

DP4+ App

<https://github.com/Sarotti-Lab/> . . .

sarotti@iquir-conicet.gov.ar

Instructive, general recommendations and case study

Content

Overview and usage recommendations	1
Probability calculations: DP4+, MM-DP4+ and Custom-DP4+	2
Prepare your files	2
Perform a calculation	3
Results output	4
Reparametrization: Custom-DP4+	5
Create a new level	5
Update a level	6
Warnings and Input control	7
Questionable values	7
Gaussian calculation files	7
Data spread sheet	8
Malfunctions report	8

Overview and usage recommendations

DP4+ App is an integrated software capable of performing already parameterized DP4+ and MM-DP4+ calculations. Furthermore, Custom-DP4+ calculations can be performed, where any level of theory required can be parameterized. Its friendly graphical interface allows easy manipulation of multiple Gaussian calculations and automatic information processing to perform the probabilistic calculus.

To use the application, it is only necessary to create a folder that contains the following files:

- Well labeled Gaussian output files of the NMR calculations of all conformers for all isomeric candidates.
- An Excel file with the experimental information and the correlation labels of each nucleus with the Gaussian calculations.

The recommendations below should be followed for optimal use of the program:

- 1) Despite DP4+ App can handle any amount of isomers, keeping the number of candidates to a minimum has several advantages, as it reduces both the overall computational cost and the probability that the calculated data for an incorrect isomer ends up having better fit with the experimental values than the correct candidate.
- 2) The conformational search should provide a good description of the conformational landscape of the system under study. Improper computational work might lead to potentially negative consequences in the overall results. Systematic sampling is always recommended, but impractical in highly flexible molecules.

In those cases, stochastic searches using a reasonably large number of steps should be carried out. To avoid missing potentially relevant conformations, all conformations within a safe energy window from the corresponding global minimum should be kept. For this application, we recommend a 10 kcal/mol cutoff value using the MMFF force field.

- 3) It is important to respect the suggested theory levels, since DP4+ and MM-DP4+ were optimized for those levels. If the desired theory level is not parameterized, you can generate your own level following the instructions of the Custom-DP4+ method.
- 4) Using unassigned or misassigned NMR data can lead to erroneous results. The chemical shifts of equivalent nuclei that show fast interconversion should be averaged (such as the case of methyl groups, or some methylene groups). Treating the signal of each individual proton independently is wrong (for example, computing three different chemical shifts for the same methyl group). Another problem arises when dealing with diastereotopic methylene protons, which are often arbitrarily correlated. Unless the discrimination of both signals as pro-R and pro-S is made using additional NMR information (such as NOE or J coupling), the most convenient way to tackle this issue is to treat them as interchangeable signals. Follow the instructions to learn how to deal with these issues.

Probability calculations: DP4+, MM-DP4+ and Custom-DP4+

Prepare your files

In order to run a correlation calculus, DP4+ App needs the selection of a working directory and an Excel file. The program has a series of controls to ensure correct data entry, so as errors in the code are avoided.

The Excel file (.xlsx) must contain the information in the "shifts" sheet. This will be the only sheet read by the program and must have the structure defined in figure 1 (see *Warnings and Input control*). The columns headers must be the same. For isomers with same labels, only three columns need to be used. In case isomers use different labels, there should be three labeling columns (label 1|label 2|label 3) for each candidate. The name of this document does not have any requirement, since it will be selected individually.

The following columns are intended to place the correlation labels

Atom
type

Experimental
chemical
shifts

if

All candidates with the same labels
Only 3 columns for all candidates

Candidates with different labels
3 columns for each candidates

	A	B	C	D	E	F	G	H	I	J	K
1	index	nuclei	sp2	exp_data	exchange	label 1	label 2	label 3	label 1	label 2	label 3
2	1	C		73.7		7					
3	2	C		46.5		8					
4	3	C	1	175.5		9					
5	4	C		11.0		13					
6	5	C	1	141.5		4					
7	6	C	1	125.9		1	5				
8	7	C	1	128.1		2	3				
9	8	C	1	127.3		6					
10	9	C		60.5		11					
11	10	C		13.9		15					
12	11	H		5.09		21					
13	12	H		2.77		22					
14	13	H		1.12		26	27	28			
15	14	H	1	7.29		16	19				
16	15	H	1	7.29		17	18				
17	16	H	1	7.29		20					
18	17a	H		4.12	a	23	24				
19	17b	H		1.21	a	29	30	31			

Experimental
index

sp2 nuclei must be
indicated with
charater "1"

Interchangeable
signals must be
pared with letters

shifts

Sheet name

Figure 1. Excel sheet (experimental information and correlation labels)

The Gaussian files must be "nmr" calculations results (.log or .out). The labeling must be as follows: $n_m_nmr.log$, where n is the isomer id, m the conformer number and $*$ a user annotation.

The selection will be made by pressing the "Select . . ." buttons via popup windows. We strongly recommend using the given example ("Create Example" button) as a template to build your working directory.

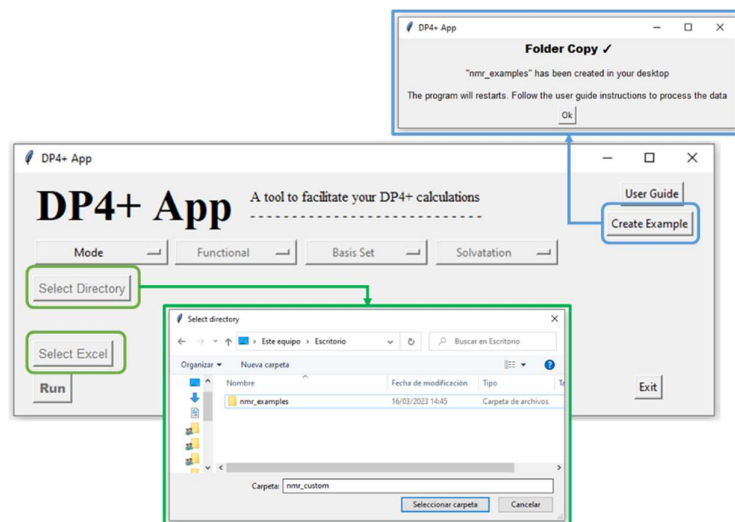


Figure 2. Entry buttons and example button

Perform a calculation

With DP4+ App it is possible to determine the probability of correlation at 60 already parameterized levels of theory. They arise from the combination of various functionals, basis set and solvation modes. Of the total, 24 levels were parameterized from geometries optimized with quantum mechanics at the B3LYP/6-31G* (*QM mode*) and the remaining 36, through molecular mechanics with the MMFF force field (*MM mode*).

QM mode theory levels combinations

Functional	Basis Set	Solvation
B3L	6-31G(d)	GAS
mPW1PW91	6-31+G(d,p)	PCM (CH ₃ Cl)
	6-311G(d,p)	
	6-311+G(d,p)	

MM mode theory levels combinations

Functional	Basis Set	Solvation
B3L	6-31G(d,p)	GAS
M06-2x	6-31+G(d,p)	PCM (CH ₃ Cl)
mPW1PW91	6-311+G(d,p)	SMD (CH ₃ Cl)
wB97XD		

Although it is allowed to carry out calculations at any selected level, for a better use of DP4+ App, the program controls the coincidence between the command lines of the Gaussian files with the selected theory level. In case the levels are not matched a warning will pop up, but it will not prevent the calc execution (figure 3).

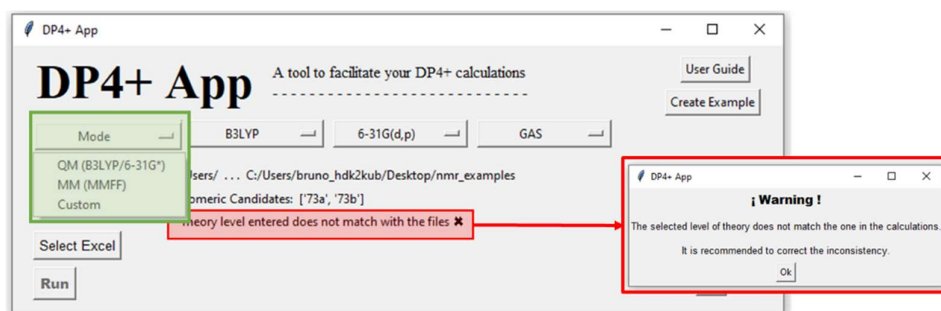


Figure 3. Modes selection and example of miss matching theory level and command lines

To perform calculations with a different theory level than those mentioned, you must first parameterize it following the instructions of the *Custom mode* (next section). For this mode, the matching of the command lines is not checked automatically, but it is possible to do it yourself on the final results sheet.

Results output

A pop-up will indicate the calculation has finished correctly and the results will be presented in an Excel file inside the selected working folder. The name of the output will correspond with the calculation mode used.

There will be five sheets, one with the probability results, two with the chemical shifts and two with the correlation errors. In the main sheet ("**results**"), you will find the probabilities of the candidates classified by their nuclei, scaling and the full version. In addition, the selected theory level, the command line of the Gaussian calculations and the automatic coincidence check are printed.

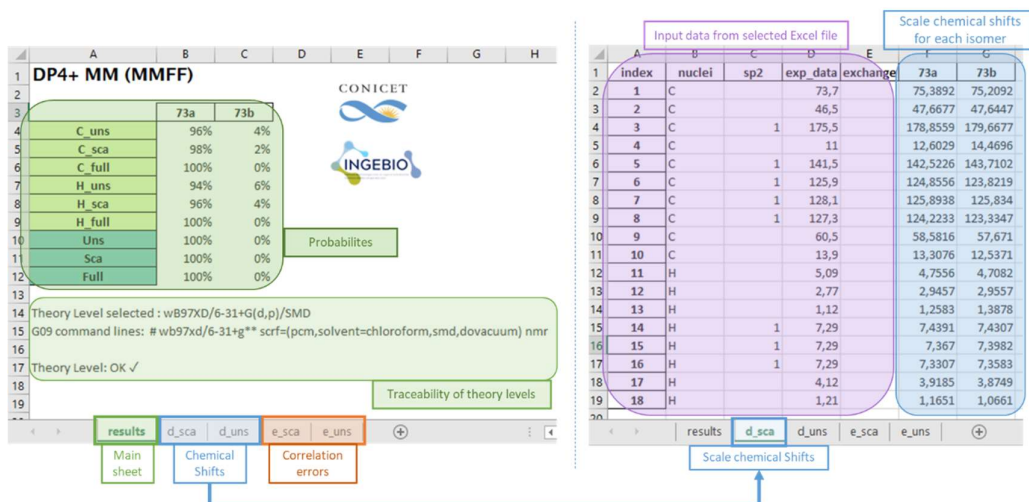


Figure 4. Output Excel file

In cases where the selected theory level does not coincide with the Gaussian calculation command line, it will warn about the misuse of the tool and the inconsistencies found (figure 5).

The *Custom mode* will indicate that the theory level cannot be compared and it is up to the user to carry out this control. In addition, below the results, the data base of the Custom level used are reported, including the standard tensors, the distribution parameters, the date and method of parametrization.

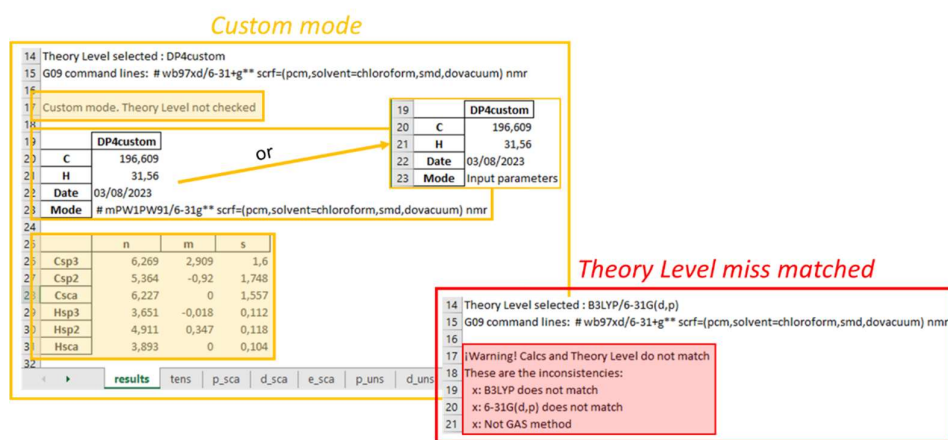


Figure 5. Examples of Custom mode traceability and theory level miss selection

Reparametrization: Custom-DP4+

Create a new level

Within the *Custom mode* of the main window, there is the "+ new" option that will redirect you to the reparameterization module. There you have to select the parameterization mode and assign a name to your custom level. The name can only have numbers and lowercase letters (do not use special characters either).

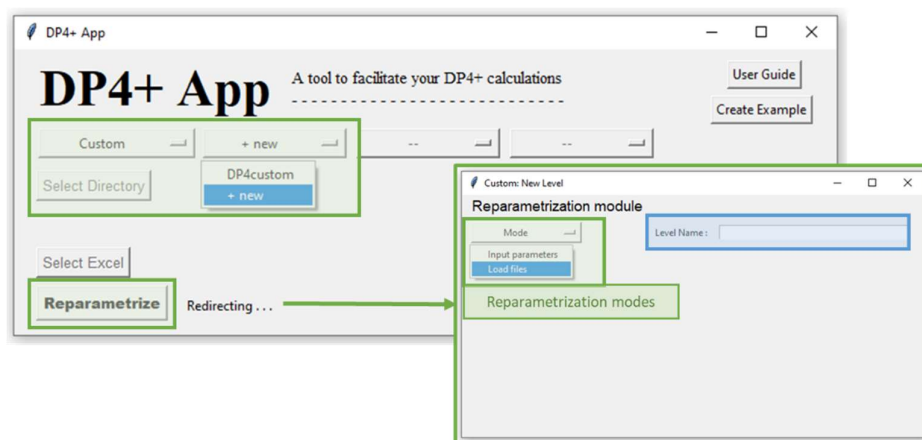


Figure 6. Selection to create a new custom level and redirecting window.

In the case you have already parameterized a theory level, you can enter the distribution parameters and tensors of the TMS by keyboard (Figure 7.a).

If, on the other hand, the distribution statistics need to be calculated, you must load a working directory and the Excel file likewise in the main window. The parameters will be determined by the app and you can see them when the custom level is used in a calculation.

The working folder must contain the nmr calculations of all the conformations of training set (table 1) and the standard TMS. To facilitate their preparation, the input files for Gaussian calculations can be downloaded in both, MM (force field MMFF) and QM (theory level B3LYP) optimization. The command line must indicate the level of theory as specified by Gaussian and the calculation instruction: "nmr" (Figure 7.b). The training set files must follow the same labeling rules used for calculations.

The TMS file should be named as: *O_*_tms.log*. It must not have the string "nmr" so as it is going to be confused as a training molecule. Its template will be downloaded too.

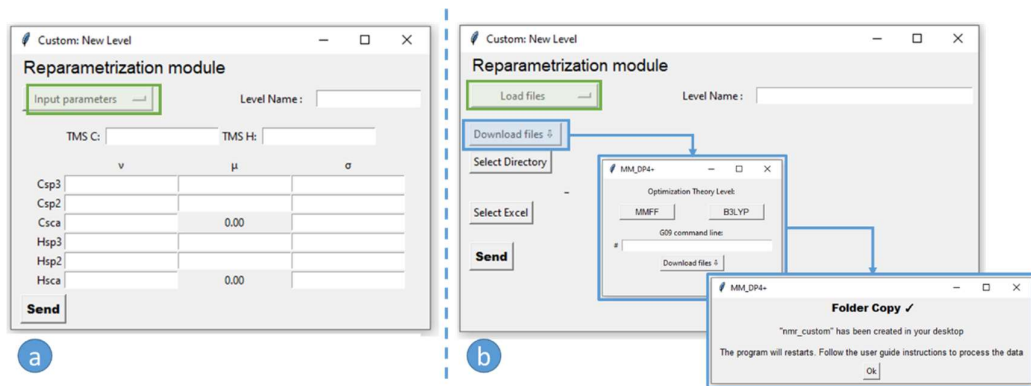


Figure 7. a) Input parameters mode: allow to entry the values of an already parametrized level. b) Load mode: capable of automatically determine the distribution parameters using the training set calculus. Also, offers templates of the training set in QM y MM optimization.

The Excel file with the experimental data and the correlation labels already assigned for will be provided with the Gaussian inputs. In it, the data set of each molecule is placed in a sheet with the associated id. The information follows the same structure as the correlation “**shifts**” sheet in figure 1.

Table 1. Training set molecules with experimental labels.

Update a level

To update a level simply follow the steps in the previous section and overwrite the name of the desire custom level. A popup will warn you about the existence of that level before updating the data. You can go back and change the name if you want to keep both parametrizations (figure 8). The update can be generated in any mode, it is not necessary to use the same one as the original.

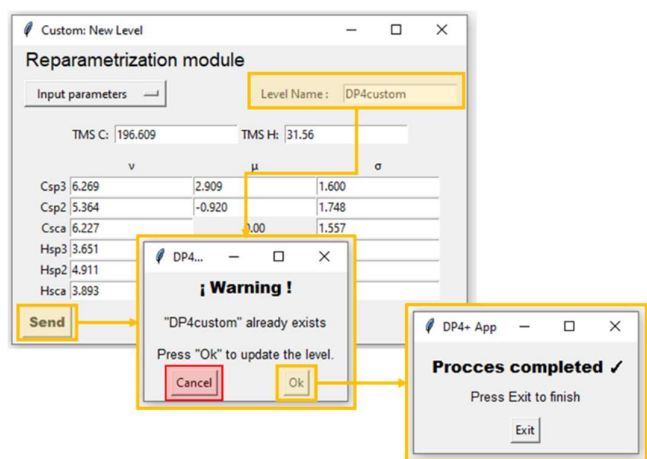


Figure 8. Example of update/overwriting a custom level

Warnings and Input control

A warning system has been incorporated to assist the user in interpreting anomalous results of DP4+ type calculations. Besides, DP4+ App has several checkpoints to ensure correct data entry. For cases that do not meet the requirements of the program, it will alert about the inconsistencies.

Questionable values

Although DP4+ App can perform calculations with any given input (which follows the requested format), there are some chemical criteria that must be taken into account for a result to be valid.

Those calculations that do not meet the following requirements will be warned about said deviations:

- $\sigma_H > 6\text{ppm}$ and $\sigma_C > 120\text{ppm}$, not marked as *sp2*
- $\sigma_H > 14\text{ppm}$, identified as ^{12}C
- $e_{\text{sca-H}} > 0.7$ and $e_{\text{sca-C}} > 10$, related to possible miscorrelation/miss assignment

The calculations can be carried out normally and the warnings will be found as highlighted cells. For DP4+ type calculations, they are printed on the *e_sca* sheet of results. For parameterization, a popup will ask if you want to continue and warnings will be found in the input spreadsheet.

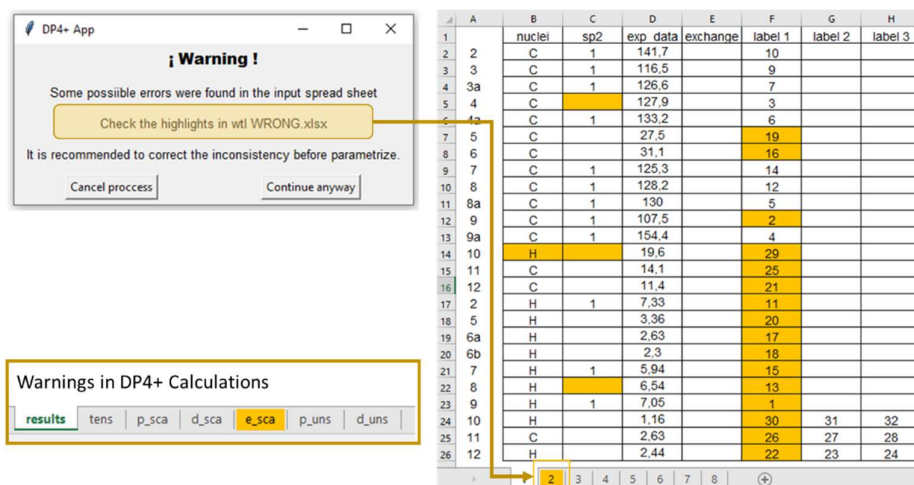


Figure 9. Example of deviant values for parametrization method

Gaussian calculation files

To ensure that the information provided by the Gaussian outputs is complete, it must be verified that their last line indicates *Normal Termination*. Those files where this string cannot be found will be separated in a *Removed Files* folder, then a popup will let you decide if you want to continue without those files or cancel to recalculate.

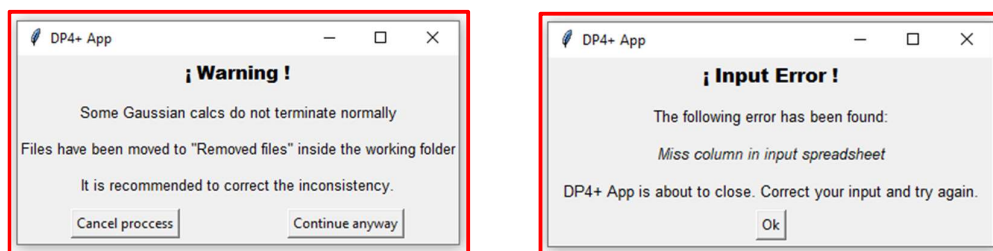


Figure 10. Examples of warning and input error

Data spread sheet

The provided spreadsheet must obey the figure 1 format, as mentioned above. The checkpoints for this file are:

- Column not found: if a header is not found
- Data not found: if miss data in columns *nuclei*, *exp_data*, *labels*
- Incorrect data: for *nuclei* data must be C or H
for *exp_data* must be number
for *labels* must be integer number
- Label index out of range: in case the label do not match any nuclei in the Gaussian calculation matrix
- Different number of candidate isomers and set of labels: in cases where multi label is used
- Mismatched diastereotopic labels

When one of these situations occurs, the program will not be able to perform the calculation. You will need to correct the inconsistency.

Malfunctions report

If you find a faulty operation of DP4+ App, please report your situation in detail to the following emails:

- brunoafranco@uca.edu.ar
- zanardi@inv.rosario-conicet.gov.ar
- sarotti@iquir-conicet.gov.ar