mQfit, a new program for analyzing quasi-elastic neutron scattering data

Nicolas Martinez[1,2,a], Francesca Natali[2,3] and Judith Peters[1,2]

[1]Univ. Grenoble Alpes, IBS, 38041 Grenoble Cedex 9, France
[2]Institut Laue-Langevin, 71 avenue des Martyrs, CS 20156, 38042 Grenoble Cedex 9, France
[3]CNR-IOM, OGG, 71 avenue des Martyrs, CS 20156, 38042 Grenoble Cedex 9, France

Abstract. Analysis of Quasi-elastic Neutron Scattering (QENS) data of complex systems such as biological or soft matter samples in a comprehensive and explicit way often requires great efforts. Most popular software only allows to fit spectra originating from one single instrument and does not permit to extract parameters from a model that is fitted simultaneously to data taken at different instrumental resolutions. We present here a new program, mQfit (multiple QENS dataset fitting), that enables to fit QENS data taken at different spectrometers (with typical resolutions between 0.01 and 0.1 meV) and momentum transfer ranges. This allows drastically reducing the number of fitting parameters. The routine is implemented with a user friendly Graphical User's Interface (GUI), and freely available. As an example, we will present results obtained on E. *coli* bacterial pellets, and compare them to values published in the literature.

1. Introduction

Quasi-elastic neutron scattering (QENS) allows to probe molecular dynamics in the ps-ns range. The incoherent neutron scattering cross section of hydrogen being much bigger than that of all other atoms present in biological matter, its contribution dominates the spectra. These characteristics make QENS an ideal tool to detect, for instance, dynamics of water in biological systems. They have indeed been extensively studied, from simple objects such as water [1], proteins [2] or lipids [3] to more complex systems such as bacteria [4] or red blood cells [5].

QENS data analysis is usually done by fitting the spectra of an individual instrument by a sum of an elastic contribution and several Lorentzians convoluted with the instrumental resolution. These curves give access to the Elastic Incoherent Structure Function (EISF), the elastic part divided by the sum of the elastic and the quasi-elastic contributions, and to the widths of the Lorentzians. Specific models to describe these quantities permit then to extract physical parameters as diffusion coefficients or residence times [6]. The described procedure is called the 'model free' method, but it has two drawbacks: when the number of Lorentzian functions, necessary to fit the data in a satisfactory way is important, subsequent analysis becomes tricky because of the difficulty to disentangle the different contributions, with consequent enhanced uncertainty on the estimation of the numerous fitting parameters. Secondly in the case where data were taken at different instrumental resolutions, each data set is fitted independently giving sometimes rise to different diffusion coefficients which in contrary should be parameters characterizing unambiguously the sample. In contrary, it has been shown in the case of lipids [7,8] that a fitting approach using models to describe existing motions from the beginning on allows a comprehensive and more accurate description of the system, which has dynamics spanning several orders of magnitude. The method was also applied to more complex systems such as neural tissue [9]. We present here a new user friendly routine capable of fitting simultaneously sets of data taken at different spectrometers with the same model. To illustrate the advantages of the program we present applications to data taken on bacterial cells, and compare the results obtained to values found in the literature.

2. Program presentation

Quasi-elastic scattering is described by the dynamical structure factor S(Q,ω) which can be expressed by:

$$S(Q, \omega) = A_0(Q)\delta(\omega) + \Sigma A_i(Q)L(\Gamma_i, \omega). \quad (1)$$

Q stands for the momentum transfer and ω designates the energy transfer from the neutron to the sample in units of $\hbar$. The first term in Eq. (1) describes particles that appear immobile in the time window of the instrument, while the second term is a sum of Lorentzian functions, each one representing a specific motion. The widths Γ_i and the amplitudes $A_i(Q)$ of the Lorentzians are functions of other parameters depending on the model used. The program calculates the structure factor, convolutes it with the instrumental resolution and performs a least square minimization in order to find the optimal parameters. The program is written in the language Python, which makes it portable on most of the operating systems. It uses for the calculations the numerical and scientific modules Numpy and Scipy [10], as well as a least squares fitting routine furnished by the University of Chicago [11]. The GUI uses the PyQT [12] module.

---

[a] Corresponding author: `martinezn@ill.fr`

2.1. The fitting routine

2.1.1. Available functions

For now, two models are available. The first model can be used to fit data collected on living organisms in H_2O. It consists of a sum of an elastic term, a slow and fast process describing bulk [1] and confined water dynamics [13], and a third component representing dynamics of other cellular components (proteins and lipids). More information on this model can be found in [9]. The second model implemented is destined to analyse data on lipids that consists of an elastic term, two components describing diffusion in a sphere of different radii [14], a third component describing jump-diffusion among two equivalent sites [6], and a large fourth component. Details on this model can be found in [7,8]. More models are foreseen to be implemented into the program within the next future.

2.1.2. Fitting procedure

The model function chosen to describe $S(Q,\omega)$ is convoluted numerically with the instrumental resolution, which can be obtained experimentally by measuring a completely incoherent elastic scatterer as for instance vanadium, and fitted to the data.

Fitting is done by minimizing the χ^2 defined by:

$$\chi^2 = \Sigma(I_{calc} - I_{meas})^2/\sigma^2 \qquad (2)$$

with I_{calc} being the calculated intensity, I_{meas} the corresponding measured intensity and σ the associated error. The minimization is done with a Levenberg-Marquardt algorithm. Other optimizers are also available on demand. Each parameter can be given an upper and lower boundary. At the end of the fitting procedure, values of the parameters and errors are printed and can be saved in an ascii file. The correct error calculation is still an issue, as it is done for the moment by inverting the curvature matrix of the χ^2 function, which cannot always be performed, but this problem will be solved.

2.2. GUI

2.2.1. Main window

The main window consists of a file menu allowing loading data, choosing the instruments, e.g. the resolutions, and the adequate model. It also has an icon bar allowing data visualization and saving.

2.2.2. Loading parameters

The loading parameter window permits to give starting values to the parameters. One can leave some parameters fixed, and also put low and high boundaries to each parameter. The starting values are updated each time a fit is performed, and can be saved and reloaded from a previous fit.

2.2.3. Visualizing data

The data visualization windows can be used to asset fit quality and check consistency. It consists of two separate windows, one showing the elastic and quasi-elastic contributions and their sum compared to the experimental data at different Q-values, the other one showing residuals as a function of Q and energy transfer.

2.2.4. Saving parameters

All output files are saved as ascii text files. When saving, the program writes a file containing the parameter values that can be later imported for a new fit, a file containing the parameter values and errors, as well as information on the instrument(s) where the fitted data were recorded. Another set of files (one per Q value) containing values of intensity for each contribution convoluted with the instrumental resolution function versus the energy transfer is generated, too.

3. Example application: Water dynamics in E. *coli*

To illustrate the program capabilities, we present here a data set taken on the instruments IN5 [15] and IN6 [16] of the Institut Laue-Langevin (ILL) in Grenoble/France. The sample was a bacterial pellet of *Eschericia coli* cells and the instrumental energy resolution was 0.01 meV for IN5 and 0.1 meV for IN6. Data reduction was carried out with the LAMP software available at the ILL [17]. Both datasets were then taken into account simultaneously for the fitting procedure. Q-values spanning from $0.4\,\text{Å}^{-1}$ to $1.1\,\text{Å}^{-1}$ and $1.3\,\text{Å}^{-1}$ to $2.0\,\text{Å}^{-1}$ were used for IN5 and IN6 respectively. Energy transfer ranging from -1.5 meV to 0.1 meV and -3.0 meV to 0.5 meV was used for IN5 and IN6 respectively.

3.1. Model used

The model used is described in [9]. It consists of a sum of an elastic term, a term describing bulk water dynamics that is the convolution of a jump-diffusion process and rotations, as these motions can happen simultaneously. The jump-diffusion process is characterized by the diffusion coefficient D_{w1} and residence time τ_1. The rotation is characterized by a radius, which is fixed to $1.98\,\text{Å}$ (corresponding to the O-H length in a water molecule) and a rotational diffusion coefficient D_{rw1}. The third contribution is attributed to interfacial water dynamics that follows the same type of dynamics as bulk water, but the values of diffusion coefficient D_{w2}, residence time τ_2 and rotation diffusion coefficient D_{rw2} are reduced compared to bulk water values. The fourth component is a Lorentzian with a full width at half maximum (FWHM) Γ constant in Q, which is attributed to CH_2 jump rotational motions [3,7] within lipids and proteins. f, p_1, p_2 and p_3 represent the fractions of hydrogen atoms belonging to bulk water, interfacial water and CH_2 groups, respectively.

3.2. Results

Figure 1 shows the fitted data and the different contributions for IN5 at a 0.01 meV instrumental energy resolution. We can see that all motions are well resolved
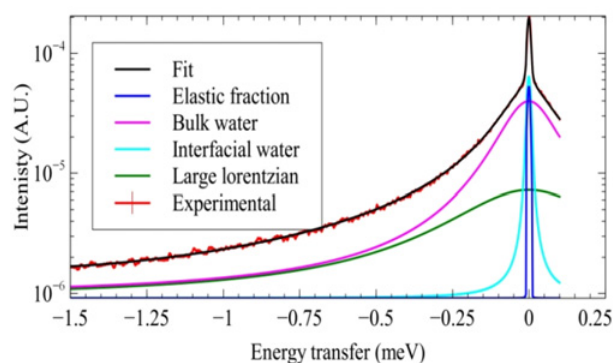
Figure 1. Experimental and calculated contributions for $Q = 0.95\,\text{Å}^{-1}$ on IN5.
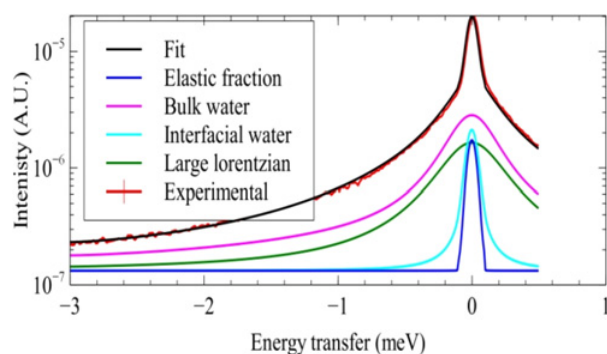

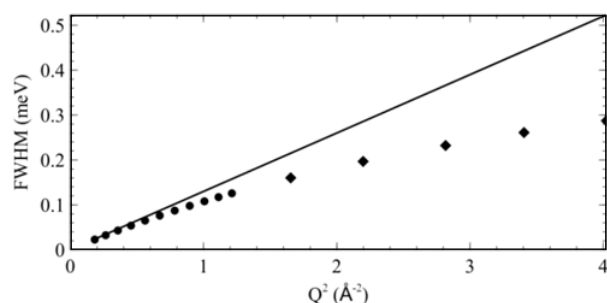**Figure 2.** Experimental and calculated contributions for $Q = 1.48\,\text{Å}^{-1}$ on IN6.

Figure 3. FWHM of the first Lorentzian as a function of Q^2. The filled circles and filled diamonds represent IN5 and IN6 Q values respectively. The solid line represents a purely translational diffusive process.

on this instrument as the quasi-elastic contributions are broader than the elastic one.

Figure 2 shows the same plot, but on IN6 at a ten folds energy instrumental resolution, namely 0.1 meV. We can see that here the interfacial water dynamics is poorly resolved, as it is close to the instrumental resolution, defined by the elastic peak, what illustrates the difficulty one can encounter when fitting the data taken on one instrument, only.

On the other hand, IN6 has a greater Q-range, and thus allows a better characterization of the bulk water residence time through the FWHM of the corresponding Lorentzian, as shown in Fig. 3.

Table 1. Parameter values.

	E. coli	Bulk water
f	0.03	
p_1	0.70	
p_2	0.08	
p_3	0.19	
D_{w1} (10^{-5} cm^2/s)	2.0	2.5
τ_1 (ps)	1.0	1.1
D_{rw1} (meV)	0.28	
D_{w2} (10^{-5} cm^2/s)	0.08	
τ_2 (ps)	NA	
D_{rw2} (meV)	0.02	
Γ (meV)	0.25	

We can see that the asymptotic behaviour of the FWHM (defining the value of the residence time) becomes clearly apparent at Q^2 values above $1.5\,\text{Å}^{-2}$, which are only available in the IN6 measurement. The obtained parameter values are shown in Table 1, along with bulk water values.

The interfacial water residence time τ_2 is not defined, since the Q range at the highest resolution is still too limited. The population fractions tell us that the organism is composed at around 80% of water, which is expected from these types of samples [18]. The values of diffusion coefficient, residence time and rotational coefficient found for bulk cellular water are similar to the one of free water [1], but slightly reduced. This result is consistent with [19] where water dynamics on deuterated E. *coli* cells was studied. The interfacial water dynamics is slowed down by roughly one order of magnitude with respect to bulk water, strongly resembling interfacial water present at the surface of proteins [13]. Finally the last component is an indicator of an average macromolecular flexibility, and could be directly compared to other cellular systems. The values of the parameters are consistent with the literature and thus establish a proof of concept of this type of analysis on bacterial systems.

4. Conclusion

We present here a new program that can be used to analyze comprehensively QENS data taken at different instrumental energy resolutions. Its GUI makes it user friendly, while being highly flexible and transparent. The results obtained on E. *coli* bacterial pellets show that using data sets originating from different spectrometers to analyze the same sample presents great advantages, especially in such samples with complex dynamics spanning timescales of several orders of magnitude. More models will be added soon to the program. The source code and a short user's manual can be found on the webpage `http://forge.epn-campus.fr/`.

References

[1] J. Teixeira, M.-C. Bellissent-Funel, S. H. Chen, and A. J. Dianoux, Phys. Rev. A **31**, 1931 (1985)

[2] W. Doster, S. Cusack, W. Petry, Nature **337** (1989)

[3] S. König W. Pfeiffer, T. Bayerl, D. Richter E. Sackrnann J. Phys. II France, **2** (1992)

[4] M. Tehei, B. Franzetti, K. Wood, F. Gabel, E. Fabiani, M. Jasnin, M. Zamponi, D. Oesterhelt, G. Zaccai, M. Ginzburg, B.-Z. Ginzburg, PNAS **104**, 3 (2007)

[5] A. M. Stadler, J. P. Embs, I. Digel, G. M. Artmann, T. Unruh, G. Büldt, G. Zaccai, JACS **130**, 50 (2008)

[6] M. Bee, Quasi-elastic neutron scattering, Principles and Applications in Solid State Chemistry, Biology and Materials Science, Adam Hilger, 1988

[7] Y. Gerelli, V. García Sakai, J. Ollivier, A. Deriu, Soft Matter **7** (2011)

[8] W. Knoll, J. Peters, P. Kursula, Y. Gerelli, J. Ollivier, B. Demé, M. Telling, E. Kemner, F. Natali, Soft Matter **10** (2014)

[9] F. Natali, Y. Gerelli, C. Stelletta, J. Peters, AIP Conf. Proc. **1518**, 551 (2013)

[10] T. E. Oliphant, Computing in Science & Engineering **9**, 90 (2007)

[11] http://lmfit.github.io/lmfit-py/index.html

[12] http://www.riverbankcomputing.co.uk/software/pyqt/intro

[13] S. Dellerue, M. C. Bellissent-Funel, Chemical Physics **258**, 2-3 (2000)

[14] F. Volino, A. J. Dianoux, Molecular Physics **41**, 2 (1980)

[15] http://www.ill.eu/en/html/instruments-support/instruments-groups/instruments/in5/description/instrument-layout/

[16] http://www.ill.eu/en/html/instruments-support/instruments-groups/instruments/in6/description/instrument-layout/

[17] http://www.ill.eu/data_treat/lamp/the-lamp-book/

[18] G. Bratbak, I Dundas, Appl. Environ. Microbiol. **48**, 4 (1984)

[19] M. Jasnin, M. Moulin, M. Haertlein, G. Zaccai, M. Tehei, EMBO Reports **9**, 6 (2008)