

***BayesApp*: a web site for indirect transformation of small-angle scattering data**

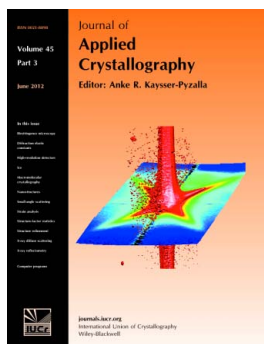
Steen Hansen

J. Appl. Cryst. (2012). **45**, 566–567

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Reproduction of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see <http://journals.iucr.org/services/authorrights.html>



Journal of Applied Crystallography covers a wide range of crystallographic topics from the viewpoints of both techniques and theory. The journal presents papers on the application of crystallographic techniques and on the related apparatus and computer software. For many years, the *Journal of Applied Crystallography* has been the main vehicle for the publication of small-angle scattering papers and powder diffraction techniques. The journal is the primary place where crystallographic computer program information is published.

Crystallography Journals Online is available from journals.iucr.org

BayesApp: a web site for indirect transformation of small-angle scattering data

Steen Hansen

Department of Natural Sciences and Environment, Faculty of Life Sciences, University of Copenhagen, Thorvaldsensvej 40, DK-1871 FRI C, Denmark. Correspondence e-mail: slh@life.ku.dk

A web site (<http://www.bayesapp.org>) for indirect transformation of small-angle scattering data is presented. When experimental data are uploaded to the server, they are processed in a few seconds and the result of the indirect information is displayed on the screen in the form of a distribution, together with the experimental data and the fit to the data. No other user input than the experimental data is necessary, but various options for the analysis may be selected. The results of the analysis can be downloaded from the web site in the form of ASCII files.

© 2012 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

Small-angle scattering (SAS) is a technique that is frequently used by a wide range of scientists for studying the structure of molecules in solution (e.g. Glatter & Kratky, 1982). The analysis of the data consists mostly of either direct fitting of a real-space model to the experimental data in reciprocal space, or indirect transformation of the data with the aim of obtaining some representative real-space distribution of interest (Glatter, 1977). Many excellent program packages are available for these tasks (e.g. *ATSAS* by Konarev *et al.*, 2006), which require download and installation on individual PCs before they can be applied to experimental data. In addition to providing a program package for various analytical tasks, the SAS toolbox (*SASTBX*) by Zwart & Liu (2010) contains online services for several useful applications, but no indirect transformation without user input.

The web site described here, <http://www.bayesapp.org>, makes indirect transformation possible without any downloading of programs or user input except for the experimental data. Parameters such as the maximum diameter or the overall noise level of the data are selected automatically by Bayesian analysis, as described by e.g. Hansen (2000, 2008). Consequently, the web site should be easy to use, but for a more detailed analysis, using more points than specified below and therefore requiring more CPU time, it may be necessary to use a program package. Alternatively, the program used by the web site may be downloaded and modified by the user.

As SAS experiments are often carried out at large-scale facilities by users of various backgrounds, the web site described here should facilitate data analysis not only by providing easy access to the software and by its applicability, but also by the short CPU time required for analysis of a data set. This short CPU time will allow for real-time analysis of many experiments, which might be useful e.g. in combination with the quantification of the information content of the data, calculated by the web site and given as a number (Vestergaard & Hansen, 2006). Using the time evolution of this number, the use of beam time may be optimized, the onset of aggregation in the sample detected *etc.*

2. Description of the web site

2.1. Input

The only user input required by the web site is the experimental data. Text lines in the data file are ignored by the program. The data

must have the format $[q_i, I(q_i), \sigma(q_i)]$, where q_i is the length of the scattering vector q at data point number i [$q = (4\pi/\lambda) \sin \theta$, where θ is half the scattering angle and λ is the wavelength of the incident radiation], $I(q_i)$ is the measured intensity at this point and $\sigma(q_i)$ is the corresponding estimated standard deviation. If the number of data points is larger than 80, the data points are rebinned so that a result can be obtained in just a few seconds of CPU time. The restriction of the number of data points to 80 was chosen to allow the user to change parameters and obtain new results in a relatively short time. Even though no user input is necessary, the user may wish to change some parameters, e.g. to include automatic background subtraction or to fix the noise estimate to a higher level owing to systematic errors in the data. Furthermore, the brief CPU time required makes it possible to carry out an analysis of an experiment while the experiment is running. Owing to the resolution of SAS experiments, data analysis using only a relatively small number of data points may often be carried out with only a negligible loss of information (Shannon & Weaver, 1949). The rebinning of k points into 1, denoted $[q_m, I(q_m), \sigma(q_m)]$, is carried out in a standard manner according to $q_m = \sum_{i=1}^k q_i/k$, $I(q_m) = \sum_{i=1}^k I(q_i)/k$ and $\sigma(q_m) = [\sum_{i=1}^k \sigma(q_i)^2]^{1/2}/k$.

Similarly, for an estimation of the distribution of interest a maximum of 100 points may be used in the web version. Consequently, the web version of the program will only be able to handle experimental intensities covering a dynamic range of about 10^3 .

For a more detailed analysis, the program allows 1000 data points to be used for the intensity and 1000 for the distribution of interest. This is more than sufficient for the resolution obtainable by SAS experiments. In this case, the user is required to download the source code and run the program on his or her own computer.

2.2. Output

After the data have been uploaded by pressing the 'Submit' button, the web site will give an estimate of the distribution of interest in about 5 s (depending upon the exact number of points used for the analysis). An error estimate is made using a Monte Carlo procedure if this option is selected. In this case, a number (to be specified by the user) of extra solutions are calculated by varying the maximum diameter, the estimated noise level and other (hyper-) parameters randomly, close to their optima. From the probabilities associated with each of the extra solutions, an error estimate on the average solution may be calculated (Hansen, 2000). The data and the

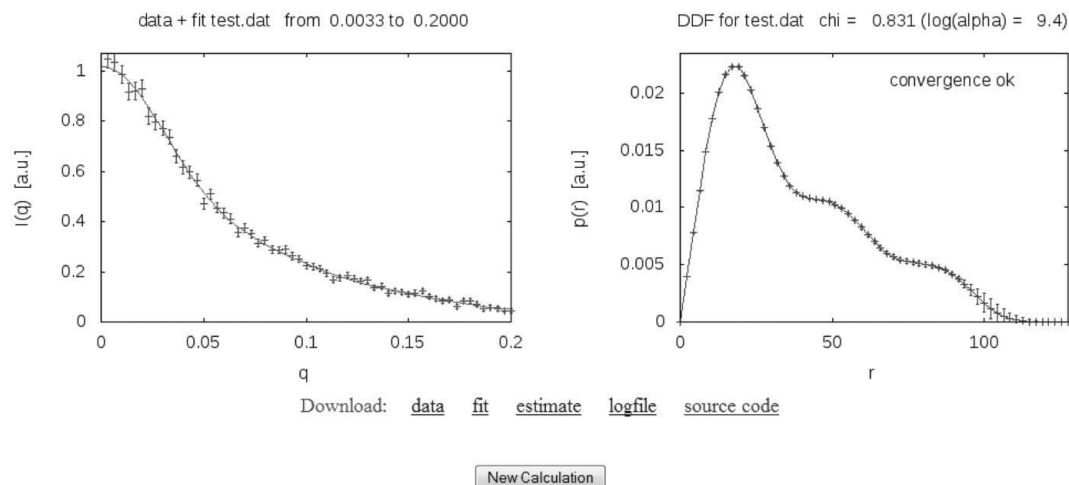


Figure 1
Example of screen output.

result are displayed as graphs using *gnuplot* (Williams & Kelley, 1986). An output example is shown in Fig. 1 with the experimental data and the fit to the data. More detailed information about the calculation and the result (estimates of the forward scattering, the Guinier radius *etc.*) can be found in the log file, which is accessible from the 'Results' web page.

The default is a simple estimation of the distance distribution function, but various other options exist with regard to limitations in the q range used, background fitting, fixation of various parameters *etc.* For non-dilute solutions, an option in the program allows simultaneous estimation of the distance distribution function and the structure factor if the user enters a nonzero estimate of the volume fraction. In this case the low information content of SAS data may lead to relatively high sensitivity towards user input, but estimates of the relevant distributions are made which are consistent with the experimental data.

The user's experimental data and the corresponding results are deleted from the server when a new data set is uploaded and analyzed. Uploading a dummy data file, such as the test data provided at the web site, will delete any previous data and results if this is required.

2.3. Availability

The program is written in Fortran90 and contains subroutines from Press *et al.* (1992). The source code can be downloaded free of charge from the web site.

References

- Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.
- Glatter, O. & Kratky, O. (1982). Editors. *Small Angle X-ray Scattering*. London: Academic Press.
- Hansen, S. (2000). *J. Appl. Cryst.* **33**, 1415–1421.
- Hansen, S. (2008). *J. Appl. Cryst.* **41**, 436–445.
- Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. (2006). *J. Appl. Cryst.* **39**, 277–286.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). In *Numerical Recipes in Fortran77. The Art of Scientific Computing*, 2nd ed. New York: Cambridge University Press.
- Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Vestergaard, B. & Hansen, S. (2006). *J. Appl. Cryst.* **39**, 797–804.
- Williams, T. & Kelley, C. (1986). *gnuplot*, <http://www.gnuplot.info/>.
- Zwart, P. & Liu, H. (2010). *The Small Angle Scattering ToolBox*, <http://sastbx.als.lbl.gov/>.