

Diffusion NMR and DOSY

1. Introduction

The diffusion of molecules in solution depends on their size and shape. This can be described quantitatively using the self-diffusion coefficient, D , which is inversely related to size; larger molecules have smaller diffusion coefficients, and vice-versa. Diffusion coefficients for small molecules (<1000 g/mol) typically range from $1 \cdot 10^{-10}$ to $1 \cdot 10^{-8}$ m^2s^{-1} . Although there exist a variety of NMR sequences for measuring diffusion, they all share a common theme of refocussing magnetisation while using gradient pulses to apply a spatially-dependent phase to the signals. **Figure 1** shows the basic scheme for diffusion measurement using the PFG (Pulsed Field Gradient) spin-echo.

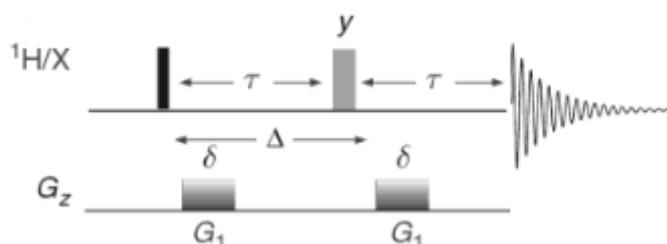


Figure 1. The PFG spin-echo sequence for measuring diffusion. The PFG (Pulsed Field Gradient) spin-echo is the most basic sequence for measuring diffusion. The diffusion during the period Δ is characterised by a series of measurements with increasing gradient pulse strengths (G_1) of fixed duration δ .

The first part of the sequence labels the nuclear spins with a phase factor that depends on their position in the NMR tube. There is then a delay Δ , during which diffusion occurs. Finally, a second gradient pulse re-focusses (or undoes) the spatial dependence of the first gradient pulse. If the molecule diffuses away from its initial position during the time period Δ , its signal(s) will not be completely refocussed, and the NMR signal(s) will be attenuated. The amount of signal that is refocussed at the end of the sequence therefore depends on the diffusion rate of the molecule; the greater the diffusion, the lower the NMR signal intensity. The Stejskal-Tanner equation describes quantitatively the signal intensity, and is shown (in simplified form) below:

$$I_G = I_{G=0} \exp \left[-(\gamma \delta G)^2 D \left(\Delta - \frac{\delta}{3} \right) \right]$$

where $I_{G=0}$ is the signal intensity at a gradient strength of zero, G is the gradient strength, D is the diffusion coefficient, δ ("little delta") is the gradient pulse duration, and Δ ("big delta") is the diffusion time. As the constants γ , δ , and Δ are known, the diffusion coefficient can be calculated.

To measure diffusion coefficients, a series of measurements are taken, where it is possible to progressively increase any of the gradient strength (G), the diffusion period (Δ), or the length of the gradient pulses (δ). Increasing Δ allows more time for the molecules to diffuse through the sample. Increasing G or δ imposes a greater degree of signal dephasing across the sample. In all cases, the amount of signal refocussed will decrease in the presence of self-diffusion, and the signal intensity will be lower. In practice, it is easiest experimentally to increase the gradient strengths (G) while keeping the diffusion period and gradient pulse lengths constant, and this is now standard practice for collecting diffusion data. The diffusion coefficient is then calculated from the Stejskal-Tanner expression using linear or non-linear regression analysis. The data may also be presented as a pseudo-

2D spectrum with chemical shift on the x axis and diffusion coefficient on the y axis; this is then known as Diffusion-Ordered Spectroscopy or DOSY.

2. Diffusion parameter sets

The following diffusion parameter sets are available, in 1D and 2D formats:

¹ H	• h1dosy_cc.crl	with convection compensation
	• h1dosy_cc_sup.crl	with convection compensation and solvent suppression
	• h1dosy.crl	no convection compensation
	• h1dosy_sup.crl	no convection compensation, with solvent suppression
¹⁹ F	• f19dosy_cc.crl	with convection compensation
	• f19dosy.crl	no convection compensation

The 1D experiments are required only for parameter optimisation; actual diffusion experiments are all collected from 2D parameters sets.

NOTES:

1. *DOSY is only available for ¹H and ¹⁹F nuclei by default.*
2. *The Achilles heel of diffusion measurements is convection, which is difficult to differentiate from diffusion but is unrelated to molecular size. It is therefore most important to allow for complete temperature equilibration of samples (> 20 mins) before attempting diffusion measurements. It is also necessary to keep sample volumes to a minimum to help avoid convection along NMR tubes. This is a greater problem for lower viscosity solvents.*
3. *When running diffusion NMR at temperatures away from ambient, the convection-compensated experiments should be used.*
4. *If there are issues with convection in your sample even when using the convection-compensated sequences (e.g. an unexpected value for D, or if there is a single diffusion coefficient where several are expected), it may be beneficial to use a 3mm tube instead of a 5 mm tube as this attenuates diffusion along the NMR tube.*

3. Experimental setup

3.1 Acquire a 1D spectrum

Setup and acquire a conventional 1D experiment, paying attention to the following parameters:

- A reasonably long relaxation delay is needed. The default **D1** is 5 seconds, and this should be the minimum value used. If the T₁ relaxation times are known, use a relaxation delay (**D1**) that is at least 3 x T₁ of the slowest-relaxing nucleus of interest, and ideally 5 x T₁.
- Use enough scans (**ns**) to obtain good signal:noise (use the “**expt**” command to estimate the experimental time).

3.2 Optimisation of Δ (d20) and δ (p30)

3.2.1 Run a 1D DOSY with 5% gradient strength

Type “ix” to create a new dataset. Load in the 1D version of the appropriate DOSY experiment. Change **D1**, **sw**, **ns**, **ds**, and **o1p** to match those of the 1D experiment obtained in step 2.1. Run the first experiment with a gradient strength (**gpz6**) of 5%. Make sure you run “rga” first. The following default settings are recommended (**figure 2**):

- **gpz6** (gradient strength) = 5 %
- **d20** (diffusion time, Δ) = **0.1s** (normal range is 0.01s up to shortest T1 of interest).
- **p30** (gradient pulse duration, δ) = **1000us** (**NEVER** go higher than 3ms! Ensure that $P30 / (D1+AQ) \leq 0.05$, or you can damage the probe!)
- **ns = 2 ds = 0** (if signal is weak, increase ns)

3.2.2 Run a 1D DOSY with 95% gradient strength

Type “iexpno” to copy the experiment to a new dataset. Set gpz6 = 95 and acquire a second spectrum using the same settings as the first spectrum (with the same **rg** value – i.e. do **NOT** run rga this time)

Overlay the two spectra by typing “.md”, or by pressing the  button. In the overlay window, select the first spectrum and manually scale it to match the second spectrum. The scaling factor should be around 5-10% (**figure 3**). If not, you will need to optimise the DOSY parameters, as described below.

3.2.3 Optimise DOSY parameters

If the intensity of the 95% gradient spectrum has not decreased sufficiently, re-run both experiments with a higher d20 and/or higher p30. Conversely, if the intensity has decreased too much, re-run the experiments with a lower d20 and/or lower p30. However, remember that **P30 should NEVER be higher than 3 ms, and P30/(D1+AQ) should NEVER be greater than 0.05**. It is normal to leave p30 as it is and start by altering d20. If you can get no further improvement by altering d20, you can then try altering p30.

PULPROG	steppgp1s1d	Pulse program for acquisition
TD	65536	Time domain size
NS	16	Number of scans
DS	2	Number of dummy scans
SWH [Hz]	10330.58	Sweep width in Hz
AQ [s]	3.1719923	Acquisition time
RG	4	Receiver gain
DW [μ s]	48.400	Dwell time
DE [μ s]	6.50	Pre-scan-delay
D1 [s]	1.00000000	Relaxation delay: $1.5 * T1$
D16 [s]	0.00020000	Delay for gradient recovery
D20 [s]	0.00000000	Diffusion time (big DELTA)
DELTA1 [s]	50000000.000000...	DELTA1=d20-p1*2-p2-p30*2-d16*3-p19
TDO	1	Dimension of accumulation loop
▼ Channel f1		
NUC1	1H	Nucleus for channel 1
P1 [μ s]	2.50	F1 channel - 90 degree high power pulse
p2 [μ s]	5.00	P2=p1*2
PL1 [dB]	2.00	F1 channel - power level for pulse (default)
PL1W [W]	11.56972027	F1 channel - power level for pulse (default)
SFO1 [MHz]	500.1330885	Frequency of observe channel
▼ Gradient channel		
GPNAM6	sine_100	SINE_100
GPNAM7	sine_100	SINE_100
GPZ6 [%]	0.00	1-100%
GPZ7 [%]	0.00	-17.13% (spoil)
P19 [μ s]	600.00	Gradient pulse 2 (spoil gradient)
P30 [μ s]	0.00	Gradient pulse (little DELTA*0.5)

Figure 2. Important acquisition parameters for 1D DOSY setup. The gradient strength (GPZ6) should be set according to the desired strength (typically either 5% or 95%). The diffusion time (d20, “big delta”) and the gradient pulse duration (p30, “little delta”) should be optimised.

3.3 Running the DOSY experiment

Create a new experiment by typing “*ixpno*” and load in the corresponding 2D version of the appropriate DOSY experiment. Set *d20* and *p30* to the values that you optimised previously. Set *ns* to be a multiple of 16 if using convection compensation, or 8 if not, ensuring that you will have a good signal:noise (as judged from the *gpz6=95%* experiment). To start the DOSY experiment, you must run the *dosy* au program **BEFORE** entering the “*zg*” command. Type “*dosy*” in the command line. A series of dialogue boxes will appear, prompting you to enter experimental details. The initial gradient value should be 5 and the final gradient value should be 95. The number of points should be >7 (typically 8-25), the default is 16. For the ramp type, choose “*q*” (for quadratic). You can then choose to start the experiment directly from the macro, or you can type “*zg*” to begin the acquisition (you can also type “*expt*” first to estimate how long the experiment will take).

Alternatively, simply type “*xau dosy 5 95 16 q y y y*” to run the *dosy* au program and start the experiment without interacting with the macro. In this case the gradient will ramp from 5 to 95, with 16 increments in a quadratic ramp. The final “*y*” instructs the macro to begin acquisition immediately.

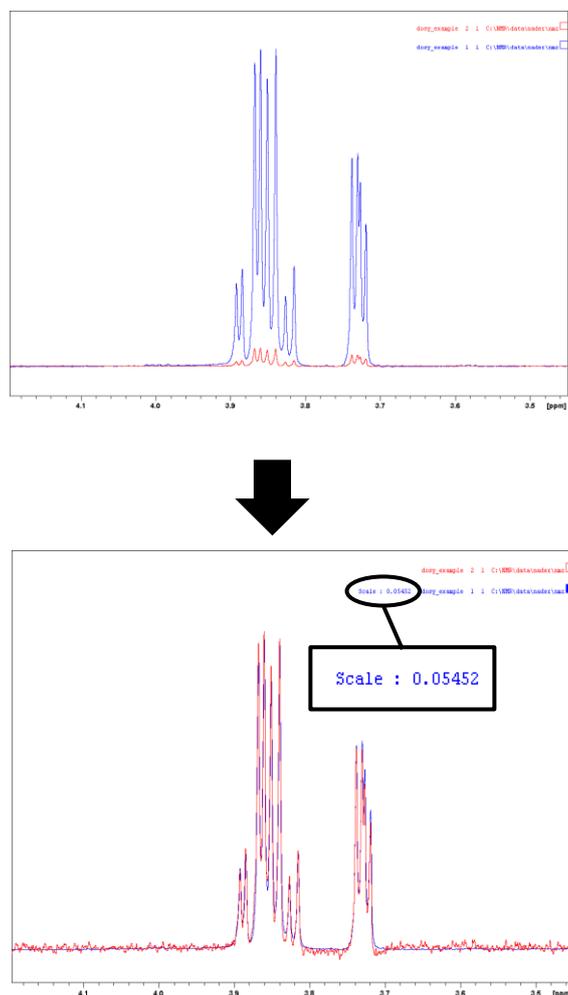


Figure 3. Optimisation of DOSY parameters. Overlay of the 5% gradient (blue) and 95% gradient (red) spectra. The scaling factor is approximately 5%, indicating that the choices of *d20* and *p30* are optimal.

4. Processing and analysing DOSY and diffusion data

4.1 Overview

Diffusion data can be analysed in two ways; regression analysis to calculate a diffusion coefficient, or DOSY analysis to generate a pseudo-2D plot with chemical shift on the x axis and diffusion coefficient on the y axis. In general, it is much easier to obtain a value for the diffusion coefficient using regression analysis than it is from a DOSY plot. It is also easier to estimate the error in the analysis, by looking at the variation in diffusion coefficients for multiple peaks from the same molecule. The alternative analysis method, DOSY, does not directly give a value for the diffusion coefficient, but instead provides a 2D spectrum, from which the coefficient can be estimated. It is good practice to use both methods for each diffusion analysis; regression analysis for obtaining accurate diffusion coefficients, and DOSY for obtaining a figure for reports / publications.

Although diffusion data can be analysed in either MestreNova or Topspin, we recommended that you use Topspin. The “T1/T2” module in Topspin can be used for regression analysis to calculate the best diffusion coefficient for each integral region or peak selected. The “DOSY” software can be used to generate a DOSY spectrum.

4.2 Processing the direct dimension

Before analysing the diffusion data, it is necessary to process the direct (F2) dimension suitably. Signal-to-noise is more important than resolution for diffusion analysis. To maximise the signal-to-noise, use a reasonable line-broadening (**LB**) of at least 0.5 Hz, and possibly larger (**figure 4**). To minimise the number of points in the direct dimension, ensure that **SI** in F2 is not too large. It can usually be set to one-half the number of points acquired in the direct dimension (**TD**) (**figure 4**).

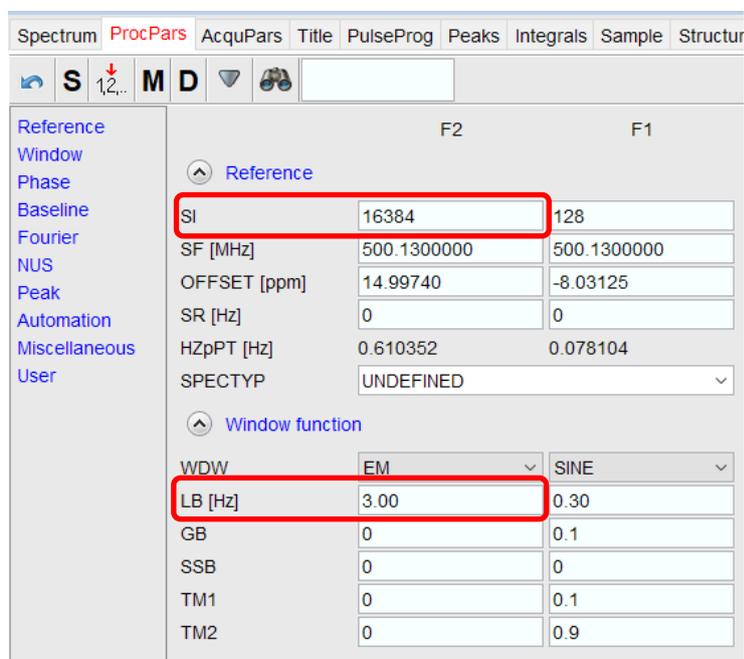


Figure 4. Processing the direct dimension for a DOSY experiment. The goal is to ensure high signal-to-noise and a small number of points. To achieve this, SI can be small, and LB should be at least 0.5 Hz.

4.3 Processing with the T1/T2 Software

- Open the 2D dataset. Phase the spectra by typing “*rser 1*” and correct the phase by running the “*apk*” command. Type “*.ph*” to enter manual phase correction mode. If necessary, correct the phasing manually. While still in manual phasing mode, press the  button to save the phasing for the other 1D spectra in the series. Save and exit manual phasing mode by pressing the  button. Type “*.2d*” (or press the  button) to return to 2D mode.
- In 2D mode, type “*xf2*” to Fourier transform and phase the dataset, then type “*abs2*” to automatically correct the baseline.
- Type “*setdiffparm*”. This AU program will transfer some important parameters to the diffusion analysis module.
- Enter the T1/T2 Analysis module by opening **Analysis | T1/T2 Relaxation** from the top menu.
- Extract a slice (press the  button, or type “*rser 1*”). Select spectrum slice 1 (which is the row with the highest intensity). The data will be opened in a new window.
- Select **Define Ranges** (or press the  button). Define integral regions (the  button) for each peak that you would like to be included in the fitting. When you have defined all the desired integral regions, export them to the relaxation module (the  button).
- Start the fitting procedure by clicking on either the  button to fit peak-by-peak, or the  button to fit all peaks simultaneously.

A typical result after processing with the T1/T2 software is shown in figure 5. You can cycle through the fitted peaks using the  and  buttons. Each integral region is displayed with its own diffusion coefficient and standard deviation.

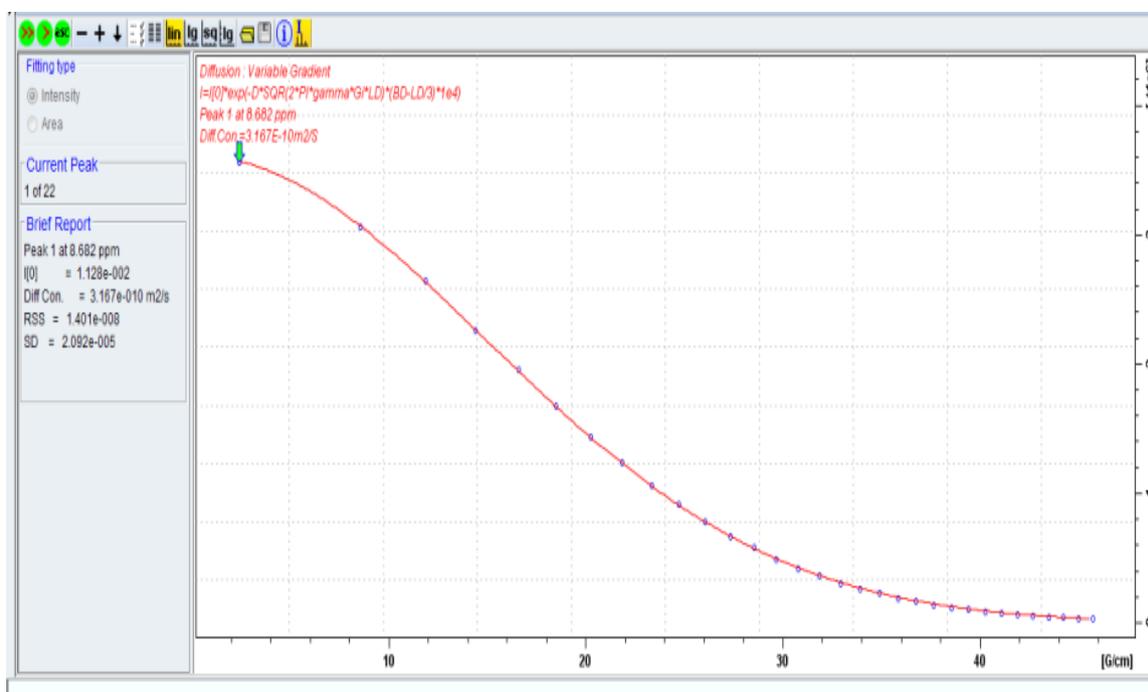


Figure 5. A typical result after processing with the T1/T2 software. The calculated diffusion coefficient is $3.17 \cdot 10^{-10} \text{ m}^2/\text{s}$. Blue points show peak intensities (or areas, according to fitting type selection) as a function of increasing gradient strength and the red line is the regression fit.

4.4 Processing with the DOSY Software

- Open the 2D dataset, and navigate to the processing parameters tab (or type “*edp*”). The number of points in the indirect dimension (**TD**[F1]) corresponds to the number of FIDs / gradient strengths used. Apply zero-filling in the F1 dimension by making sure **SI** in F1 is at least twice the number of points (**TD**) in F1 (the default **TD**[F1] is 16, so **SI**[F1] should be 32, unless you have used a different number of gradients).
- Read the first FID by typing “*rser 1*”, and process by typing “*efp*” and then “*absn*”. Correct the phase by running the “*apk*” command. Type “*.ph*” to enter manual phase correction mode. If necessary, correct the phasing manually. While still in manual phasing mode, press the  button to save the phasing for the other 1D spectra in the series. Save and exit manual phasing mode by pressing the  button. Type “*.2d*” (or press the  button) to return to 2D mode.
- In 2D mode, type “*xf2*” to process and phase the dataset in the F2 dimension.
- Set ABSF1 and ABSF2 to 1000.0 and -1000.0, respectively (simply type “*ABSF1*” and “*ABSF2*” to set these values via a dialogue box). Execute the “*abs2*” command to correct the 2D baseline.
- Type “*setdiffparm*”. This AU program will transfer some important parameters to the diffusion analysis module.
- Type “*eddosy*” to open the DOSY processing module. All parameters should be pre-set with meaningful defaults.
- Type “*dosy2d setup*” to run through the data and estimate the diffusion coefficient range.
- Type “*dosy2d*” to perform the DOSY transform and generate the DOSY spectrum.

The DOSY software produces a 2D DOSY spectrum with chemical shifts on the F2 axis and diffusion coefficients on the F1 axis. An example is shown in figure 6. The centre of each peak corresponds to the calculated diffusion coefficient, and the width of each peak correlates with the fitting error. The F1 axis will show either [log m²/s] (logarithmic scale), or [m²/s] (linear scale). If you have chosen a logarithmic scale, just read the values as log D values in m²/s (i.e. -8.6 means 10^{-8.6} = 2.5*10⁻⁹ [m²/s]). In the case of a linear scale, the diffusion coefficient can be read directly from the F1 axis.

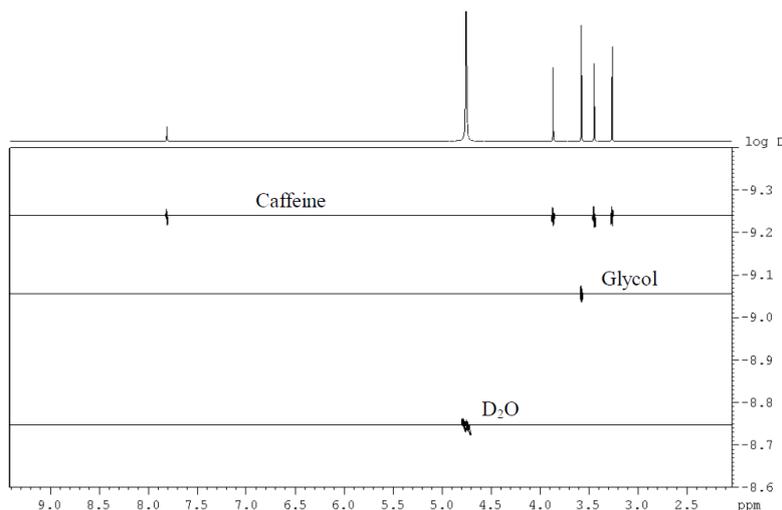


Figure 6. Example of a DOSY spectrum. The sample is a mixture of ethylene glycol and caffeine dissolved in D₂O. Each compound has a different diffusion coefficient, given by the centre of each peak in the F1 dimension

5. Estimating Molecular Weight from Diffusion Data

It is possible to estimate the molecular weight of a compound from its diffusion coefficient, using the SEGWE (Stokes-Einstein Gierer-Wirtz Estimation) method. This calculation is most accurate for small molecules (<1000 g/mol) that are composed of light atoms (sulphur or lighter). For larger molecules (including peptides and proteins), or those composed of heavier atoms, an estimation for molecular weight may still be made, but the interpretation should in general be limited to an assessment of whether molecules are interacting or if oligomeric species are present. This method is only suitable for dilute solutions (or pure solvents).

A spreadsheet for SEGWE calculations (developed at the University of Manchester) is available for download from the NMR website (**figure 7**). In the “D ->M” section of the spreadsheet, select the name of the solvent from the drop-down list, and enter the experimental temperature (in Kelvin) and the calculated diffusion coefficient (in m^2s^{-1}). The estimated molecular weight will be displayed in the “Predicted MW” field. It is also possible to estimate a diffusion coefficient from the molecular weight using the “M->D” section of the spreadsheet. In this case, provide the temperature and solvent as above, and enter the known molecular weight in the “MW” field.

Diffusion Coefficient Calculator RE 2016			
Sample:	Test Sample		
Type of Calculation	M --> D	Top section predicts diffusion coefficient from molecular weight	
Solvent:	methanol-d4	<div style="border: 1px solid black; padding: 2px;"> methanol-d4 DMSO-d6 toluene-d8 CDCl3 tetrahydrofuran-d benzene-d6 cyclohexane-d12 userdefined </div>	Set experimental conditions
Experimental T	298		- solvent from dropdown menu (can add own data)
MW:	46		- sample temperature
g mol ⁻¹			- enter expected molecular weight
Diffusion Coefficient:	2.22E-09		
m ² s ⁻¹			
Type of Calculation	D --> M	Bottom section predicts diffusion coefficient from molecular weight	
Solvent:	methanol-d4		Set experimental conditions
Experimental T	298		- solvent from dropdown menu (can add own data)
Diffusion Coefficient	2.22E-09		- sample temperature
m ² s ⁻¹			
Predicted MW	46.0		- enter experimental diffusion coefficient
g mol ⁻¹			

Figure 7. SEGWE spreadsheet for estimating the relationship between molecular weight and diffusion coefficient for small molecules. In the lower portion of the spreadsheet an estimated molecular weight of 46 g/mol was calculated for a compound with a diffusion coefficient of $2.22 \times 10^{-9} \text{ m}^2\text{s}^{-1}$ in methanol-d4 at 298 K. The reverse calculation is shown in the upper portion of the spreadsheet, where a diffusion coefficient of $2.22 \times 10^{-9} \text{ m}^2\text{s}^{-1}$ has been calculated for this same compound, from the molecular weight, solvent, and temperature.

6. Further Reading

- Claridge TDW, **High-Resolution NMR Techniques in Organic Chemistry**, Elsevier, 3rd Ed, Chapter 10
- Johnson CS, **Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications**, Progress in NMR Spectroscopy, 1999, 34, 203–256
- Evans R, Dal Poggetto G, Nilsson M, Morris GA, **Improving the Interpretation of Small Molecule Diffusion Coefficients**, Analytical Chemistry, 2018, 90 (6), 3987-3994
- Evans R, **The interpretation of small molecule diffusion coefficients: Quantitative use of diffusion-ordered NMR spectroscopy**, Progress in NMR Spectroscopy, 2020, DOI: 10.1016/j.pnmrs.2019.11.002