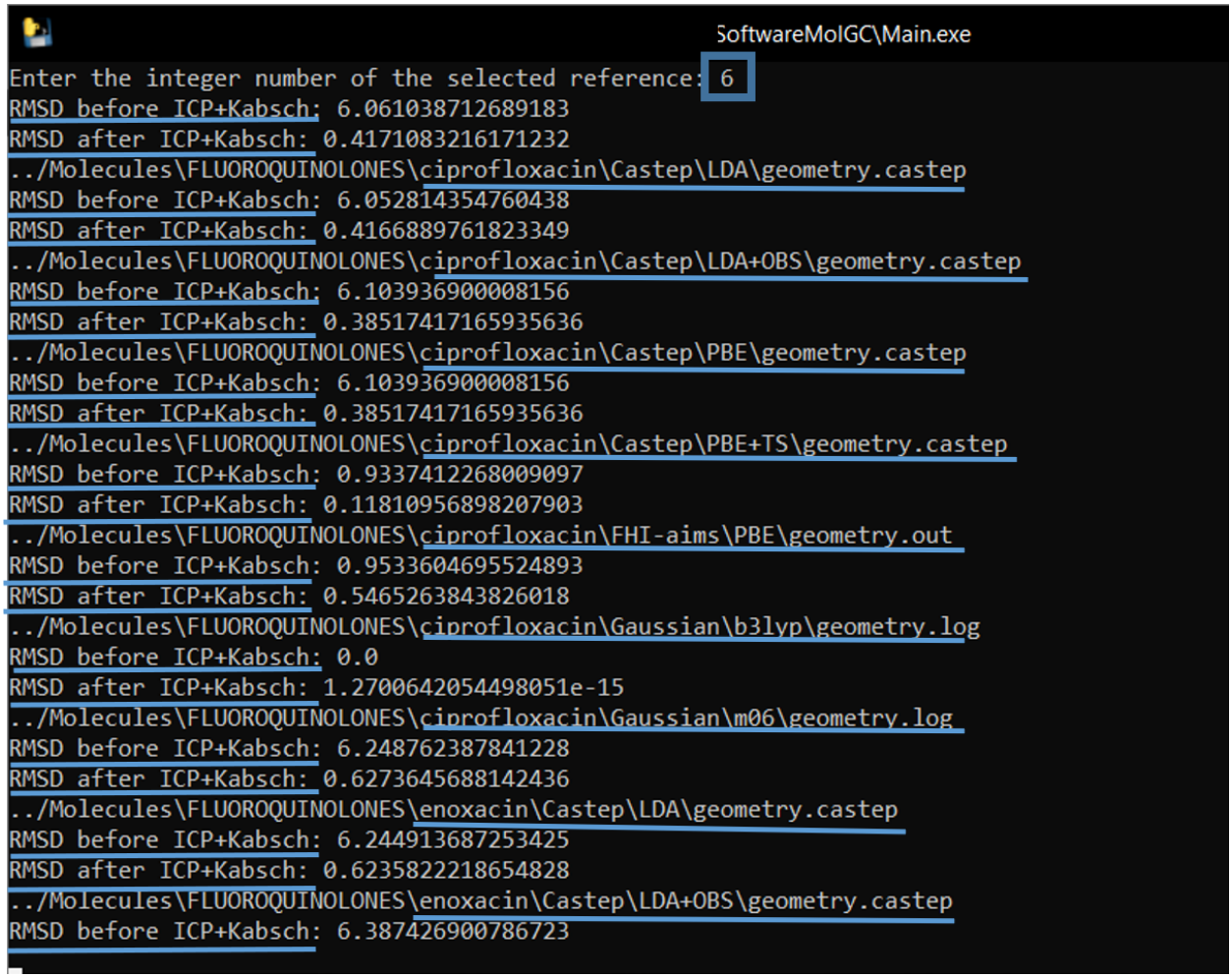
Option 1. Cross Validation over all molecules - Performs analysis considering all available functionals.

7. When selecting **Option 0**, MolGC will identify all possible functionals and software available in the ***Molecules*** folder.

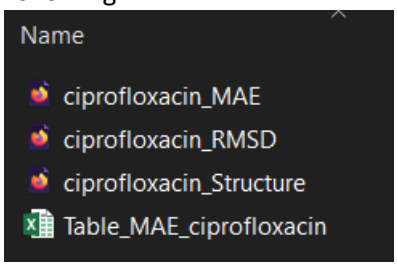


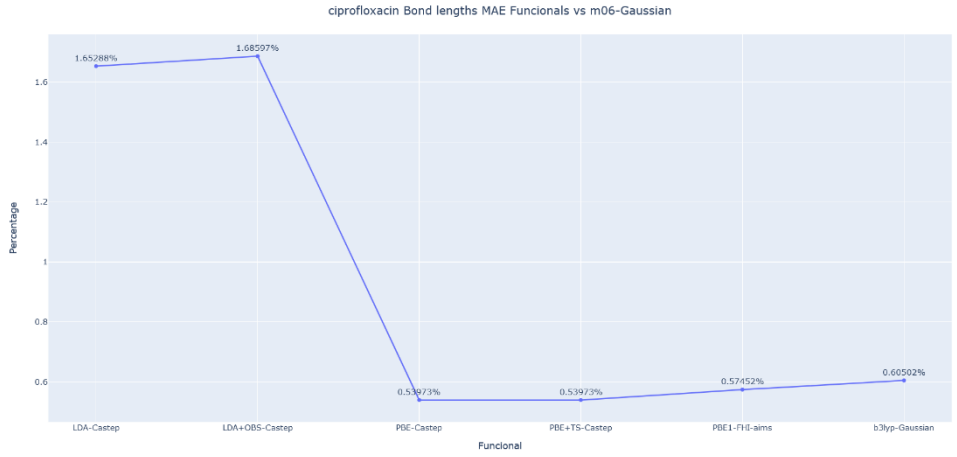
As can be observed, the molecules in the GitHub repository have seven functionals acquired with three different software.

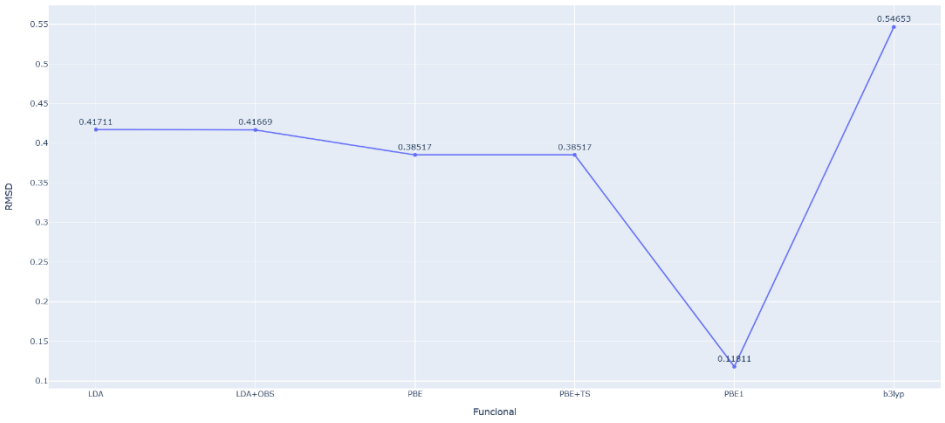
For example, when considering using the **m06** functional, you must select **option 6**, and MolGC will start analyzing this functional considering all the molecules placed in the corresponding folders. In the figure below, we can observe that MolGC will compute the Root Mean Square Deviation (RMSD) before and after the alignment process implemented with the Iterative Closest Point (ICP) and the Kabsch algorithms, minimizing the RMSD value after its computation.



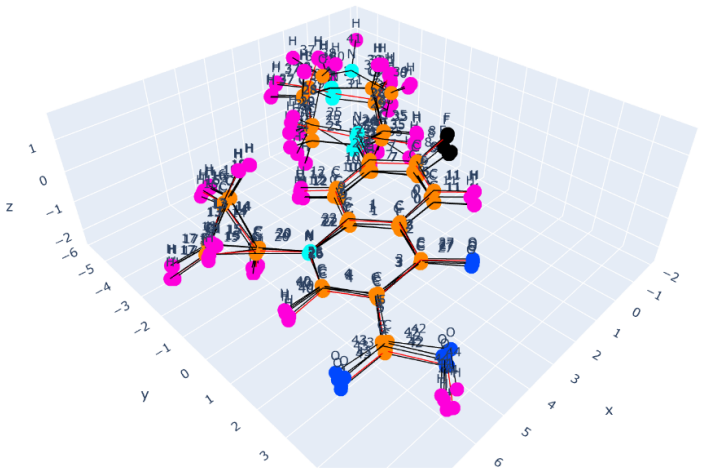
8. Once MolGC completes the analysis of each molecule, a resulting file will be generated in the ***Files_Ind_Eval*** folder. For example, for the **ciprofloxacin** molecule, MolGC will produce the following:



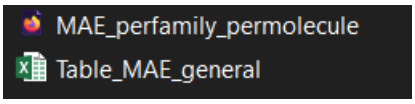


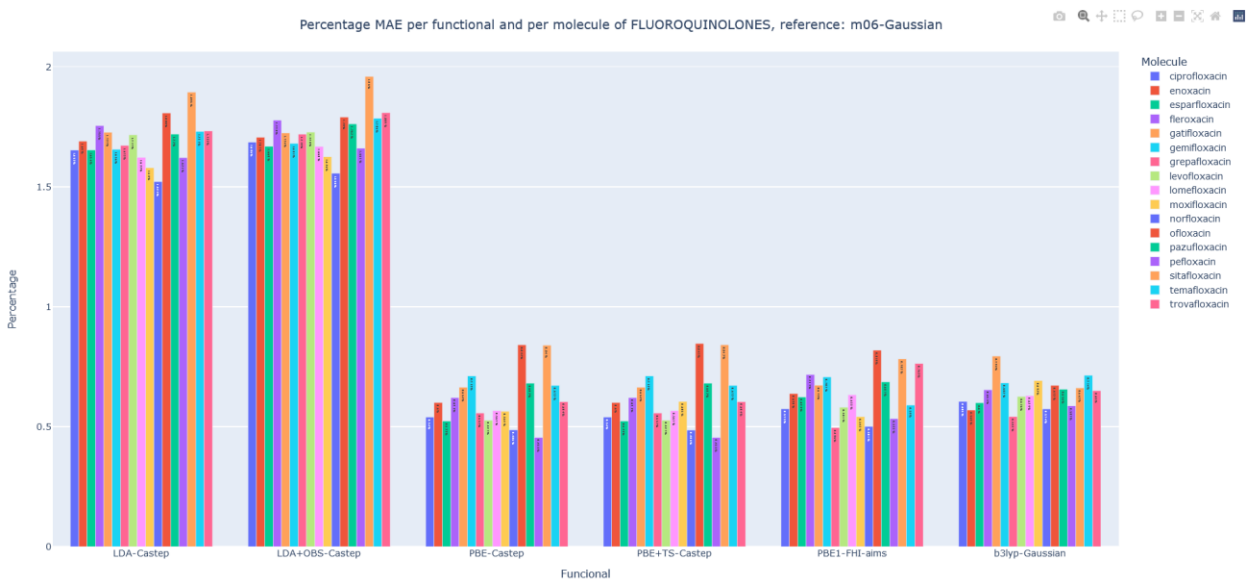


../Molecules\FLUOROQUINOLONES structure



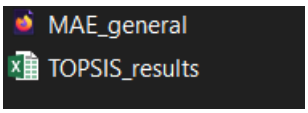
9. Once MolGC finishes the analysis with all the molecules of one family, it will produce the corresponding output files as presented below:

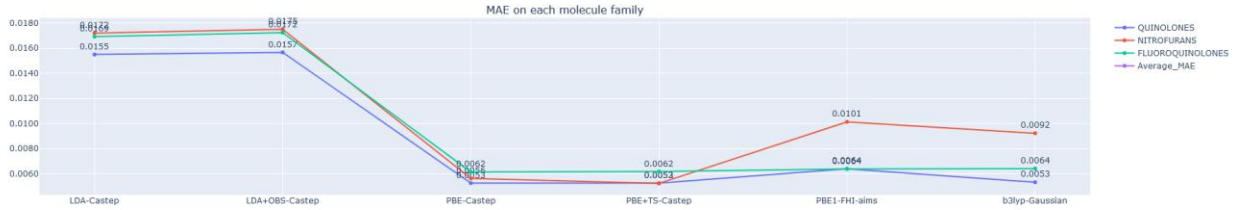


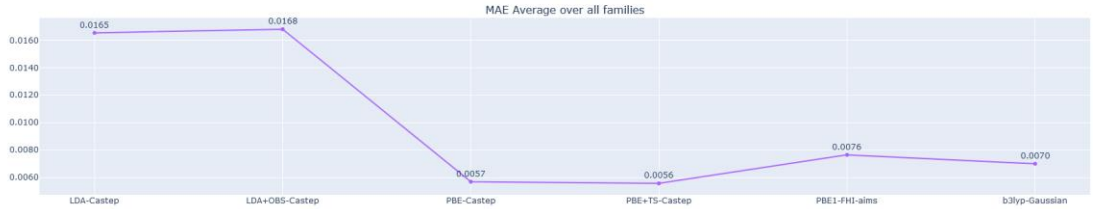


	LDA-Castep	LDA+OBS-Castep	PBE-Castep	PBE+TS-Castep	PBE1-FHI-aims	b3lyp-Gaussian	Molecule	Family
0	0.0165288216924504 16	0.016859723324091 11	0.00539725264209783	0.00539725264209783	0.00574524210112493 3	0.00605021182966641	ciprofloxacin	FLUOROQUINOLONES
1	0.016902701192035 18	0.017066298057715 215	0.0060019540745550074	0.0060019540745550074	0.00638045613049102	0.00569226564211480 9	enoxacin	FLUOROQUINOLONES
2	0.016536815192467 02	0.016685721930499 907	0.0052340829204226775	0.0052340829204226775	0.00623894945535118 6	0.00599548404068337 6	esparfloxacin	FLUOROQUINOLONES
3	0.017546309461680 173	0.017780809206504 246	0.006212558482878911	0.006212558482878911	0.00717496228878319 3	0.00654129766352290 5	fleroxacin	FLUOROQUINOLONES
4	0.017269511864017 012	0.017242005577764 265	0.006642071750293062	0.006642071750293062	0.00672809809035783 5	0.00794270346878455	gatifloxacin	FLUOROQUINOLONES
5	0.016561963119439 51	0.016809485605691 27	0.007116387233869877	0.007116387233869877	0.00707245032065164 2	0.00682459813230274 2	gemifloxacin	FLUOROQUINOLONES
6	0.016720890737453 933	0.017194589450720 31	0.005571157875219633	0.005571157875219633	0.00495961311304698 9	0.00541896339732944 5	grepafloxacin	FLUOROQUINOLONES
7	0.017168217199623 85	0.017283556705951 088	0.0052698664482611025	0.0052698664482611025	0.00580657993749361 75	0.00625891978019952	levofloxacin	FLUOROQUINOLONES
8	0.016224899978936 988	0.016681615164573 636	0.005667528591977015	0.005667528591977015	0.00632871626958851 5	0.00629482502651588 3	lomefloxacin	FLUOROQUINOLONES
9	0.015795413795934 934	0.016255168212467 124	0.005638791423038884	0.006049949891571753	0.00541636120832648 25	0.00692351503162904 1	moxifloxacin	FLUOROQUINOLONES
10	0.015216977293247 601	0.015561975242971 232	0.0048640746758740585	0.004858937355262944	0.00500685608151427 3	0.00574207021374974 9	norfloxacin	FLUOROQUINOLONES
11	0.018075790013578 59	0.017903358215017 602	0.008415445941188307	0.00846856890075896	0.00818569336493066 8	0.00671904960618727	ofloxacin	FLUOROQUINOLONES
12	0.017200229515288 514	0.017621245174907 443	0.006815838416580974	0.006815838416580974	0.00687008785285716 2	0.00655979792614716 9	pazufloxacin	FLUOROQUINOLONES
13	0.016214084978247 45	0.016605462393605 004	0.004552288362052176	0.004552288362052176	0.00533408896340603 4	0.00586761575040486 9	pefloxacin	FLUOROQUINOLONES
14	0.018946316003832 363	0.019602598517123 693	0.008397051260293786	0.008416498446575044	0.00782372594370249 8	0.00661572245174558 9	sitafloxacin	FLUOROQUINOLONES
15	0.017310312533571 44	0.017855573158562 038	0.006716629686821507	0.006716629686821507	0.0058989103585745 2	0.00713919704772186	temafloxacin	FLUOROQUINOLONES
16	0.017325725663072 75	0.018090745215077 28	0.00604325272230901	0.00604325272230901	0.00763161481277693 35	0.00649811582157571 1	trovafloxacin	FLUOROQUINOLONES

10. Finally, the general Mean Absolute Error (MAE) is generated considering all families



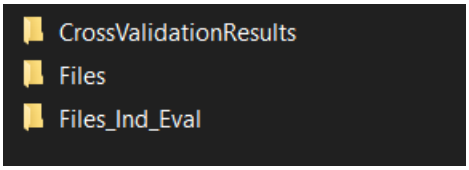




TOPSIS results

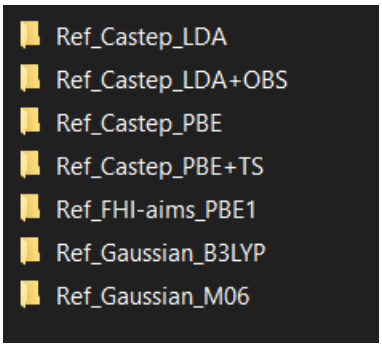
Functional	FLUOROQUINOLONES	NITROFURANS	QUINOLONES	D_noIdeal	D_ideal	Weigthing_closness_to_Idea	Ranking
PBE+TS-Castep	0.006178519047147498	0.005235331995911316	0.005261415873931377	0.7209266653290396	0.0010339666631554445	0.0014321648380557137	1
PBE-Castep	0.006150366618101988	0.005634317245374592	0.005263037635743979	0.7135693066920319	0.013689573669502644	0.0188235221860712	2
b3lyp-Gaussian	0.006416726637075346	0.009229532495670923	0.005331777986653676	0.6431890673778856	0.1374215775247408	0.17604368890188885	3
PBE1-FHI-aims	0.006388376840763383	0.010135665723436346	0.006399936139980013	0.6029336166787604	0.17455232734255266	0.2245086598475762	4
LDA-Castep	0.016914410602051632	0.017201978325728104	0.015499639234510865	0.017306133047545413	0.7046781763658682	0.9760297657138755	5
LDA+OBS-Castep	0.017241172420778966	0.0175066687314938	0.01566755649115396	0	0.721509898210694	1	6

11. When selecting **Option 1**, MolGC will create two new folders: ***Files*** and ***CrossValidationResults***.



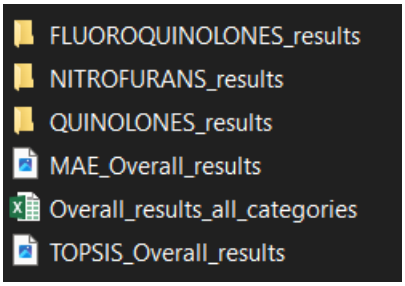
*The ***Files_Ind_Eval*** folder was previously created with **Option 0**.

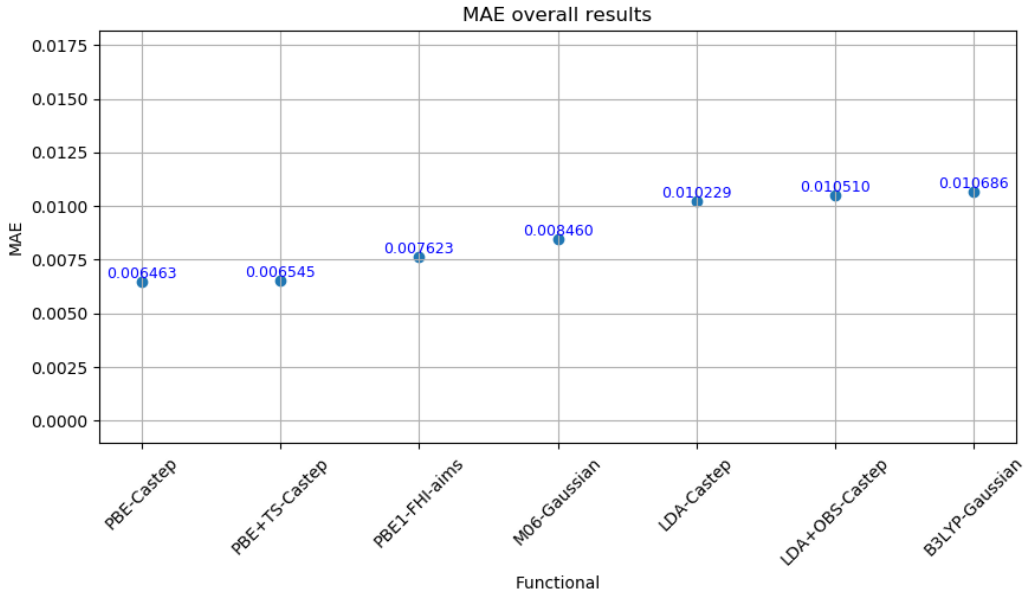
The ***Files*** folder contains all the evaluations considering that each of the functionals was selected as an individual reference

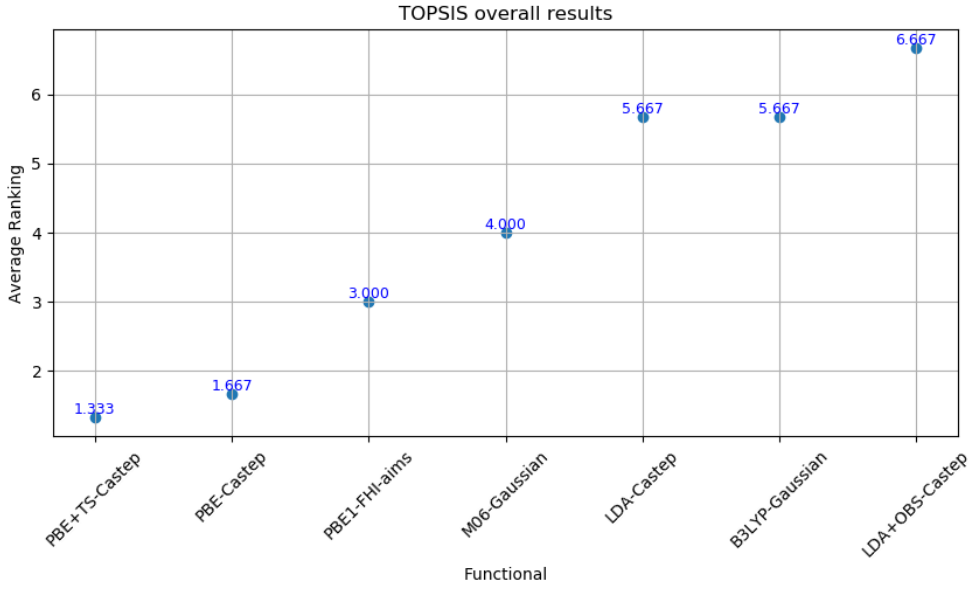


*For example, steps 7 to 10 explain the MolGC process when selecting as reference **m06** (last folder of the previous figure).

The ***CrossValidationResults*** folder considers all the results of the ***Files*** folder and computes the **Comprehensive Multiple-Criteria Decision Analysis**.







Average MAE over all molecules families:

Functional	FLUOROQUINOLONES	NITROFURANS	QUINOLONES	Overall_Average_MAE
PBE-Castep	0.006272455	0.007495376	0.005622547	0.006463459
PBE+TS-Castep	0.006252219	0.007763014	0.005619821	0.006545018
PBE1-FHI-aims	0.007077254	0.009266967	0.006523744	0.007622655
M06-Gaussian	0.008459293	0.009277642	0.00764195	0.008459628
LDA-Castep	0.010542267	0.010713386	0.009430511	0.010228721
LDA+OBS-Castep	0.010911655	0.010954483	0.009665268	0.010510469
B3LYP-Gaussian	0.010134196	0.013146264	0.008778837	0.010686433

TOPSIS Average ranking over all molecules families:

Functional	FLUOROQUINOLONES	NITROFURANS	QUINOLONES	Overall_Average_Rank
PBE+TS-Castep	1	2	1	1.333333333
PBE-Castep	2	1	2	1.666666667
PBE1-FHI-aims	3	3	3	3
M06-Gaussian	4	4	4	4
LDA-Castep	6	5	6	5.666666667
B3LYP-Gaussian	5	7	5	5.666666667
LDA+OBS-Castep	7	6	7	6.666666667