TUTORIAL FOR NMR FINGERPRINTING TOOL

1 Overview and required spectra

In this tutorial, we give step-by-step instructions on how to use the NMR Fingerprinting (NMR-FP) tool provided on this website, which allows an automated prediction of the structural groups in unknown samples based only on standard NMR spectra of the samples. Specifically, for each studied sample, the following spectra are required:

- ^{1}H NMR spectrum
- ^{13}C NMR spectrum (optionally: quantitative)
- ^{13}C DEPT NMR spectra with 90° and 135° phase angle

Furthermore, using prior information about the general presence or absence of labile protons in each sample can be used, which is recommended if the respective information is available. However, if this information is not available, the tool can also be used by selecting the appropriate option as described below. In the present version of the NMR-FP tool, thirteen structural groups can be distinguished, which are summarized in Table 1; furthermore, the tool can also differentiate between multiple substitution degrees for many of the groups, which is also indicated in Table 1.

Table 1: Structural groups that can be distinguished by the NMR-FP tool. The last column denotes the substitution degree, i.e., primary (P), secondary (S), tertiary (T), and quaternary (Q) carbon atoms that can be differentiated for the respective structural group.

Label	Group name	Distinguished substitution degrees
$\mathrm{CH_3}$	Methyl	P
CH_x	Alkyl	S, T, Q
cyCH_x	Cyclic alkyl	S, T, Q
$\mathrm{CH}_x\mathrm{OH}$	Alcohol	P, S, T, Q
$\mathrm{CH}_x\mathrm{O}$	Ether	P, S, T, Q
$\mathrm{CH}_x{=}$	Aliphatic double bond	S, T, Q
$\mathrm{CH}_x^{\mathrm{ar}}{=}$	Aromatic carbon	T, Q
$\mathrm{RO{-}CH}_x^{\mathrm{ar}}{=}$	Aromatic carbon with oxygen substituent	T, Q
COOR	Ester/lactone/anhydride carbonyl	T, Q
ROOCH_x	Alkyl next to ester/lactone oxygen	P, S, T, Q
COOH	Carboxylic acid	T, Q
$\mathrm{CO^{ald}}$	Aldehyde	T
$\mathrm{CO^{ket}}$	Ketone	Q

There are generally no constraints regarding the studied samples. However, if a sample contains structural groups that are not listed above, they cannot correctly be identified by the NMR-FP tool. However, since the predictions are based on chemical shifts, it can be expected that chemically similar structural groups from Table 1 are predicted in this case.

2 Preprocessing of NMR spectra

All NMR spectra should be preprocessed in a standard manner, including phase and baseline correction. The chemical shift of the NMR spectra should be fixed to a reference peak with a known chemical shift relative to a referencing standard such as Tetramethylsilane (TMS) or Trimethylsilylpropanoic acid (TMSP). We furthermore recommend using an apodization of the spectra and a subsequent (automatic) peak picking routine to determine the number and the position of all peaks in the spectra. In Figures 1 and 2, examples for suitably preprocessed ^{13}C and ^{1}H NMR spectra are shown.

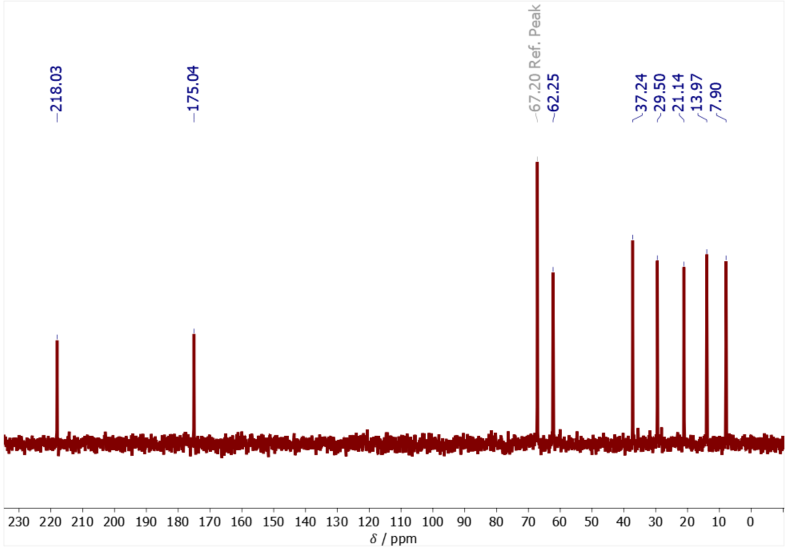


Figure 1: Preprocessed ^{13}C NMR spectrum of an exemplary mixture. The chemical shift of the known reference component, which was added as standard here, was set to 67.20 ppm relative to TMSP.

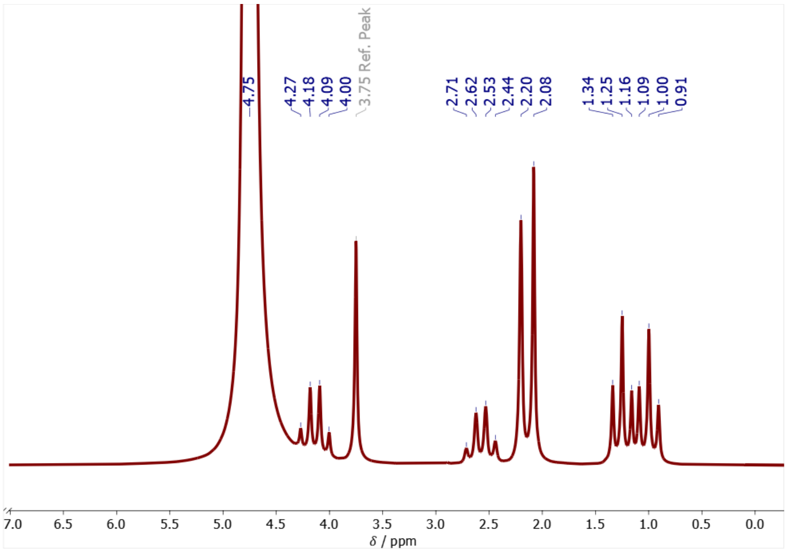


Figure 2: Preprocessed ^{1}H NMR spectrum of an exemplary mixture. The chemical shift of the known reference component, which was added as standard here, was set to 3.75 ppm relative to TMSP.

If an automatic peak picking technique is used, it is advisable to review the results to prevent artifacts that could impair the predictions of the NMR-FP tool. If a quantitative evaluation is requested, the peak areas in the ^{13}C NMR spectrum can be determined either by direct integration of clearly separated peaks or by using a peak fitting routine.

3 Obtaining information on the substitution degrees of the carbon atoms

The spectra obtained from the DEPT experiments have to be processed in the same manner as the ^{13}C NMR spectrum of the sample. Information on the substitution degree of each carbon atom can then be determined manually based on the ^{13}C DEPT NMR experiments with 90° and 135° pulse angle (called DEPT 90 and DEPT 135, respectively, in the following). Specifically, the intensities, including their signs, of all peaks in these spectra need to be considered together with the scheme described in Table 2. Figure 3 shows an example.

Table 2: Typical intensities of peaks in DEPT 90 and DEPT 135 spectra depending on the substitution degree of the associated carbon atom. '+' indicates positive peak intensity, '0' indicates very low intensity compared to the same peak in the ^{13}C NMR spectrum, and '-' indicates negative peak intensity.

	Primary	Secondary	Ternary	Quaternary
Abbreviation	P	S	T	Q
No. of bonded hydrogen atoms	3	2	1	0
Peak intensity in ^{13}C spectrum	+	+	+	+
Peak intensity in DEPT 90	0	0	+	0
Peak intensity in DEPT 135	+	-	+	0

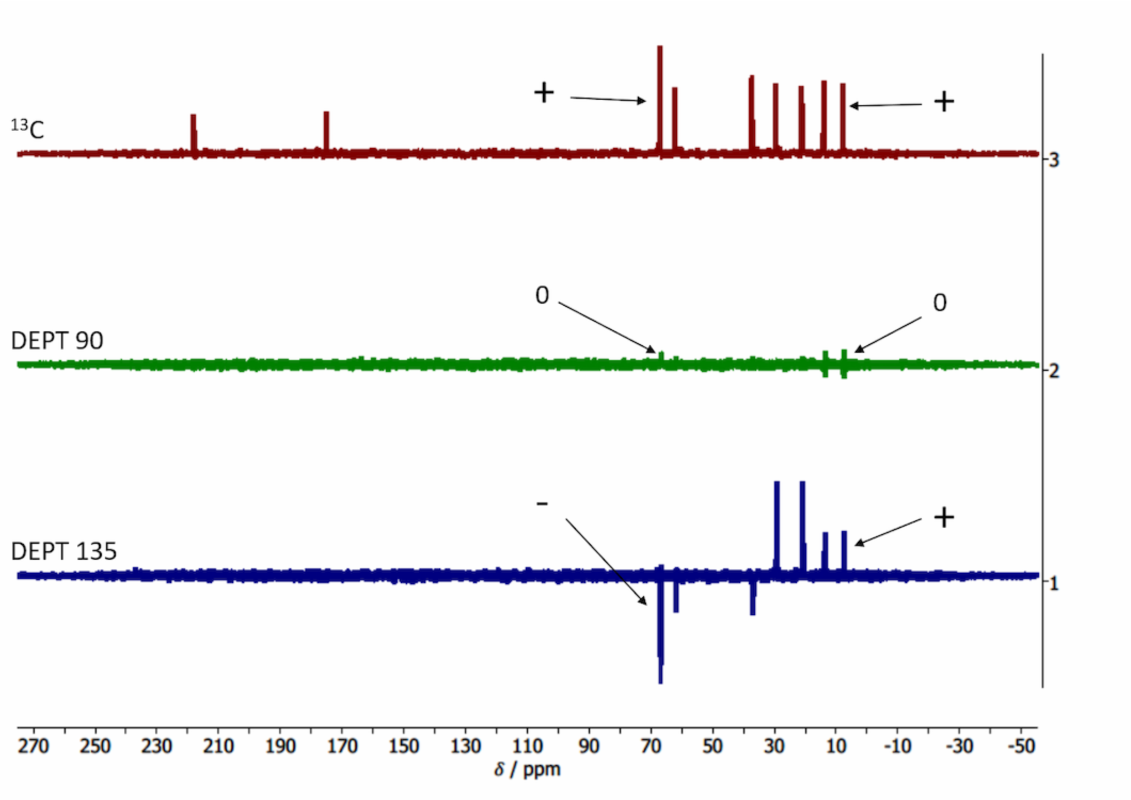


Figure 3: ^{13}C and DEPT NMR spectra of an exemplary mixture. For some peaks, the intensities are indicated by arrows for demonstration. The other intensities can be determined analogously.

There are two options for providing information on the substitution degrees in the NMR-FP tool:

- Direct provision of the substitution degree of each carbon as determined from the DEPT NMR spectra as described above (or in any other way, e.g., based on heteronuclear single quantum coherence spectroscopy (HSQC)-multiplicity edited experiments)
- Provision of the peak intensity ('+', '0', '-') of each carbon in the DEPT 90 and DEPT 135 NMR spectra

4 Detection of labile protons

While labile protons can often directly be recognized by the user based on the ^{1}H NMR spectrum due to their usually broad peak form, there are also other ways to retrieve information about the general presence or absence of labile protons in a sample, e.g.:

- HSQC experiment

- So-called D_2O shake
- pH measurement

Using ^{1}H-^{13}C-HSQC experiments is particularly helpful if a sample contains only the elements carbon (C), hydrogen (H), and oxygen (O), since labile protons are only present in hydroxyl groups. As an example, the HSQC spectrum of such a mixture is shown in Figure 4, where the ^{1}H dimension is plotted on the horizontal and the ^{13}C dimension is plotted on the vertical axis. To identify labile protons, the correlation between proton peaks and carbon peaks needs to be considered. In Figure 4, the proton peak at 11.8 ppm (indicated by the magenta line) does not show a correlation with any carbon peak and can, thus, be classified as a labile proton.

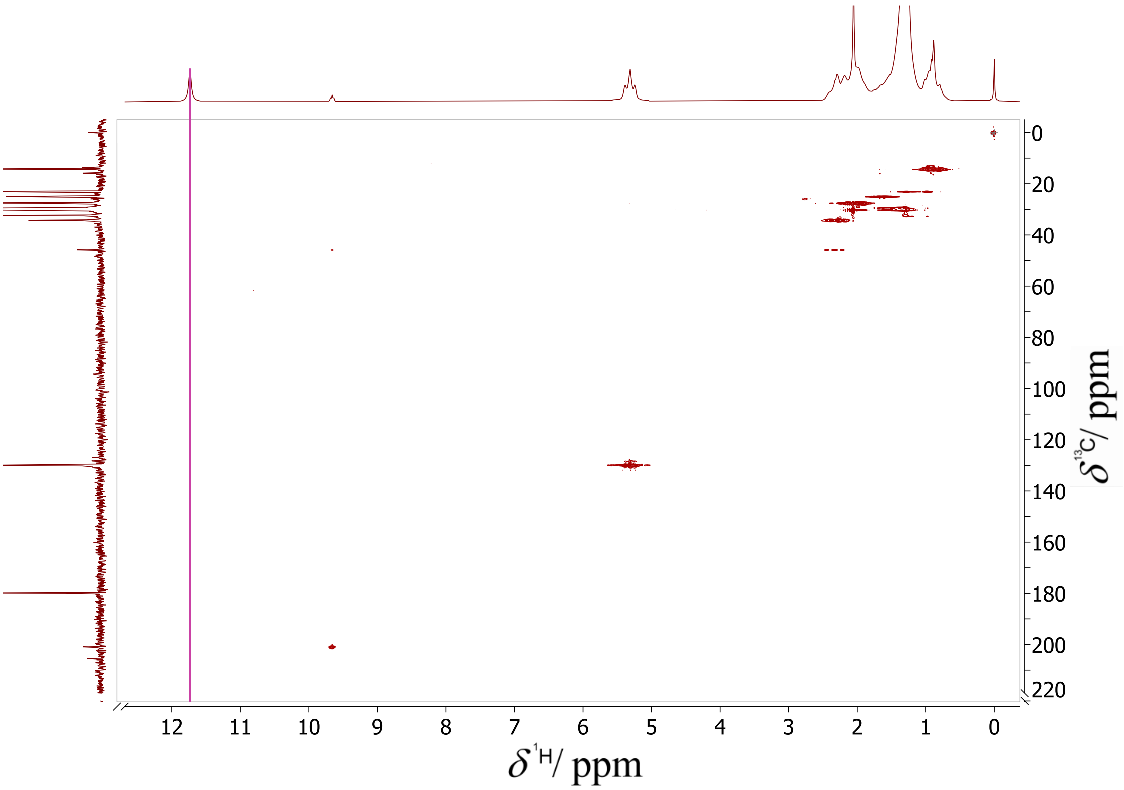


Figure 4: HSQC spectrum of an exemplary mixture containing only C, H, and O atoms. The magenta line indicates a proton peak not showing any correlation with any carbon peak.

Using the D_2O shake approach requires adding a small amount of deuterated water to the sample, which causes the peaks of labile protons to vanish or at least strongly change their position and/or amplitude in the ^{1}H NMR spectrum in comparison to the spectra taken before the addition of D_2O. Measuring the pH value is especially suitable for the detection of carboxylic acids.

5 Using the NMR Fingerprinting tool

There are two general ways to use the NMR-FP tool, which differ in the form in which the above-described input data are provided:

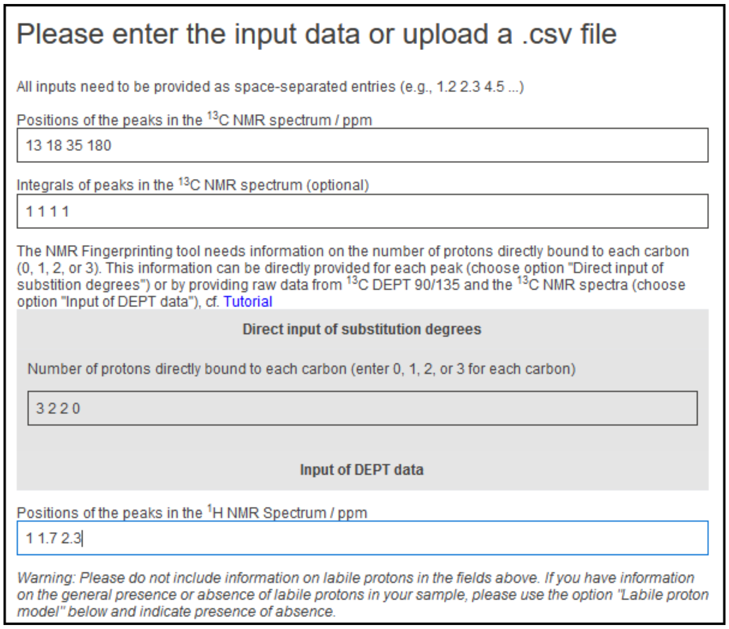
a) Via the user interface on the "Application" page of the website

b) Via uploading a .csv file

a) Input via user interface

In Figures 5 and 6, the input section of the user interface, including exemplary input data, is shown. In each field, the appropriate information on the peaks in the respective NMR spectra needs to be entered. The second field is optional and only required if quantitative results are requested.

All entries need to be separated by spaces. It is important that the number of peaks in the ^{13}C NMR spectrum, for which information is entered in the first three fields of the user interface, is the same and that the order of the peaks is always the same. The information on the substitution degree of the carbon atoms can be provided in two ways, as described in Section 3 above.



Please enter the input data or upload a .csv file

All inputs need to be provided as space-separated entries (e.g., 1.2 2.3 4.5 ...)

Positions of the peaks in the ^{13}C NMR spectrum / ppm

13 18 35 180

Integrals of peaks in the ^{13}C NMR spectrum (optional)

1 1 1 1

The NMR Fingerprinting tool needs information on the number of protons directly bound to each carbon (0, 1, 2, or 3). This information can be directly provided for each peak (choose option "Direct input of substition degrees") or by providing raw data from ^{13}C DEPT 90/135 and the ^{13}C NMR spectra (choose option "Input of DEPT data"), cf. Tutorial

Direct input of substitution degrees

Number of protons directly bound to each carbon (enter 0, 1, 2, or 3 for each carbon)

3 2 2 0

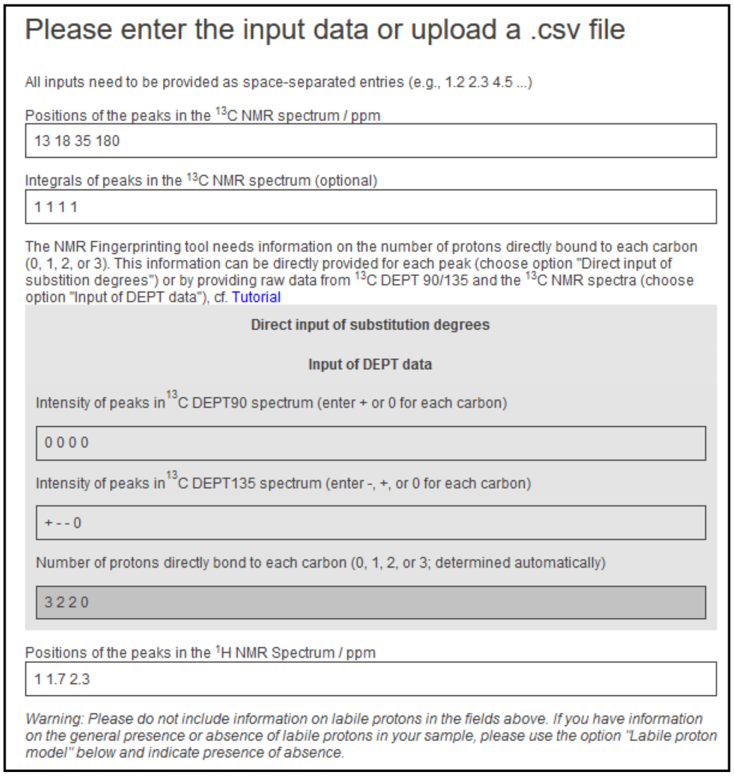
Input of DEPT data

Positions of the peaks in the ^{1}H NMR Spectrum / ppm

1 1.7 2.3

Warning: Please do not include information on labile protons in the fields above. If you have information on the general presence or absence of labile protons in your sample, please use the option "Labile proton model" below and indicate presence of absence.

Figure 5: Screenshot of the user interface of the NMR-FP with exemplary input data and using the option to directly provide the information on the substitution degree of each carbon atom.



Please enter the input data or upload a .csv file

All inputs need to be provided as space-separated entries (e.g., 1.2 2.3 4.5 ...)

Positions of the peaks in the ^{13}C NMR spectrum / ppm

13 18 35 180

Integrals of peaks in the ^{13}C NMR spectrum (optional)

1 1 1 1

The NMR Fingerprinting tool needs information on the number of protons directly bound to each carbon (0, 1, 2, or 3). This information can be directly provided for each peak (choose option "Direct input of substition degrees") or by providing raw data from ^{13}C DEPT 90/135 and the ^{13}C NMR spectra (choose option "Input of DEPT data"), cf. Tutorial

Direct input of substitution degrees

Input of DEPT data

Intensity of peaks in ^{13}C DEPT90 spectrum (enter + or 0 for each carbon)

0 0 0 0

Intensity of peaks in ^{13}C DEPT135 spectrum (enter -, +, or 0 for each carbon)

+ - - 0

Number of protons directly bond to each carbon (0, 1, 2, or 3; determined automatically)

3 2 2 0

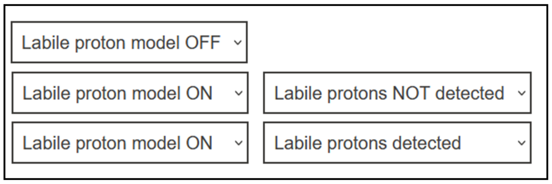
Positions of the peaks in the ^{1}H NMR Spectrum / ppm

1 1.7 2.3

Warning: Please do not include information on labile protons in the fields above. If you have information on the general presence or absence of labile protons in your sample, please use the option "Labile proton model" below and indicate presence of absence.

Figure 6: Screenshot of the user interface of the NMR-FP with exemplary input data and using the option to provide the peak intensities ('+', '0', or '-') from the DEPT 90 and DEPT 135 NMR spectra.

In Figure 7, all possible options for the inclusion of information on labile protons are shown. If the information on labile protons is available, the "Labile proton model" can be used, and the presence or absence should be indicated by selecting the appropriate option, cf. Figure 7. Note that if the labile proton model is used, the chemical shifts of labile protons should not be included in the peak list. If no information on the presence or absence is available, the option "Labile proton model OFF" should be used, cf. Figure 7.



Labile proton model OFF

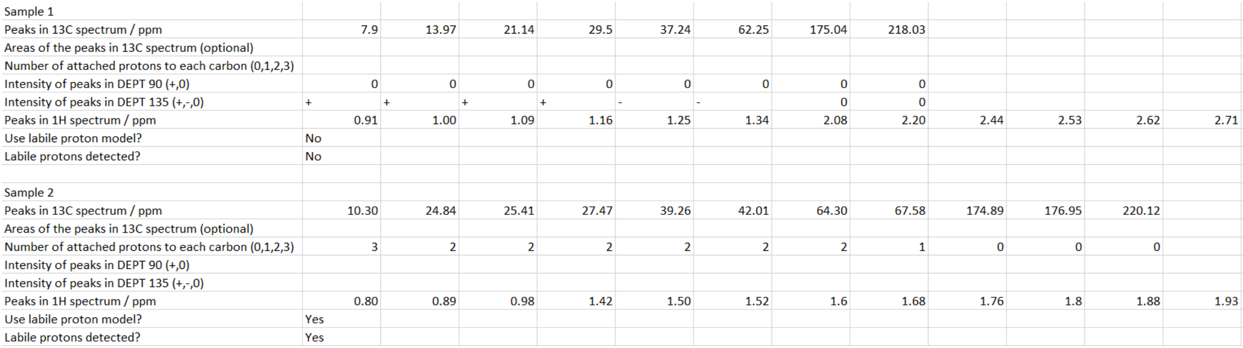
Labile proton model ON — Labile protons NOT detected

Labile proton model ON — Labile protons detected

Figure 7: Possible appearances of the user interface for including information on the presence or absence of labile protons in the studied sample.

b) Input via .csv file

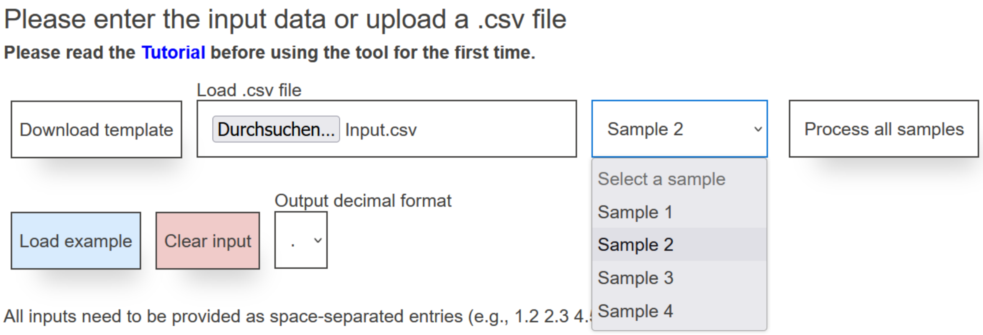
As an alternative, the input data can also be provided in the form of a .csv file. For this purpose, the template available for download on the "Application" page should be used, in which the respective information needs to be entered as shown in Figure 8. Note that if the labile proton model is used, the chemical shifts of labile protons should not be included in the peak list in the .csv file. Using the template also allows processing multiple samples at once, which should be separated by blank lines, cf. Figure 8. The predictions of the structural groups for multiple samples can be downloaded by using the "Process all samples" button. Also, via the .csv file, both above-described ways of providing information on the substitution degrees of the carbon atoms can be used, cf. Figure 8.



Sample 1												
Peaks in 13C spectrum / ppm	7.9	13.97	21.14	29.5	37.24	62.25	175.04	218.03				
Areas of the peaks in 13C spectrum (optional)												
Number of attached protons to each carbon (0,1,2,3)												
Intensity of peaks in DEPT 90 (+,0)	0	0	0	0	0	0	0	0				
Intensity of peaks in DEPT 135 (+,-,0)	+	+	+	+	-	-	0	0				
Peaks in 1H spectrum / ppm	0.91	1.00	1.09	1.16	1.25	1.34	2.08	2.20	2.44	2.53	2.62	2.71
Use labile proton model?	No											
Labile protons detected?	No											
Sample 2												
Peaks in 13C spectrum / ppm	10.30	24.84	25.41	27.47	39.26	42.01	64.30	67.58	174.89	176.95	220.12	
Areas of the peaks in 13C spectrum (optional)												
Number of attached protons to each carbon (0,1,2,3)	3	2	2	2	2	2	2	1	0	0	0	
Intensity of peaks in DEPT 90 (+,0)												
Intensity of peaks in DEPT 135 (+,-,0)												
Peaks in 1H spectrum / ppm	0.80	0.89	0.98	1.42	1.50	1.52	1.6	1.68	1.76	1.8	1.88	1.93
Use labile proton model?	Yes											
Labile protons detected?	Yes											

Figure 8: Screenshot of the .csv template for providing the input for the NMR-FP tool. The template allows for providing information for multiple samples at the same time. For sample 1, the peak intensities in the DEPT NMR spectra are provided, whereas, for sample 2, the substitution degree of each carbon is directly provided.

After uploading the .csv file, the sample of interest can then be selected in the dropdown menu as shown in Figure 9 (if the .csv file contains information on multiple samples).



Please enter the input data or upload a .csv file

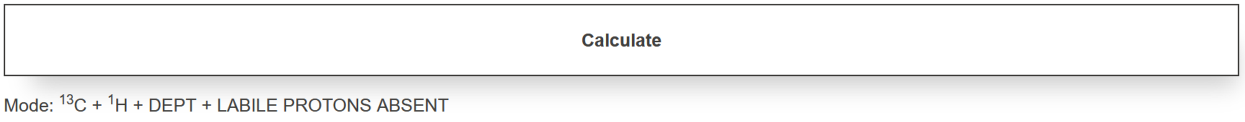
Please read the Tutorial before using the tool for the first time.

Load .csv file: Input.csv

Sample 2

All inputs need to be provided as space-separated entries (e.g., 1.2 2.3 4.

Figure 9: Screenshot of the "Application" page with the selection of one of four samples provided via the uploaded "Input.csv" file.



Mode: ^{13}C + ^{1}H + DEPT + LABILE PROTONS ABSENT

Figure 10: Screenshot of the "Application" page with the "Calculate" button at the bottom of the website.

In any case, after providing the input data to the NMR-FP tool, the output for one sample is obtained by pressing the "Calculate" button at the bottom of the website, cf. Figure 10.

6 Retrieval and interpretation of results

The NMR-FP tool generates the following output for each processed sample:

- The predicted structural groups assigned to all peaks in the ^{13}C NMR spectrum
- Information on the (internal) decision function of the classification algorithm
- The predicted group mole fractions in the studied sample

By default, only the predicted structural groups and their assignment to the peaks in the ^{13}C NMR spectrum are shown, while the detailed output (decision function and group mole fractions) can be retrieved by clicking the "Additional Information" button. Figures 11 and 12 show examples. The results can be downloaded as a .csv file by clicking on the respective download buttons beneath the plots, cf. Figures 11 to 13.

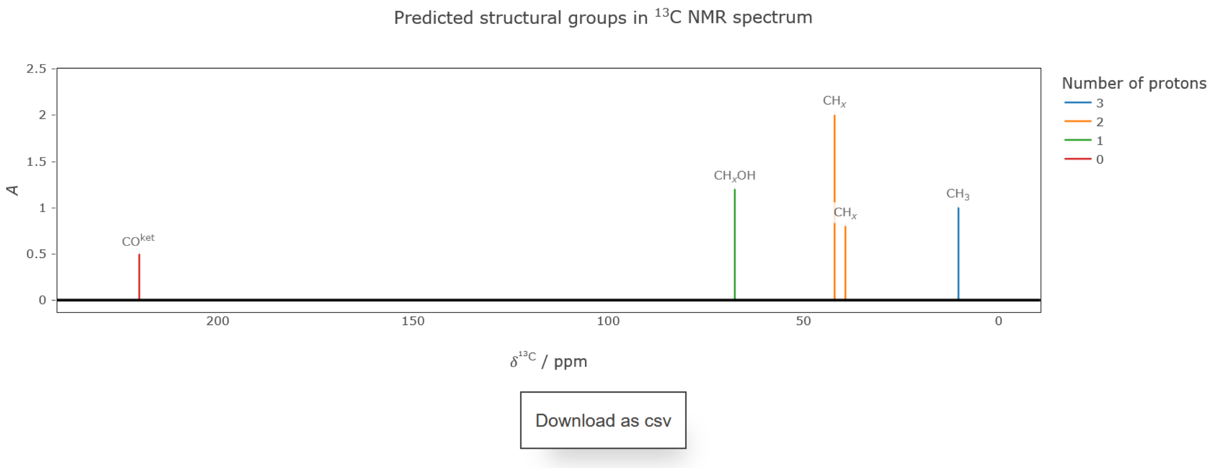


Figure 11: Prediction of structural groups and assignment to peaks in the ^{13}C NMR spectrum of an exemplary sample with the NMR-FP tool. The area A of the peaks is indicated by their height.

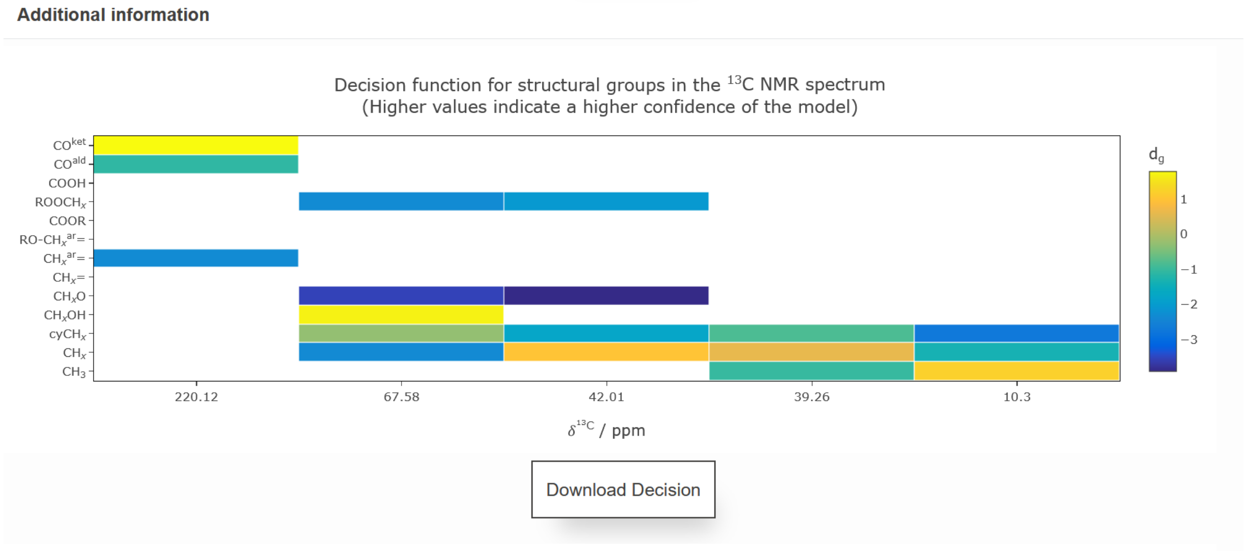


Figure 12: Example for additional output information on the decision function. Larger values of the decision function indicate larger model confidence for the respective predictions.

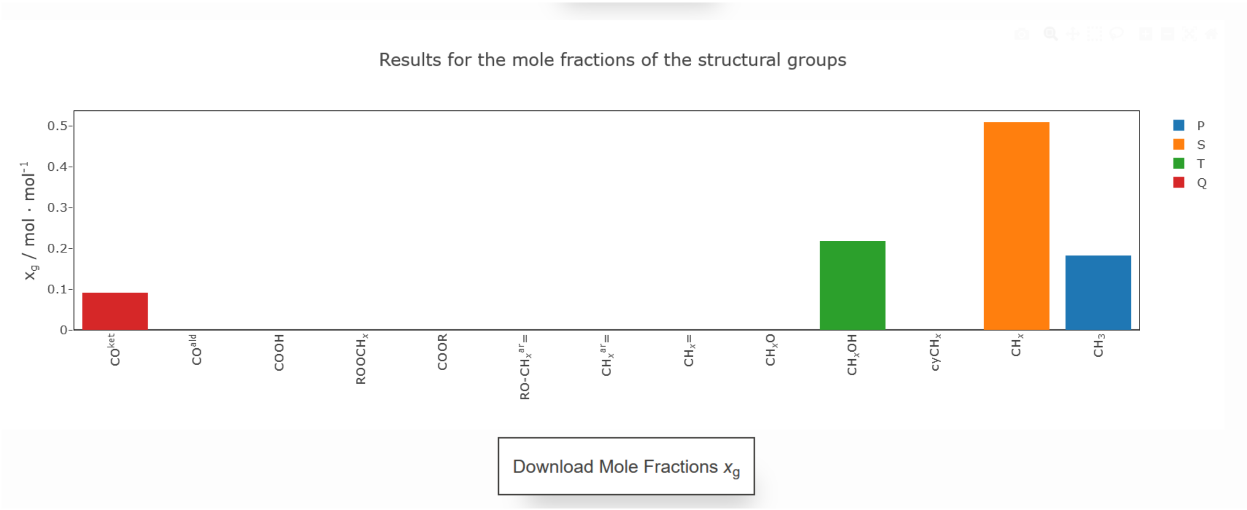


Figure 13: Example for additional output information on the predicted structural group mole fractions.