GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES VOL. 11, NO. 1, JANUARY 2005: 141 - 148
 ISSN 1118-0579

DATA ANALYSIS OF SMALL ANGLE X-RAY SOLUTION SCATTERING AND ITS APPLICATION TO HUMAN TUMOUR NECROSIS FACTOR.

S. P. MALLAM and T. C. AKPA

(Received 18 July 2000; Revision Accepted 23 August 2001)

ABSTRACT

Small Angle X-ray Scattering analysis was used for the study of the protein, Human Tomour Necrosis Factor (TNF) homogeneously dispersed in solution. The experiment consisted in sending a well collimated beam of synchrotron radiation of wavelength, λ through the sample and measuring the variation of the intensity as a function of the scattering angle θ through the scattering vector $s = 2\sin\theta/\lambda$. The scattering curve recorded at very low angles was first extrapolated to the origin, joined to data recorded at so called wide-angles and extrapolated to infinity to obtain a complete scattering curve for the analysis using computation subroutines that can run iteratively. The complete scattering curve was then used to extract the size, shape and the organisation of the macromolecules in solution. The results obtained reveals that in solution, the TNF is a trimer with a radius of gyration of 25.5 ± 0.3Å and a maximum diameter of at least 84 Å.

Keywords: Small Angle X-Ray Scattering; Tumour Necrosis Factor; Radius of Gyration; Maximum Diameter.

INTRODUCTION

Tumour necrosis factor (TNF), is a protein with a large number of biological effects. It has been shown to cause in-vivo necrosis of tumour cells and in-vitro lysis of certain transformed cell lines. TNF activity has been reported to include anti-viral effects, involvement in cell differentiation and modulation of cellular immune response in-vitro (Mestan *et al.*, 1986, Takeda *et al.*, 1986). The polypeptide chain of human TNF has a molecular weight of 17356 daltons (Marnenout *et al*, 1985). In solution, the active form of the protein is suspected to be a compact trimer (Wingfield *et al.* 1987).

Given this wide spectrum of effects and the growing interest in its clinical potential, it is important to determine the three dimensional structure of the protein. Small Angle X-ray Scattering (SAXS) analysis which is a technique for studying structural features of macromolecules of sizes from 0.5 to 30.0 nm dispersed in a homogeneous medium has proved useful for such studies (Cantor and Schimmel, 1980). In this paper, we present the theoretical basis of the technique, the computational aspects used in deriving the structural parameters and the results obtained by applying these to human TNF.

THEORY OF SMALL ANGLE X-RAY SOLUTION SCATTERING

Small Angle X-Ray Scattering (SAXS) is a particular case of coherent elastic scattering. Each electron is a source of secondary scattering, contributing to the final scattering pattern. In ideal experimental scattering conditions, it may be assumed that the sample under study is homogeneous, isotropic and with small dimensions relative to the sample detector distance. The

S. P. MALLAM, Health Physics and Radiation Biophysics Section, Centre for Energy Research and Training, ABU, Zaria - Nigeria
T. C. AKPA, Health Physics and Radiation Biophysics Section, Centre for Energy Research and Training, ABU, Zaria - Nigeria

sample is considered perpendicular to the direction of propagation of the well-collimated monochromatic radiation beam, which it intercepts completely. The radiation is assumed to be scattered to a cone of half angle at the top of 2θ such that $sin2\theta = 2\theta$ and $cos2\theta = 1$. If in addition, absorption is considered negligible and the scattering curve is corrected for parasitic scattering from the solvent and collimator, the scattered intensity, I(s) may be written as (Luzzati, 1960; Luzzati, *et. al.* 1976);

$$I(s) = <\int p(\underline{r})\exp(-2i\pi \underline{s}.\underline{r})dV_{\underline{r}} > = \int\{p(r)[\sin(2\pi rs)/2\pi rs]\}\, dV_r \qquad (1)$$

where $p(\underline{r})$ is the auto-correlation function of $\rho(\underline{r})$, the electronic density distribution of the sample defined by:

$$p(\underline{r}) = \rho(\underline{r})^*\rho(-\underline{r}) = \int\rho(\underline{r})\rho(\underline{r}+\underline{R})\, dV_{\underline{r}} \qquad (2)$$

and $\underline{s}$ is the scattering vector with magnitude $s = 2sin\theta/\lambda$, and λ the wavelength of the X-rays.
The distance distribution function is given by (Glatter and Kratky, 1982):

$$r^2p(r) = 2r\int s.i_n(s)\sin(2\pi rs)ds \qquad (3)$$

where $i_n(s) = I(s)/V\rho$ is the relative scattering intensity normalised to one electron, V the volume of the sample and ρ, the average electronic density of the sample. Therefore, the scattering intensity is determined by the spatial arrangement of the electrons in the sample and depends only on the scattering angle 2θ through the scattering vector $\underline{s}$ since $\underline{r}$ has been averaged out.
The study of macromolecules in solution may be developed through the hypothesis of an ideal solution of identical particles so that no correlation exists between the particles. This can be approximated by a solution of infinite dilution. From these assumptions, the normalised scattering intensity $i_n(s)$ by a sample containing N particles may be expressed as a function of the scattering intensity by one particle $i_1(s)$ as (Vachette, 1979):

$$i_n(s) = N.\, i_1(s) \qquad (4)$$

The intensity $i_1(s)$ is defined by the electronic density contrast between the particle and the solvent.
For a solution of globular particles, it has been shown that the scattering intensity $i_n(s)$, near the origin can be written as a series in s^2 such that (Luzzati, *et. al.* 1976):

$$i_n(s) / i_n(0) = 1 - (4/3)\pi^2R_g^2\, s^2 + \ldots\ldots \qquad (5)$$

where R_g^2 is the square of the radius of gyration R_g. Very close to the origin, $(s\rightarrow 0)$, the first two terms give the so called Guinier approximation:

$$i_n(s) = i_n(s)\exp(-4\pi^2R_g^2\, s^2/3) \qquad (6)$$

By plotting the logarithm of the scattering intensity as a function of the square of the scattering vector s^2, the radius of gyration of the particles can be determined from the slope. The Guinier law can be exploited in practice to extrapolate the scattering curve to the origin, $(s = 0)$, a domain not accessible experimentally.

In addition to the foregoing discussion, a supplementary assumption can be made that the electron density is uniform in the volume V_1 associated to the particle. This volume will contain the particle and

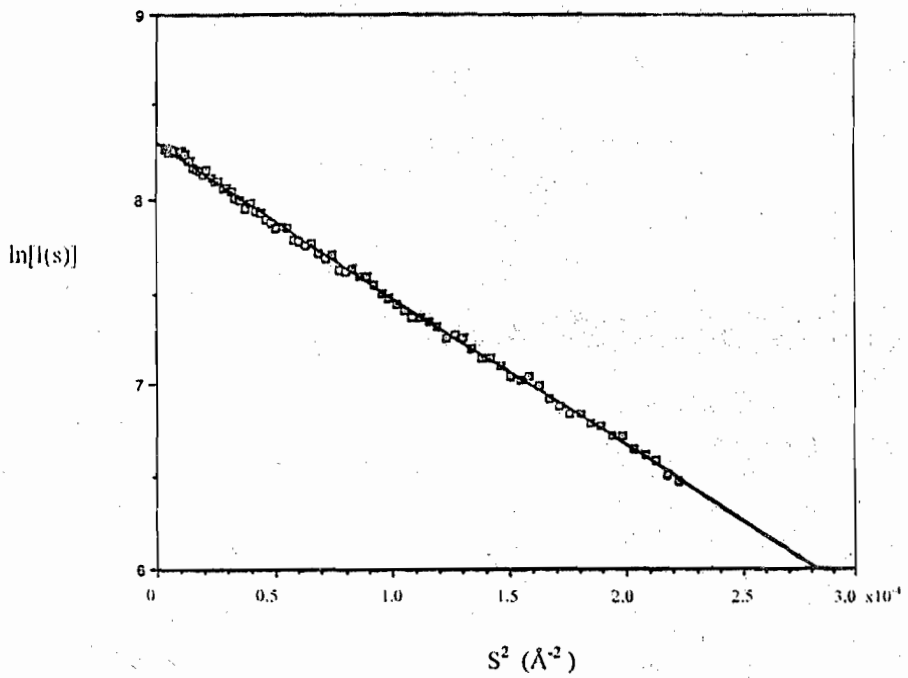

Figure 1: Guinier Plot of Tumour Necrosis Factor Solution Scattering Data at very Small Angles

some water but does not represent the sample volume V. At the resolution of the experimental set-up, this hypothesis can be applied to a globular protein molecule. In this case, the asymptotic form of the scattering intensity may be written as (Murthy and Knox, 1977; Glatter and Kratky, 1982):

$$\lim_{s\to\infty} 8\pi^3 [s^4 i_h(s)/c_e] = (A + Bs^4)/m \qquad (7)$$

where A and B are constants, c_e is the electron concentration of the sample and m is the number of electrons in the particle. In this expression called Porod's law, the constant A is related to the surface S, between the particle with ρ_1 electrons and the solvent of ρ_0 electrons by (Vachette, 1979):

$$A = S(\rho_1 - \rho_0)^2 \qquad (8)$$

The constant B is a factor which takes into account the variations in the electron density at small distances, an eventual polydispersity of the particles and the errors in $i_h(s)$ at large s. For cases where there exists a linear variation of the intensity, B can be calculated and subtracted out from $i_h(s)/c_e$ for subsequent calculations. The determination of the asymptotic form of the scattering curve using Porod's law can be used to test the hypothesis of uniform electron density and also to extrapolate the curve to infinity, a region also not accessible experimentally. This extrapolation allows for the complete determination of the scattering curve from the origin to infinity used for the determination of the distance distribution function $r^2p(r)$ by Fourier transformation.

Therefore, from the X-ray scattering curve of the particles in solution, the radius of gyrations R_g, the particle surface area S, the particle hydration volume v_1 may be determined. In addition, the distance distribution function can give an estimate of the maximum diameter D_{max} of the particle. It should, however be noted that the determination of S depend essentially on the intensity at large angles and is therefore poorly determined. On the other hand, the distance distribution function can only give a lower limit for the maximum diameter since the upper limit is affected by residual oscillations in the Fourier Transform.

DATA ANALYSIS SUBROUTINES

Subroutines for the determination of the asymptotic form of the scattering curve, splicing of the two curves recorded at different conditions of concentration and sample-detector distance to give a single continuous curve on the same scale, the Fourier transformation of the complete scattering curve and the calculation of the distance distribution function were written bearing in mind the need for them to be interactive, user friendly and capable of being presented in graphical form. The different menus and questions were designed to make them very explicit. Default values were provided for all questions to allow for the continuation of the calculation in cases where better values are not known (Golding, 1982; Boulin, *et al.*,1986; Mallam, 1987).

To run the subroutines ASYMPT which traces the scattering curve $s^4i(s)$ as a function of s^4 after background subtraction, the initial limits for the Porod law extrapolation are entered from either the key board or the cursor. ASYMPT, by a linear regression in this region, then calculates the slope B and the extrapolation constant A, the regression coefficient and draws the curve and the regressed line of best fit. When a satisfactory result is obtained, the subroutine subtracts B from the experimental intensity, channel by channel, and stores the result in a file. This subroutine requires as input principal parameters, the solvent multiplication factor for background subtraction, the last channel to be used, the limits of the regression region, channel width in $Å^{-1}$ and the last channel of the experimental intensity to be kept.

As seen earlier, scattering intensity at wide angles (up to about 0.05 $Å^{-1}$) are one or two orders of magnitude weaker than at low angles. Thus it is important that the wide-angle data be recorded at high concentrations and sometimes, at a reduced sample-detector distance. The extrapolation to the origin (Guinier) and to infinity (Porod) are therefore made on two different experimental curves both in terms of intensity as well as scattering vector s. The subroutine SPLICE is used to reduce these scattering curve to the same scale. To do this, the subroutine REPACK is used to interpolate the low angle data to the wide-angle data with the same channel width. The two curves are then spliced together in the overlap region by taking either a simple average or weighted average of the experimental intensities for each channel and stores the resulting complete scattering curve in a file. The principal parameters required for running these subroutines are the channel widths of the two scattering curves to be spliced, the concentration ratios, and the overlap region for the splicing and the last channel to be stored.

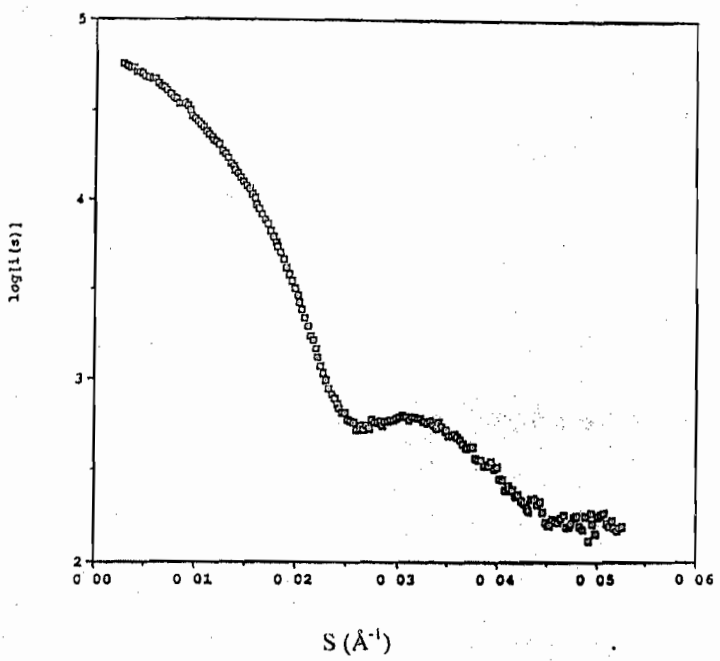

Figure 2: Wide Angle Scattering Pattern for Tumour Necrosis Factor in Solution

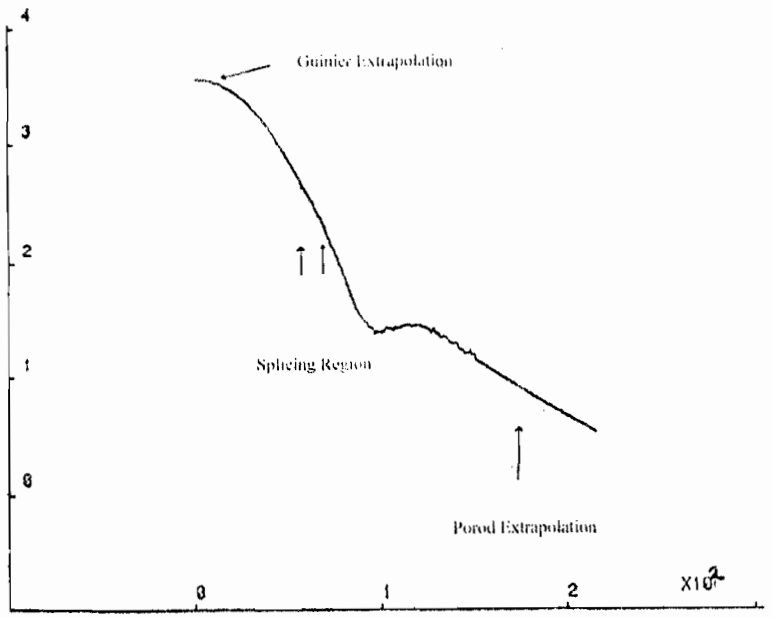

Figure 3: Complete Scattering Curve of TNF Solution after Splicing the Wide Angle Data to the Data from much Small Angle (S→0)

Starting from the complete scattering curve from the splicing subroutine, the subroutine MAXIDIS calculates the distance distribution function $r^2p(r)$ by carrying out its Fourier Transform and storing the results in a file. The distance distribution function obtained will depend on the quality of the extrapolation, the precision of the splicing between the two scattering curves and, of course, the precision of the experimental intensities. These calculations are carried out on the raw data without any smoothing. Beyond the values of an upper limit D_{sup} of r corresponding to the molecule size on the distance distribution function, there exist therefore parasitic oscillations which will affect also the former and therefore introduce an error in the determination of the maximum diameter D_{max} of the molecule.

The complete scattering curve $i_h(s)$ is a function defined on [0, ∞] and width Δs. Its Fourier Transform is

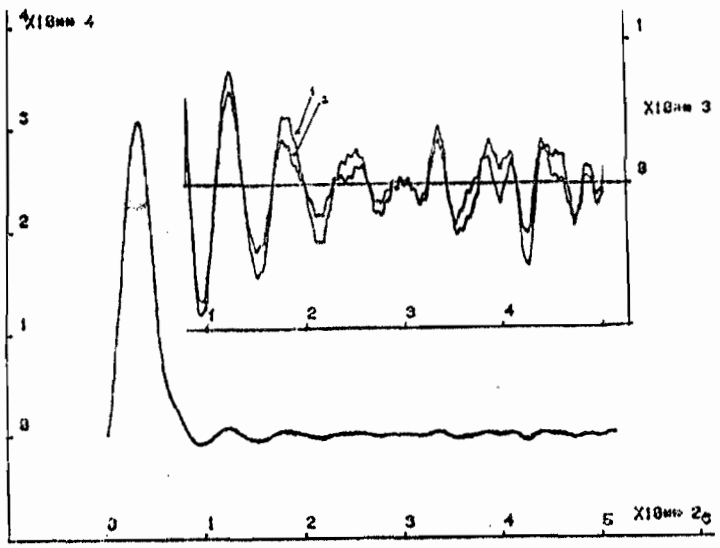

Figure 4: Distance Distribution Pattern of TNF in Solution both before and after Inverse Fourier Transformation. The Two Curves are Enlarged for Distances between 80 and 512 Å so as to render the Differences Visible. In insert, 1 is the Calculated Curve from Data in Figure 3 while 2 is the Calculated Curve after Inverse Fourier Transform.

therefore a periodic function with frequency $1/\Delta s$ and symmetric about $1/(2*\Delta s)$. Beyond a reasonable upper limit for the estimation of the maximum diameter, the only contribution to the values of $r^2p(r)$ are due to the above mentioned errors. To correct for these parasitic oscillations, the inverse Fourier transform of the distance distribution function between D_{sup} and $1/(2*\Delta s)$ can be carried out using the subroutine INVERSEFT. The resulting intensity of the channels with high intensity can then be subtracted from the corresponding intensity $i_n(s)$ of these channels in the complete scattering curve and the corrected curve used as input to calculate the distance distribution and thus the maximum diameter, D_{max}. Two or three of this iterative process is sufficient to eliminate the residual oscillations due to either imperfect splicing (which is often the case) or some abnormal intensity such as parasitic radiation passing through the discriminator barrier during the experiment.

The key parameters require for running both MAXIDIS and INVERSEFT include, the maximum channel of the recorded intensity in the complete curve, the value of the intensity at the origin $i_n(0)$, the value of the extrapolation constant A in the Porod region, the channel width of the complete scattering curve Δs, real space sampling length Δr, an upper limit of r for the calculation of p(r), the first channel for use in the calculation of the inverse transform and the region of the scattering curve to be calculated by INVERSEFT.

APPLICATION TO HUMAN TUMOUR NECROSIS FACTOR

Small Angle X-ray Scattering data were recorded for the recombinant tumour necrosis factor, produced in *E. coli* using phage lambda promotor, at the D24 experimental line at Loboratoire d'Utilisation du Rayonnement Ionisantes (LURE) in Orsay, France. For the low angle data, varying protein concentrations of between 2 and 17 mg/ml were used at a sample to detector distance of 795 mm. This corresponds to $\Delta s_1 = 1.66 \times 10^{-4}$ Å^{-1}/channel. The wide angle data was recorded at a concentration of 83 mg/ml and a sample - detector distance of 496 mm corresponding to $\Delta s_2 = 2.67 \times 10^{-4}$ Å^{-1}/channel.

The low angle data was used to derive the radius of gyration R_g of the molecule using the Guinier approximation as shown in Figure 1. Plotting the radius of gyration so obtained as a function of solution concentration and extrapolating to zero (ideal solution assumption) gives the value of R_g as 25.5±0.3Å. From the intensity at the origin, s = 0 Å^{-1}, a molecular weight of between 50 and 53 kDa were obtained for the protein. This shows the protein to be present in solution as a trimer in agreement with sedimentation equilibrium determinations (Wingfield *et.al.*, 1987).

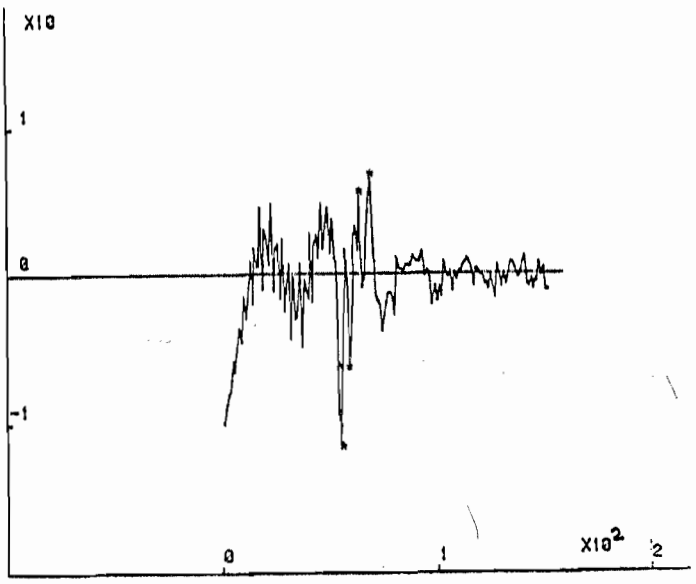

Figure 5: Inverse Fourier Transform of the Distance Distribution Function for r between 150Å and $1/(2\Delta S_2)$

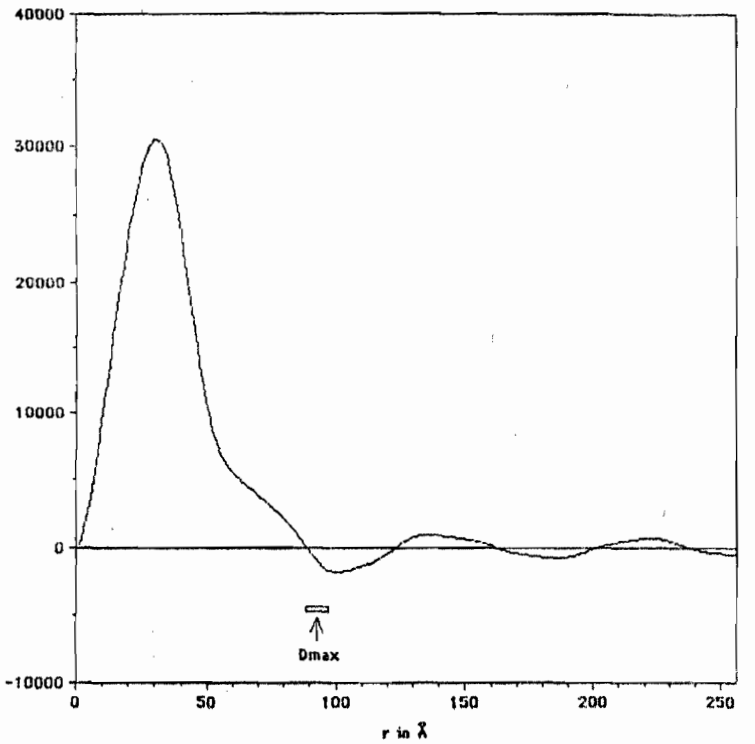

Figure 6: Plot of the Distance Distribution Function after Correction of Certain Channels in the Complete Scattering Curve of Figure 3

The wide-angle scattering data is presented in Figure 2. The secondary maximum near $s = 0.03$ Å^{-1} in Figure 1 may be associated with the oligomeric structure of the molecule. The subroutines ASYPT and SPLICE (with an option to REPACK) were used to extrapolate the scattering curve in Figure 2 to infinity before splicing it to the data collected at much smaller angles ($s \rightarrow 0$) on the same scale. The complete curve is shown in Figure 3. Using MAXIDIS, the distance distribution function was calculated from the complete scattering curve and is shown in Figure 4. Residual oscillations beyond the particle size are clearly visible on this Figure with a pseudo-period of about 60Å. The amplitude of the first minimum represents less than 3% of the maximum of the function. An inverse Fourier Transform of the distance distribution function between 150Å and $1/(2*\Delta s_2)$ is shown in Figure 5. As would be expected, this figure shows some channels with particularly high intensities at s values corresponding to multiples of (1/60) Å^{-1}. This is the region where the two curves were spliced and matched into a single curve. The values of the intensities of the inverse transform at these channels (marked * in Fig. 5) were subtracted from the corresponding channels in the complete scattering curve and the Fourier Transform repeated. The resulting curve, marked 2 on the enlarged insert in Figure 4, shows a reduction of less than 15% of the first minimum of the residual oscillation. The reduction is however, more significant beyond the first minimum. The distance distribution function intersects the abscissa at about 84Å. Taking into account the error in the estimation of the radius of gyration, this value is taken as the lower limit of maximum diameter, which may probably be comprised between 84 and 100Å. This lower limit will support an elongated model, such as the trimer of prolate ellipsoids, for the shape of the protein in solution.

CONCLUSION

From an X-ray scattering curve of the particles in solution, the radius of gyration R_g, the molecular mass and the hydration volume V_1 may be determined among other particle parameters. In addition, the distance distribution function can give an estimate of the maximum diameter D_{max} of the particle. Computation subroutines developed in this paper allow the calculations to be carried out interactively. The application these techniques to the analysis of small angle scattering data of human tumour necrosis factor reveals that the protein is probably a trimer of molecular mass of between 50 and 53

kDa and a R_g of about 25.5 ± 0.3Å. The maximum diameter D_{max} of human TNF, taken into account the residual oscillations in the Fourier Transform and the error inherent in the determination of the radius of gyration R_g, is estimated to be at least 84Å.

REFERENCES

Boulin, C., Kempt, R., Koch, M. H. J., Mclaughlin, S. M., 1986. Data Appraisal, Evaluation and Display for Synchrotron Radiation Experiments: Hardware and Software. Nucl. Instr. Meth. A249: 399 - 345

Cantor, C. R., Schimmel, P. R., 1980. Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function. Freeman & Company, New York, 137pp

Glatter, O., Kratky, O., (Eds) 1982. Small Angle X-Ray Scattering. Academic Press. London, 203pp

Golding, F., 1982. CATY: A System for Experimental Control, Data Collection, Data Display and Analysis. Nucl. Instr. Meth. 201: 231 - 235

Luzzati, V., 1960. Interpretation des Measures Absolues de Difusion Centrale des Rayons-X en Solution en Collimation Pontuelle ou Lineaire: Solution des Particules Globulaires et de Batonnets. Acta Cryst. 13: 939 - 945

Luzzati, V., Tardieu, A., Mateu, L., Stuhrmann, H. B., 1976. Structure of Human Serum Lipoproteins in Solution: I - Theory and Techniques of an X-ray Scattering Approach using Solvents of variable Density. J. Mol. Biol. 101: 115 - 127

Mallam, S., 1987. La Diffusion des Rayons-X par des Macromolecules en Solution: Analyse des Donnes et Etude du Fragment S1 de la Myosine. Rapport de Stage de DEA, Universite Pierre et Marie Curie, Paris-VI, France. 25pp

Marnenout, A., Fransen, L., Tavernier, J., Van Der Heyden, J., Tizard, R., Kawashima, E., Shaw, A., Johnson, M.J., Seman, D., Mueller, R., Ruysschaert, M. J., Van Vliet, A., Fier, W., 1985. Molecular Cloning and Expression of Human Tumour Necrosis Factor and Comparison with Mouse Necrosis Factor. Eur. J. Biochem. 152: 515 - 522

Mestan, J., Digel, W., Mittnacht, S., Hillen, H., Blohm, D., Moller, A., Jacobsen, H., Kirchner, H., 1986. Antiviral Effect of Recombinant Tumour Necrosis Factor in-vitro. Nature, 323: 816 - 819

Murthy, N. S. and Knox, J., 1977. On Soule-Porod Plots of Protein X-ray Scattering Data. J. Appl. Cryst. 10: 137 - 140

Takeda, K., Iwamoto, S., Sugimoto, H., Takuma, T., Kawatani, N., Noda, M., Masaki, A., Morise, H., Arimura, H., Konno, K., 1986. Identity of Differentiation Inducing Factor and Tumour Necrosis Factor. Nature, 323: 338 - 340.

Vachette, P., 1979. Contribution a l'Etude de la Structure du Ribosome d'*Escherichia coli*. Ph.D Thesis, Universite de Strasbourg, France 125pp.

Wingfield, P., Plain, R.H., Craig, S., 1987. Tumour Necrosis Factor is a Compact Trimer, FEBS Lett. 211: 179 -184