Eur Biophys J (1994) 22:379–389

European Biophysics Journal

The aggregation behavior of zinc-free insulin studied by small-angle neutron scattering

Jan Skov Pedersen [1], **Steen Hansen** [2], **Rogert Bauer** [2]

[1] Department of Solid State Physics, Risø National Laboratory, DK-4000 Roskilde, Denmark
[2] Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark

Received: 23 August 1993 / Accepted in revised form: 11 October 1993

Abstract. The aggregation behavior of zinc-free insulin has been studied by small-angle neutron scattering as a function of pH and ionic strength of the solution. The pair distance distribution functions for the 12 samples have been obtained by indirect Fourier transformation. The results show that the diameter of the aggregates is 40 Å at pH 11 and 10 mM NaCl, independent of the protein concentration. The largest diameter of about 120 Å is found for pH 8, 100 mM NaCl, and a protein concentration of 10 mg/ml. Estimates of the pair distance distribution functions, free of inter-particle correlation effects, were obtained by an indirect Fourier transformation, omitting the data at small scattering vectors, which are influenced by these effects. By this procedure the weight-averaged molecular mass and the average radius of gyration were determined. These parameters vary from 1.3 times the monomer mass and 14 Å, to 6.8 times the monomer mass and 31 Å, respectively. The mass distribution between the oligomers was determined by a model based on the crystal structure of zinc-free insulin. The results from this model and the Fourier transformations have been compared to an equilibrium model recently introduced by Kadima et al. (1993). The neutron scattering results agree well with the predictions of this model except that broader mass distributions are suggested by neutron scattering.

Key words: Zinc-free insulin – Aggregation – Small-angle neutron scattering

1. Introduction

The association/dissociation of the protein hormone insulin has been investigated by several experimental techniqes, such as equilibrium sedimentation (Jeffrey and Coates 1966 a, b), static light scattering (Doty et al. 1952; Steiner 1951; Bohidar and Geissler 1984), concentration difference UV spectroscopy (Lord et al. 1973; Strazza et al. 1985), rapid kinetics (Coffman and Dunn 1988), circular dichroism (Goldman and Carpenter 1974; Pocker and Biswas 1981), osmotic pressure (Hansen 1991), and nuclear magnetic resonance (Palmieri et al. 1988; Roy et al. 1990; Kadima et al. 1992). The monomeric species of insulin has a molecular weight of 5800 Daltons. Higher species of insulin are of interest, both biochemically and clinically for diabetes therapy. This is because the hormone functions as the monomer at very low concentration, but is injected at a very high local concentration.

The best characterized structure of insulin is the zinc-insulin hexamers studied by X-ray crystallography (Baker et al. 1988; Hill et al. 1991). The pattern of association of the crystalline hexamer has served as a guide in studies of the association of insulin in solution. Several investigations by sedimentation techniques (Jeffrey and Coates 1966 a, b; Pekar and Frank 1972) and kinetic techniques (Coffman and Dunn 1988) proposed an equilibrium mixture of monomers, dimers, tetramers and hexamers. Higher aggregates of the hexamers have also been proposed (Pekar and Frank 1972). The structure of zinc-free insulin has also been determined (Dodson et al. 1978; Badgar and Caspar 1991; Badgar et al. 1992). This structure is relevant for the present studies of zinc-free insulin.

The aggregation behavior of insulin has recently been studied by dynamic light scattering over a large range of ionic strengths and pH values (Kadima et al. 1993). The results were modeled by a simple equilibrium model having both an attractive part of Gibbs free energy as well as a repulsive term. The latter is dependent on the effective charge which is a function of both the ionic strength and pH.

In the present paper we report the results of a small-angle neutron scattering (SANS) study of zinc-free insulin in solution. The aggregation behavior has been studied as a function of pH and ionic strength of the solutions and as a function of the insulin concentrations. The analysis of the data is divided into two parts. First we have analyzed the data by the (model-independent) indirect Fourier transformation (IFT) method (Glatter 1977) and by this

Correspondence to: J. Skov Pedersen

obtained real space information in terms of the pair distance distribution function $p(r)$. This function is the correlation function of the scattering length density distribution multiplied by the square of the distance. By omitting from the IFT the part of the measured spectra corresponding to small scattering vectors we have also estimated the $p(r)$ functions free of inter-particle correlation effects. These functions can be used for determining the average molecular mass and average radius of gyration of the particles. From the $p(r)$ functions determined by these two approaches information on the inter-particle correlation effects can be derived. In the second part of the analysis we have determined the mass distribution between the various oligomers. This analysis requires a structural model for the oligomers. The model we used was based on the crystal structure of zinc-free insulin (Dodson et al. 1978). Finally we compared the results to those obtained by Kadima et al. (1993) by light scattering experiments and to the predictions of the equilibrium model of Kadima et al. (1993).

2. Material and methods

Sample preparation

All chemicals were of analytical grade. Native human insulin powders were obtained from Novo Nordisk A/S, Novo Allé, DK-2880 Bagsværd, Denmark.

Samples of insulin were prepared by dissolving 50 mg/ml of insulin powder at pH 9.5 and ionic strength 10 mM of NaCl in 99% D_2O. The samples were prepared in D_2O to avoid the large incoherent scattering from hydrogen and to improve the scattering contrast. The solutions were then ultrafiltered (UF) at 4500 rpm in a centrifuge using an Ultrafree CL ultrafiltration unit from Millipore with a cutoff at 100 kDaltons. Subsequently the pH was adjusted to the desired value and NaCl was added as needed. The concentrations of the fractions were determined from absorbance measured at 280 nm after dilution in 20 mM phosphate buffer. The extinction coefficient under these conditions is $E_{1\,\mathrm{cm}\,1\%} = 10.5$ (Sober 1970). Samples were prefiltered on 0.22 μm Millipore filters prior to measurement. The pH values given were direct readings of the pH meter. For comparison with light scattering data in H_2O 0.4 pH units were subtracted. For calculation of the hydrodynamic diameter in D_2O the viscosity of D_2O were used.

Below we will refer to the samples by their ionic strength, pH, and nominal protein concentration C.

Small-angle neutron scattering

The SANS measurements were carried out using the facility at the DR3 reactor at the Risø National Laboratory in Denmark. The instrument is situated in a neutron guide hall and utilizes a cold neutron source. The neutrons are monochromatized by a mechanical velocity selector, which in the present experiments was set at a spread $\Delta\lambda/\lambda = 18\%$ (Full-width-at-half-maximum value). The wavelength λ as well as the collimation length L and sample-detector distance l can be varied. In the present experiments the scattering patterns were recorded for the length of the scattering vector q in the range 0.013–0.500 Å^{-1}. The parameter q is given by $q = 4\pi \sin\theta/\lambda$, where θ is half the scattering angle. The range of q was obtained by two instrumental settings: (*i*) $\lambda = 3.05$ Å, $L = l = 1$ m. (*ii*) $\lambda = 6.73$ Å, $L = l = 3$ m. The radius of the source and sample apertures were 0.8 cm and 0.35 cm, respectively. The effective resolution of the area-sensitive detector of the instrument was 1.0 cm.

The samples were contained in quartz cuvettes (Hellma) with a pathlength of 2 mm. The raw spectra were corrected for background from buffer, cuvette, and other sources by conventional procedures. The intensities were converted to an absolute scale and corrected for variations in detector efficiency by dividing by the scattering spectrum of pure water. The data were normalized by the protein concentration, which was determined by UV absorption.

We performed the following four series of measurements:

1) Concentration series for $C = 1$, 2, 5, 10 mg/ml with 10 mM NaCl at pH 11
2) Concentration series for $C = 1$, 2, 5, 10 mg/ml with 100 mM NaCl at pH 8
3) pH variation for pH 8, 9.5, 11 at $C = 10$ mg/ml with 10 mM NaCl
4) pH variation for pH 8, 9.5, 11 at $C = 10$ mg/ml with 100 mM NaCl.

The concentrations are given as nominal values. The actual values are given in the tables.

Throughout the data analysis the smearing of the ideal cross section $\mathrm{d}\sigma(q)/\mathrm{d}\Omega$ by the instrumental resolution was included as described in Skov Pedersen et al. (1990) and in Hansen and Skov Pedersen (1991).

3. Theory of small-angle scattering

General theory

In the following some theoretical results, which are necessary for the interpretation of the measured small-angle scattering data, are given. As general references for small-angle scattering we suggest e.g. Guinier and Fournet (1955) and Jacrot (1976). For a collection of identical particles the scattering cross section $\mathrm{d}\sigma(q)/\mathrm{d}\Omega$ can be written in terms of the pair distance distribution function $p(r)$:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(q) = 4\pi \int p(r) \frac{\sin(qr)}{qr}\,\mathrm{d}r \tag{1}$$

where $p(r)$ is the self-convolution of the scattering density distribution multiplied by r^2.

For systems with several kind of objects (e.g. different oligomers) $p(r)$ is the weighted sum of the pair distance distribution function of each kind. The values are weighted by the number distribution.

Using the $p(r)$ the forward scattering $d\sigma(q=0)/d\Omega$ can be expressed as:

$$\frac{d\sigma}{d\Omega}(q=0) = 4\pi \int p(r)\, dr, \tag{2}$$

and the average radius of gyration $R_{g,av}$ as:

$$R^2_{g,av} = (\int p(r)\, r^2\, dr)/(2\int p(r)\, dr). \tag{3}$$

The weight-averaged molar mass M_{av} is given by the forward scattering. Using the units mg/ml for the protein concentration C:

$$M_{av} = \frac{d\sigma}{d\Omega}(q=0)\,\frac{10^3}{C N_A \Delta\varrho_m^2}, \tag{4}$$

where N_A is Avogadros number and $\Delta\varrho_m$ is the excess scattering length per unit mass (Dalton) of the protein. The average value $R_{g,av}$ is given as the weighted sum of the radius of gyration $R^2_{g,i}$ of the ith type of particles. The values are weighted by the square of the mass of the particles and the number distribution.

For systems with high concentrations of particles or inter-particle interactions one has:

$$\frac{d\sigma}{d\Omega}(q) = C N_A \Delta\varrho_m^2 M \{P(q)^2\} S(q) \tag{5}$$

where M is the molecular mass and $P(q)$ is the normalized particle form factor:

$$P(\boldsymbol{q}) = \int \varrho(\boldsymbol{r}) \exp(i\boldsymbol{r}\cdot\boldsymbol{q})\, d\boldsymbol{r} / \int \varrho(\boldsymbol{r})\, d\boldsymbol{r}, \tag{6}$$

which fulfills $P(0)=1$. The symbol $\{.\}$ means averaging over all orientations of the molecules. $S(q) = [1 + H(q)\{P(q)\}^2/\{P(q)\}]$ is the effective structure factor and $H(q)$ is the inter-particle structure factor. Except for very strong interactions or very high particle concentration $S(q)$ deviates from unity only at very small values of q. The interference effects are usually observed as a reduction of the intensity at low q values whereas the scattering is unaffected at large q values. If one determines a $p(r)$ function for such a system by IFT one will observe negative values for $p(r)$ at large r values. This is due to a volume around the particle with scattering length density smaller than the average scattering density of the sample caused by exclusion of other particles from this volume by the interactions. For strongly interactions systems one might observe positive values in $p(r)$ at even larger values r due to characteristic distance to a shell of nearest neighbors.

For a polydisperse system the expression corresponding to (5) is:

$$\frac{d\sigma}{d\Omega}(q) = C N_A \Delta\varrho_m^2 \sum_i f_i M_i \{P_i(q)^2\} S(q), \tag{7}$$

where $S(q)$ is an effective structure factor, which is more complicated than for the monodisperse case. The parameter f_i is the mass fraction of the particles with mass M_i and form factor $P_i(q)$. The effective structure factor $S(q)$ behaves similarly to the one for the monodisperse case. In consequence the particle interference effects mainly influence the scattered intensity at low q.

Indirect Fourier transformation of SANS data

We have used the indirect Fourier transformation method of Glatter (1977) for obtaining the pair distance distribution function $p(r)$, with some modifications which we will describe below. In the method by Glatter $p(r)$ is described as a sum of cubic spline functions defined on the interval $[0; D_{max}]$ where D_{max} is the maximum range of the correlations in the system. The Fourier transform of $p(r)$, smeared by instrumental resolution, is fitted to the measured data by a least-squares procedure with a smoothness constraint. We have included an extra parameter to describe the residual background in the data originating from the hydrogen in the insulin. From our application of the method to the data sets with poor counting statistics (low protein concentrations) we have found it useful to change the original constraint of Glatter to:

$$N_c = \sum_{n=1}^{N-1} (a_{n+1} - a_n)^2 + a_1^2 + a_N^2, \tag{8}$$

where $a_1, \ldots, a_N$ are the coefficients of the spline functions. This constraint is a measure of the total length of the curve describing $p(r)$. With this we avoid systematic errors in $p(r)$ from oversmoothing of the solution for data with poor statistics, that are occasionally found for the original constraint of Glatter (1977), which does not have the two last terms in Eq. (8).

In our applications of the IFT method we have chosen the smallest value of D_{max} which gives a smooth solution for $p(r)$. The errors on $p(r)$ and the parameters $d\sigma(q=0)/d\Omega$ and $R_{g,av}$ derived from $p(r)$ were estimated by the Monte-Carlo procedure described by Svergun and Skov Pedersen (1993).

In some of the $p(r)$ functions we have determined for the insulin samples, we have negative values in $p(r)$ due to particle interactions. This effect in $p(r)$ from the interactions (via the structure factor in Eqs. (5, 7)) can be eliminated, as described by Glatter (1979), by omitting the low q part of the data, where the structure factor modifies the data. At the same time the value of D_{max} should be reduced. We estimated the value from the $p(r)$ functions obtained by fitting the full q range. The value of D_{max} was chosen as the r value for which $p(r)$ has the lowest value. The lower limit for the q values included in the fit was chosen so that agreement with the data was obtained in the entire fitted q range for this value of D_{max}.

From the IFT of the full and of the restricted q range we get estimates of the $p(r)$ functions of the system with and without particle interactions, respectively. From the corresponding two values of the forward scattering we obtain an estimate of the effective structure factor $S(q)$ for $q=0$ (cf. Eqs. (6, 7)). The value of $S(0)$ is given as the forward scattering fo the full q range divided by the forward scattering for the restricted q range. The average radius of gyration $R_{g,av}$ is obtained from the fit to the restricted q range. All the results from the IFT method are *model independent* in the sense that they do not depend on a structural model for the oligomers.

Mass distribution from SANS data

In principle the mass distribution f_i between the various oligomers can be determined using (7). However, this would require the form factor $P(q)$ of the oligomers and the effective structure factor $S(q)$ to be known. The influence of the latter can be eliminated by fitting to the data in the restricted q range that was used for the IFT for obtaining the $p(r)$ functions free of particle interaction effects. The form factors of the oligomers can be calculated only if one assumes structural models for the oligomers. This makes the determined mass distribution *model dependent.*

The crystal structure of zinc-free insulin is known (Dodson et al. 1978; Badgar and Caspar 1991; Badgar et al. 1992). We have therefore based the structural models on the crystal structure of zinc-free pig insulin with the cubic structure (Dodson et al. 1978). Note that the crystal structure of zinc-free insulin is distinctly different from the structure of zinc-containing insulin, in which the zinc atoms bind the insulin dimers together in hexamers (see e.g. Dodson et al. 1966; Baker et al. 1988). The structure of zinc-free insulin is shown in Fig. 1, which contains a schematic drawing of the unit cell. The insulin is present as dimers which are the elliptically shaped units in the figure. The dimers form rows by lining up in their longitudinal direction in all three spatial directions. Furthermore the structure has open channels running through the crystal. This makes it a very open structure with a mass fraction of water as high as 61%.

The form factors $P_i(q)$ of the oligomers are calculated by taking a subset of the unit cell shown in Fig. 1. As the model is based on the cubic structure for insulin we will only include the oligomers with aggregation numbers $n_a = 1, 2, 4, 6, 8, \ldots$. These numbers correspond to the indices $i = 1, 2, 3, 4, \ldots$, respectively, in (7). There are several ways of selecting e.g. a hexamer from the crystal structure, but for simplicity we will assume that a hexamer can be represented by one particular subset. The numbers in Fig. 1 indicate the order in which the dimers are included in the oligomers. The hexamer, for example, is taken as the dimers with numbers 1, 2, and 3.

As previously noted, the mass distributions are dependent on the assumed oligomer structure. The structures of the oligomers based on the crystal structure are quite open and the analysis using these structures will give one particular set of mass distributions. If structures with a more compact arrangement of the monomers and dimers are assumed one will obtain different mass distributions, which are broader.

The size and shape of the monomer can be estimated from the crystal structure (see e.g. Baker et al. 1988). The monomer has a size of about $40 \times 20 \times 10\,\text{Å}^3$. We will approximate the monomer structure by a set of eight spheres with a radius of 5.0 Å placed as shown in Fig. 2. The distance between the nearest spheres is 10 Å. This "coarse graining" of the structure is sufficiently accurate as the real space resolution of the data is $\pi/q_{\max} = 6.3$ Å, according to the sampling theorem. The parameter $q_{\max}$ is the largest q value probed in the scattering measurements. The form factor of the monomer is calculated by the

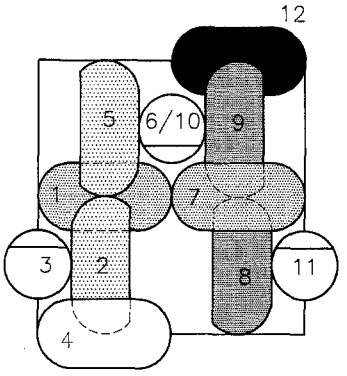

Fig. 1. Unit cell of zinc-free insulin (after Dodson et al. 1978). Each of the ovals represents an insulin dimer. The shading indicates the position of the dimers in the direction perpendicular to the plane of the paper. A darker shading means that the dimer is at a lower position. The circles indicate the position of the dimers which have the long axis perpendicular to the plane of the paper

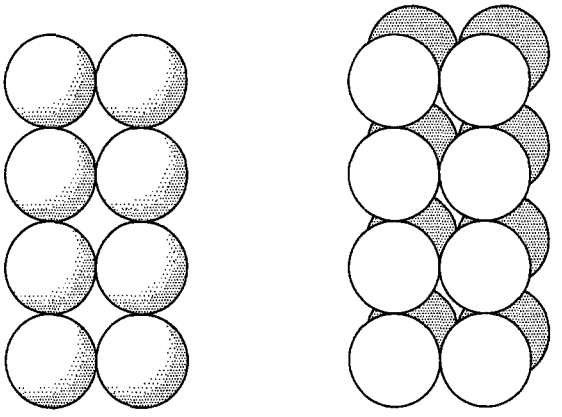

Fig. 2. *Left-hand side:* The spheres used for calculating the form factor of a monomer. *Right-hand side:* The spheres used for calculating the form factor of a dimer

Debye formula:

$$\{P(q)\}^2 = \frac{\Phi(q)^2}{N_p^2} \sum_{i,j}^{N_p} \frac{\sin(q r_{ij})}{q r_{ij}} \tag{9}$$

where r_{ij} is the distance between the ith and the jth particle, N_p is the number of subunits, and $\Phi(q)$ is the form factor of a sphere, $\Phi(q) = 3[\sin qR - qR\cos qR]/(qR)^3$, where R is the radius of the sphere. The form factor $\Phi(q)$ has its first zero at $q \approx 4.5/R$, which is far outside the range of the measured data. Therefore the characteristic of the form factor of the spheres are not directly observed in the measured q range.

In analogy with the calculation for the monomer, the form factor of the dimer is calculated by placing two monomers to form the structure shown in Fig. 2 with the size of about $40 \times 20 \times 20\,\text{Å}^3$. The form factor of the higher oligomers were calculated similarly by placing the dimer in accordance with the structure shown in Fig. 1.

The scattering cross section of the model is given by (7). As both the molecular mass of the insulin monomer (5800) and the contrast factor of the protein are known the absolute value of the form factors of the oligomers can

be used in (7). For the contrast of the protein in D_2O we used the value $\Delta\varrho_m = 4.06 \times 10^{-14}$ cm/D.

The distribution of the masses between the various oligomers is a result of a thermodynamic equilibrium, and one can expect the distribution to be a smooth function. Therefore the mass distribution parameters f_i were calculated by a constrained least-squares method similar to the one used for determining the parameters for the IFT method. The constraint was:

$$N_c = \sum_{i=1}^{N-1} (f_{i+1} - f_i)^2 + f_N^2. \tag{10}$$

This gives a solution which goes smoothly to zero at large i. As the parameters f_i describe mass fraction, these parameters are non-negative. This was included in the fitting as a non-negativity constraint using a procedure similar to the one described by Lawson and Hanson (1974). In addition to the parameters for the mass distribution we included a parameter to describe residual background in the spectra. This parameter was unconstrained. The weight of the constraint relative to the mean-squared residual was determined by performing a variation of the weight over several decades (see Glatter 1977). Plotting the logarithm of N_c versus the logarithm of the weight gives a curve which displays a point of inflection. The sum $\sum_i f_i$ attained its maximum value for a value of the weight which was slightly smaller than the value corresponding to the inflection point. For this value there is already an increase in the mean-squared residual by about a factor of two. Thus it was concluded that the solution is influenced by the constraint at this value. For all the fits the weight of the constraint was chosen at the value where $\sum_i f_i$ has its maximum value.

As a consistency check of the model one can see if the sum $\sum_i f_i$ is equal to unity. From the mass distribution one can also calculate an alternative estimate for the weight-averaged molecular mass:

$$M_{av} = \sum_i f_i M_i. \tag{11}$$

Furthermore one can calculate the variance σ_M of the mass distribution from:

$$\sigma_M^2 = \sum_i (f_i M_i - M_{av})^2. \tag{12}$$

The parameter σ_M / M_{av} is a measure of the polydispersity of the samples.

The errors on the mass distribution f_i were determined by conventional error calculations for least-squares methods (see e.g. Bevington 1969; Glatter 1977). We will not take these errors too literally as they are dependent on the weight of the constraint. The errors are in fact mainly determined by the systematic errors that might appear owing to inaccurate or wrong estimates of the form factors of the oligomers.

4. Results

Two typical small-angle neutron scattering spectra are shown in Fig. 3, for a protein concentration of 10 mg/ml. The two spectra are for (a) an ionic strength of 10 mM NaCl pH 11 and (b) an ionic strength of 100 mM NaCl pH 8, respectively. These two samples represent the two extreme situations of the aggregation behavior as will be shown in greater detail by the data analysis. However, one can observe significant differences directly from the spectra. The scattering intensity is much higher for (b) than for (a) and the form of the scattering curve is much narrower for (b) than for (a). Both of these features show that the particles are much larger in (b) than in (a). For the data (a) one observes a dip in the intensity at low q which indicates the presence of particle interactions.

Indirect Fourier transformation of the SANS data

Figure 4 shows the results for $p(r)$ from IFT for the four series of measurements. The lines in Fig. 3 are the fits from IFT. The full lines are the cross section smeared by instrumental resolution and the dotted curve is the non-smeared cross section. The full measured range of the data was used. For the concentration series with 10 mM NaCl and at pH 11 (a), the functions nearly coincide for $r < 30$ Å, where the functions are determined by the particle self-correlations. At large distances the $p(r)$ functions for $C = 2$, 5, 10 mg/ml have negative values owing to inter-particle correlations. The structure of the functions indicate a maximum size of the particles of about 40 Å in agreement with the size of the monomer in the crystal structure (Dodson et al. 1978; Baker et al. 1988).

Part (b) of Fig. 4 shows $p(r)$ for the concentration series with 100 mM NaCl and pH 8. As expected from the scattering curves these functions reveal a significantly larger size of the particles. In this case the size depends on the protein concentration. For the four concentrations $C = 1$, 2, 5, 10 mg/ml the maximum sizes of the particles

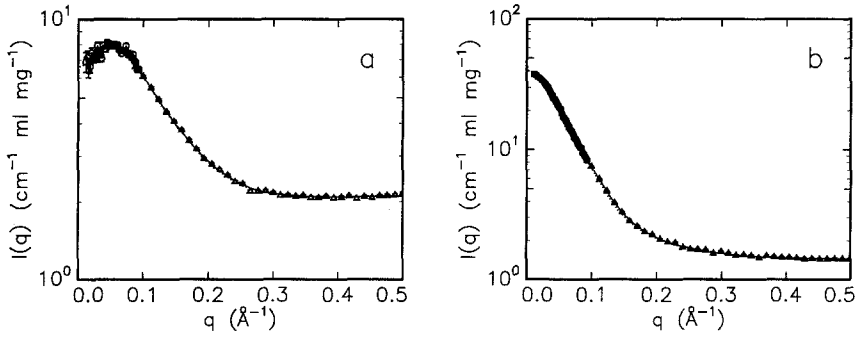

Fig. 3 a, b. Typical SANS spectra. Both data sets are for a protein concentration of 10 mg/ml. **a** Ionic strength of 10 mM NaCl pH 11. **b** Ionic strength of 100 mM NaCl pH 8

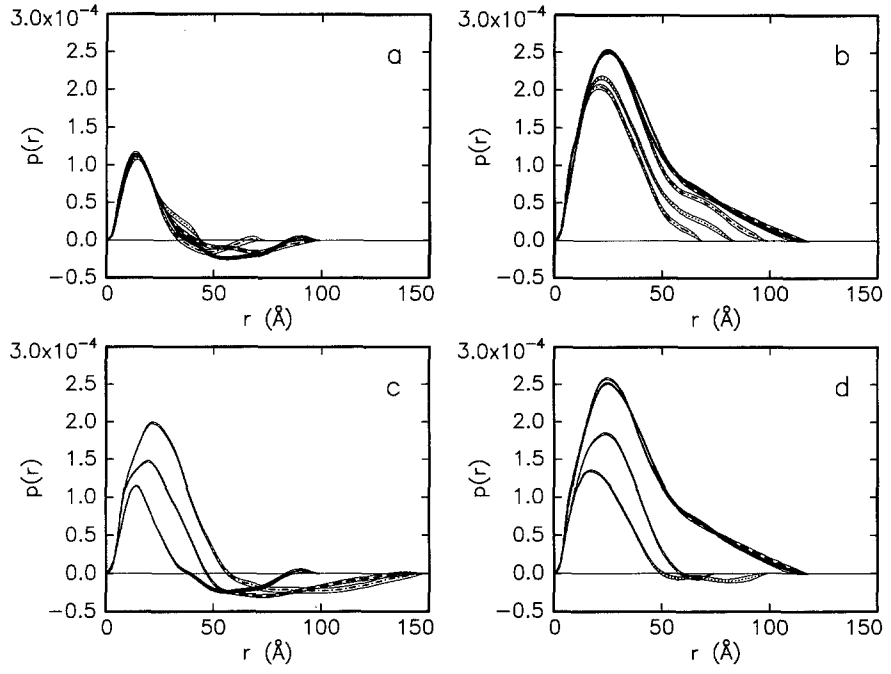

Fig. 4a–d. The pair distance distribution functions $p(r)$ for the four series of measurements obtained by IFT of the full range of the measured data. The error bands are indicated by the light full lines. **a** Concentration series for 10 mM NaCl at pH 11. $C = 1$ mg/ml: (···), $C = 2$ mg/ml: (–·–·–), $C = 5$ mg/ml: (– – –), and $C = 10$ mg/ml: (——). **b** Concentration series for 100 mM NaCl pH 8. $C = 1$ mg/ml: (·–·–·–·), $C = 2$ mg/ml: (···), $C = 5$ mg/ml: (–·–·–·), $C = 10$ mg/ml: (– – –) and (——). **c** pH variation for pH 8 (–·–·–·), 9.5 (– – –), and 11 (——) at $C = 10$ mg/ml with 10 mM NaCl. **d** pH variation for pH 8 (– – –) and (——), 9.5 (–·–·–·), 11 (···) at $C = 10$ mg/ml with 100 mM NaCl

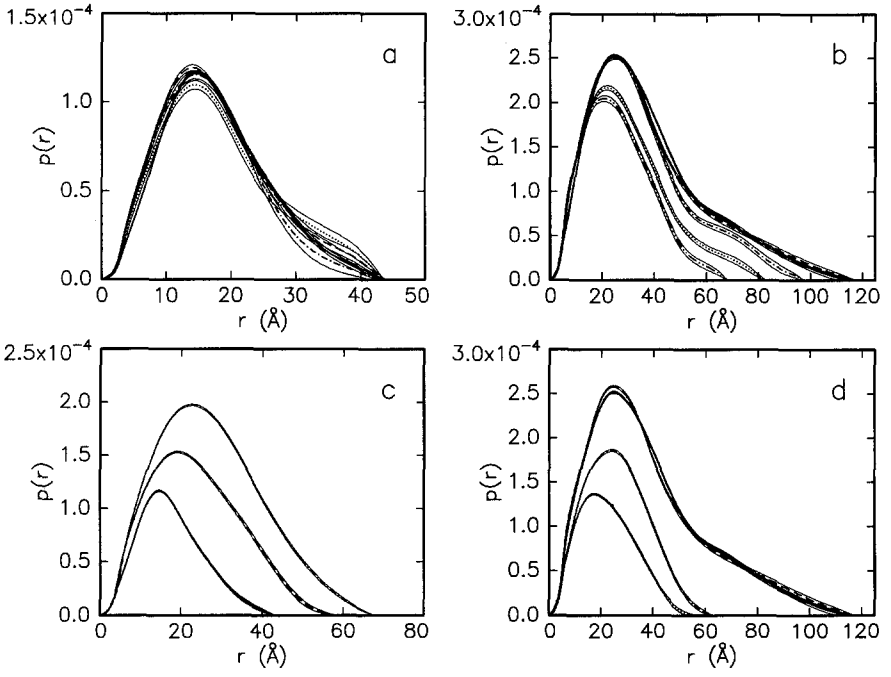

Fig. 5 a–d. The pair distance distribution functions $p(r)$ for the four series of measurements obtained by IFT of the restricted ranges of the measured data, which give positive $p(r)$. The error bands are indicated by the light full lines. **a** Concentration series for 10 mM NaCl at pH 11. $C = 1$ mg/ml: (···), $C = 2$ mg/ml: (–·–·–), $C = 5$ mg/ml: (– – –), and $C = 10$ mg/ml: (——). **b** Concentration series for 100 mM NaCl pH 8. $C = 1$ mg/ml: (·–·–·–·), $C = 2$ mg/ml: (···), $C = 5$ mg/ml: (–·–·–·), $C = 10$ mg/ml: (– – –) and (——). **c** pH variation for pH 8 (–·–·–·), 9.5 (– – –), and 11 (——) at $C = 10$ mg/ml with 10 mM NaCl. **d** pH variation for pH 8 (– – –) and (——), 9.5 (–·–·–·), 11 (···) at $C = 10$ mg/ml with 100 mM NaCl

are $D_{max} = 70$, 85, 100, 120 Å, respectively. The largest changes in $p(r)$ are observed between the three lowest concentrations. The functions do not show any signs of inter-particle correlations.

Part (c) of Fig. 4 shows $p(r)$ for the pH variation with $C = 10$ mg/ml and 10 mM NaCl. All functions show the presence of inter-particle correlations. The functions further show that lowering of pH leads to an increase in the maximum particle size, from 40 Å at pH 11, to 65 Å at pH 9.5, and to 75 Å at pH 8. Part (d) of Fig. 4 shows the results for the pH variation with $C = 10$ mg/ml and 100 mM NaCl. The increase in pH leads to changes in the maximum particle size from 120 Å at pH 8, to 75 Å at pH 9.5, and to 55 Å at pH 11. The $p(r)$ functions for pH 8 and 9.5 both show that inter-particle correlations are present.

Figure 5 shows the results for $p(r)$ obtained from the fits to the restricted q ranges, which give positive $p(r)$ functions. As already mentioned these functions are estimates of the $p(r)$ functions expected for similar particles in the absence of particle interactions. For 100 mM NaCl and pH 11 it was not necessary to restrict the q range. The result is that all the $p(r)$ functions from the concentration variation series with 10 mM pH 11 (a) nearly coincide. This shows that the particles shape is nearly independent of the protein concentration, and furthermore it gives confidence that the inter-particle correlation effects are eliminated by the applied procedure. The $p(r)$ functions

Table 1. Results from indirect Fourier transformation of the SANS spectra. The subscript "c" refers to parameters corrected for correlation effects by the use of a restricted q range for the IFT. M_{mono} is the monomer molecular mass (5800). The parameter D_h is the hydrodynamic diameter as measured by dynamical light scattering. The results from light scattering are based on the measurements and model by Kadima et al. (1993)

				Neutron scattering					Light scattering + model				
#	NaCl [mM]	pH	C [mg/ml]	$\frac{d\sigma(q=0)}{d\Omega}$ [ml(mg cm)$^{-1}$]	$\left(\frac{d\sigma(q=0)}{d\Omega}\right)_c$ [ml(mg cm)$^{-1}$]	$S(q=0)$	$(M_{av})_c$ [M_{mono}]	$(R_{g,av})_c$ [Å]	$S(q=0)$	M_{av} [M_{mono}]	D_h [Å]	$2R_{g,av}/D$	Z_{eff}
1	10	11	1.2	7.52±0.28	7.52±0.28	1.00	1.31	14.5±0.50	0.84	1.2	32	0.91	5.8
2	10	11	2.2	5.82±0.28	7.56±0.40	0.76	1.33	13.1±0.80	0.73	1.3	33	0.73	5.7
3	10	11	5.0	5.94±0.77	8.03±0.16	0.74	1.39	13.9±0.27	0.50	1.6	35	0.79	5.6
4	10	11	9.1	4.50±0.31	7.56±0.10	0.60	1.31	13.9±0.17	0.37	1.7	36	0.77	5.4
5	10	9.5	9.8	7.54±0.43	14.9±0.22	0.51	2.59	18.6±0.20	0.51	2.5	42	0.87	3.2
6	12	8	9.6	16.1±0.62	22.0±0.18	0.73	3.83	21.6±0.12	0.69	4.1	51	0.85	1.8
7	100	8	1.1	22.3±0.62	22.3±0.62	1.00	3.88	21.2±0.70	0.99	4.0	51	0.83	1.0
8	100	8	2.3	27.1±0.46	27.1±0.46	1.00	4.71	24.4±0.50	0.98	4.8	54	0.90	1.0
9	101	8	5.3	34.9±0.71	34.9±0.71	1.00	6.05	28.1±1.60	0.96	5.5	57	0.99	1.0
10	102	8	9.7	39.4±0.39	39.4±0.39	1.00	6.84	31.3±0.50	0.92	6.2	60	1.04	1.0
11	100	9.5	9.8	20.0±0.43	20.5±0.16	0.98	3.89	20.5±0.40	0.85	4.1	51	0.80	1.8
12	100	11	8.7	11.7±0.30	12.9±0.17	0.91	2.25	17.8±0.23	0.76	2.7	43	0.83	3.0

for (c) and (d) in Fig. 5 confirm the conclusions drawn from Fig. 4 concerning the maximum particle sizes.

The results derived from the $p(r)$ functions obtained by IFT are summarized in Table 1. The subscript "c" on $d\sigma(q=0)/dM$, M_{av}, $R_{g,av}$ refers to parameters corrected for correlation effects by the use of a restricted q range for the IFT. The values of M_{av} are calculated by (11) using $\Delta\varrho_m = 4.06\times10^{-14}$ cm/D, and they are given in units of the monomer mass of 5800. The values of the effective structure factor $S(q=0)$ are calculated as the ratio between the uncorrected and corrected forward scattering. We estimate the errors on M_{av} and $S(0)$ to be about 10–20% including both statistical and systematic errors. These are connected with the determination of the protein concentrations, the absolute calibration of the scattering data, and the value of the excess scattering length.

Except for the lowest protein concentration ($C=1$ mg/ml), the scattering for systems with a low ionic strength (10 mM NaCl) is influenced by inter-particle correlation effects. This is seen in $S(q=0)$ which has values in the range 0.5–0.76. For the samples with high ionic strength the particle correlations effects are much smaller, as the $S(q=0)$ values are close to unity.

For 10 mM NaCl and pH 11 the weight-averaged molecular mass M_{av} is about 1.3 M_{mono} independent of the protein concentration C. It is therefore likely that the samples contain only monomers and dimers. Lowering pH to 8 gives an increase in M_{av} to 3.8 M_{mono}. For 100 mM NaCl and pH 8 M_{av} varies from 3.9 to 6.8 M_{mono} with an increase in C from 1 mg/ml to 10 mg/ml. Increasing pH to 11 gives rise to a decrease in M_{av} to 2.3 M_{mono}.

For low ionic strength and pH 11 the radius of gyration $R_{g,av}$ is about 14 Å. Lowering pH gives an $R_{g,av}$ of 22 Å at pH 8. For high ionic strength $R_{g,av}$ varies from 21 Å at $C=1$ mg/ml to 31 Å for 10 mg/ml. Increasing pH changes $R_{g,av}$ to 18 Å at pH 11.

For monodisperse systems the compactness of the objects can be estimated from the dependence of R_g on the molecular mass M:

$$R_g \propto M^{\alpha}. \tag{13}$$

For $\alpha=1/3$ one has compact objects, whereas $\alpha=1/2$ corresponds to the open structure of a random Gaussian coil. A plot of $R_{g,av}$ versus M_{av} for all 12 samples gives the parameter $\alpha\approx0.45$. This corresponds to relatively open structures, however one should bear in mind that changes in polydispersity might influence the value of α.

Mass distributions from the SANS data

For the fitting of the model described in the previous section the restricted q ranges of the data, used in the IFT for eliminating correlation effects, were used. This is necessary as the model does not include the particle correlation effects. Figure 6 shows the fits of the model to two typical SANS spectra. These are for the samples with $C=10$ mg/ml and, respectively, (a) 10 mM NaCl pH 11 and (b) 100 mM NaCl pH 8. The fit to (a) is perfect, whereas some smaller deviations are seen for (b) around $q=0.3$ Å^{-1}.

The mass distributions from the model fits are shown in Figs. 7 and 8. Figure 7 shows the dependence on protein concentration. Part (a) and (b) show the results for the two concentrations $C=1$ and 10 mg/ml, respectively, with 10 mM NaCl at pH 11. It is seen that the two samples have nearly equal amounts of monomers and dimers and only a small fraction of tetramers. Part (c) and (d) show the results for $C=1$ and 10 mg/ml, respectively, with 100 mM NaCl at pH 8. The smooth mass distribution increases in width for increasing concentration. For the concentration $C=1$ mg/ml the mass distribution goes to zero at $n_a=4$, where n_a is the aggregation number. For

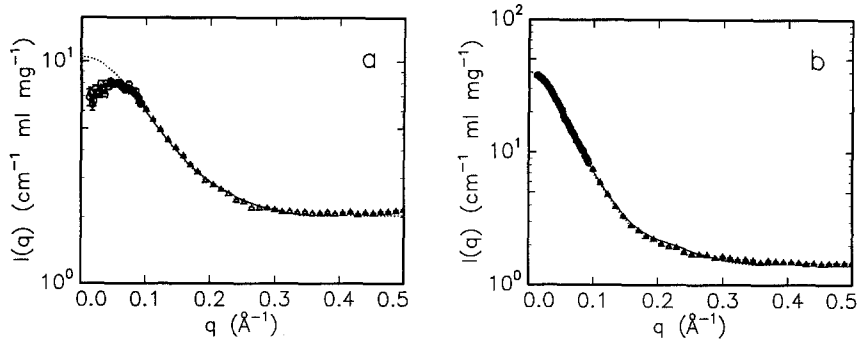

Fig. 6 a, b. Fit of the mass distribution model to two typical SANS spectra. Both data sets are for a protein concentration of 10 mg/ml. **a** Ionic strength of 10 mM NaCl pH 11. **b** Ionic strength of 100 mM NaCl pH 8

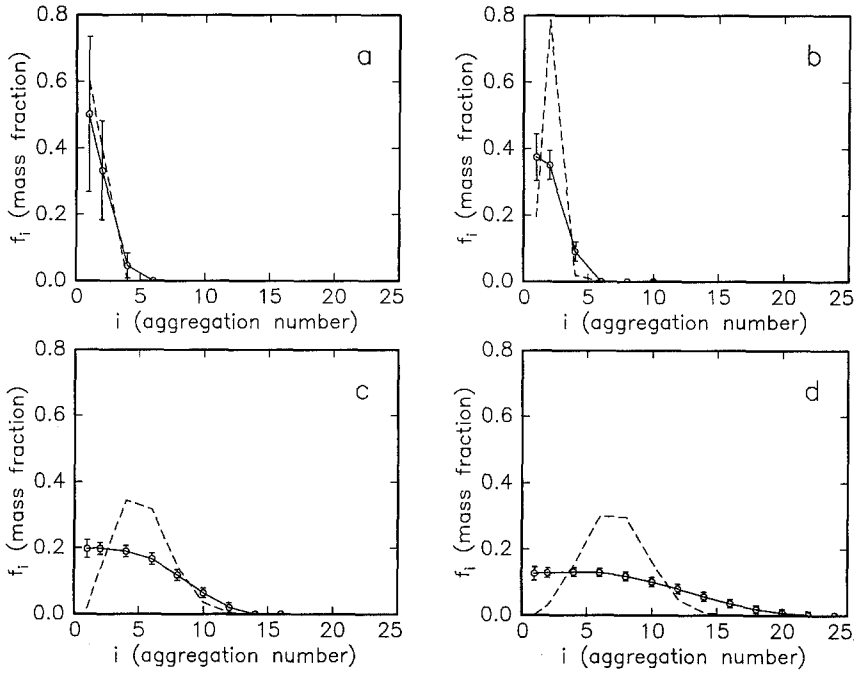

Fig. 7 a–d. Mass distribution f_i obtained by fitting the model described in Sect. 3. **a** 10 mM NaCl, pH 11, and $C = 1$ mg/ml. **b** 10 mM NaCl, pH 11, and $C = 10$ mg/ml. **c** 100 mM NaCl, pH 8, and $C = 1$ mg/ml. **d** 100 mM NaCl, pH 8, and $C = 10$ mg/ml. The broken curves are the mass distributions calculated by the equilibrium model of Kadima et al. (1993)

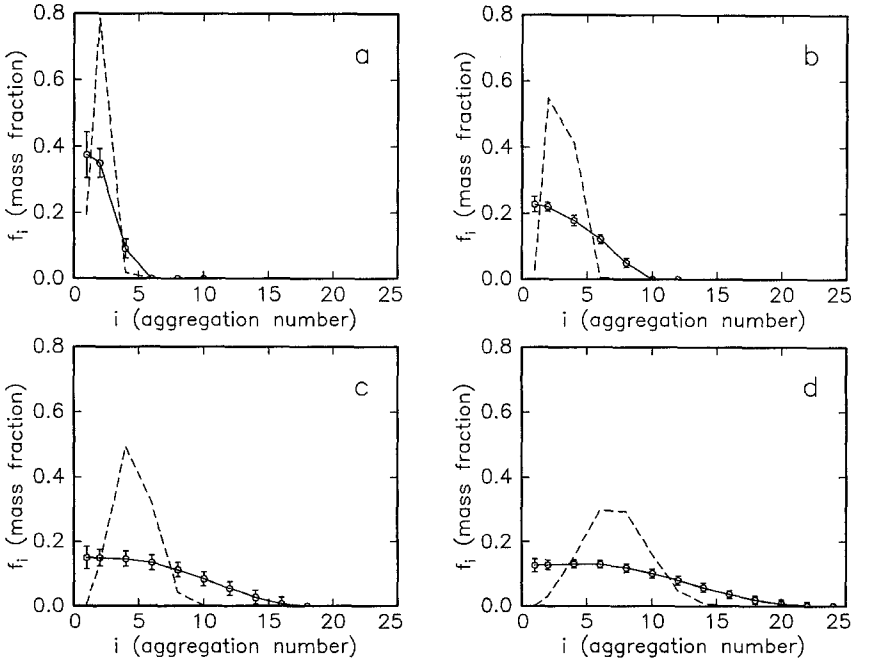

Fig. 8 a–d. Mass distribution f_i obtained by fitting the model described in Sect. 3. **a** 10 mM NaCl, pH 11, and $C = 10$ mg/ml. **b** 100 mM NaCl, pH 11, and $C = 10$ mg/ml. **c** 10 mM NaCl, pH 8, and $C = 10$ mg/ml. **d** 100 mM NaCl, pH 8, and $C = 10$ mg/ml. The broken curves are the mass distributions calculated by the equilibrium model of Kadima et al. (1993)

Table 2. Results from the mass distributions from the fits of the model described in Sect. 3. The values of $S(q=0)$ are calculated as the forward scattering from IFT of the full q range divided by the forward scattering from the fits of the mass distribution model. M_{mono} is the monomer molecular mass (5800). The results from light scattering are based on the measurements and model by Kadima et al. (1993)

				Neutron scattering						Light scatt. + model		
#	NaCl [mM]	pH	C [mg/ml]	$\left(\frac{d\sigma(q=0)}{d\Omega}\right)_c$ [ml (mg cm)$^{-1}$]	$S(q=0)$	$(M_{av})_c$ [M_{mono}]	σ_M [M_{mono}]	$\sigma_M/(M_{av})_c$	$\sum_i f_i$	$S(q=0)$	M_{av} [M_{mono}]	Z_{eff}
1	10	11	1.2	7.76	0.97	1.53	0.60	0.37	0.88	0.84	1.2	5.8
2	10	11	2.2	7.59	0.77	1.44	0.50	0.35	0.92	0.73	1.3	5.7
3	10	11	5.0	8.36	0.71	1.56	0.58	0.37	0.94	0.50	1.6	5.6
4	10	11	9.1	8.28	0.54	1.76	0.68	0.39	0.82	0.37	1.7	5.4
5	10	9.5	9.8	18.4	0.41	3.87	1.84	0.48	0.82	0.51	2.5	3.2
6	12	8	9.6	27.2	0.59	5.45	2.72	0.50	0.86	0.69	4.1	1.8
7	100	8	1.1	24.2	0.92	4.93	2.08	0.42	0.96	0.99	4.0	1.0
8	100	8	2.3	27.5	0.99	5.18	2.50	0.48	0.92	0.98	4.8	1.0
9	101	8	5.3	35.9	0.97	6.45	3.30	0.51	0.97	0.96	5.5	1.0
10	102	8	9.7	38.1	1.03	7.27	3.72	0.51	0.91	0.92	6.2	1.0
11	100	9.5	9.8	25.3	0.79	5.43	2.68	0.49	0.81	0.85	4.1	1.8
12	100	11	8.7	14.5	0.81	3.14	1.41	0.45	0.80	0.76	2.7	3.0

$C = 10$ mg/ml the largest oligomers have increased in size to $n_a = 20-22$.

Figure 8 shows the dependence on ionic strength and pH. Part (a) and (b) of Fig. 8 show the mass distribution of the ionic strength of 10 and 100 mM NaCl, respectively, for $C = 10$ mg/ml and pH 11. The width of the mass distribution increases for increasing ionic strength. For 10 mM NaCl the distribution goes to zero at $n_a = 6$, whereas for 100 mM NaCl it goes to zero at $n_a = 10$. Part (c) and (d) show the corresponding changes at pH 8 for $C = 10$ mg/ml. Part (c) is for 10 mM NaCl and (d) is for 100 mM NaCl. In this case the increase in ionic strength also leads to broader mass distributions.

The changes upon a decrease of pH can be seen by comparing Fig. 8 a with c and Fig. 8 b with d. A decrease in pH gives rise to broadening of the mass distribution and formation of higher oligomers.

The parameters derived from the mass distributions are given in Table 2. In addition the table contains estimates of $S(0)$ calculated as the forward scattering from IFT of the full q range divided by the forward scattering determined by the mass distribution fits. The values of the forward scattering $(d\sigma(q=0)/d\Omega)_c$ are similar to those determined from the IFT of the restricted q range. Therefore the values for $S(0)$ are also similar to the values from IFT. Also the values of the average molecular mass M_{av} are similar to the values determined by the IFT. The errors on M_{av} and $S(0)$ are in this case also about 10–20%. The polydispersity factor σ_M/M_{av} varies between 0.35 and 0.51. In general, the polydispersities are larger for the high ionic strength as compared to the low ionic strength. This could indicate that the polydispersity is larger for samples with larger M_{av}. The sum of the mass fractions varies between 0.80 and 0.97, which means that some of the mass is not properly accounted for in the model. A possible source of systematic errors is the smoothness constraint, which is likely to give rise to an underestimation of the fraction of the smallest oligomers. As $I(0)$ is proportional to the mass, the oligomers give only a small contribution to the total scattering. Additional sources of errors are the determination of the protein concentrations and the absolute calibration of the scattering data, and inaccuracy of the value of the excess scattering length.

5. Discussion

Recently, Kadima et al. (1993) have studied the aggregation properties of zinc-free insulin by static and dynamic light scattering. Kadima et al. introduced an equilibrium model, which we will use in the interpretation of the SANS results. The insulin molecules are charged and have significant interactions particularly in the highly charged state. The quantities measured by light scattering are the *apparent* molecular mass and the *apparent* hydrodynamic diameter. In order to give a useful interpretation of these quantities a model for the interactions and the aggregation had to be proposed (Kadima et al. 1993). The model is a thermodynamic model with a Gibbs free energy expression with a positive contribution from charge repulsion between the oligomers and a negative contribution from entropy or energy gain upon association. The latter contribution was taken as a constant per monomer, whereas the repulsion term was taken to depend on charge and numbers of monomers in the various oligomers. Three empirical parameters enter the model: The two characteristic energies and a parameter describing the dependence of the charge of the insulin molecule on the ionic strength of the solvent. When the values of these parameters and the characteristic of the solution (pH, ionic strength, and protein concentration) are specified, the mass distribution of the oligomers can be calculated.

The model provides a means for expressing the various pH values and ionic strengths in terms of one parameter, which is the effective charge Z_{eff} of the monomers. The

charge Z_0 of the insulin molecules in 100 mM KCl as a function of pH has been determined by proton titration (Kaarsholm et al. 1990).

The effective charge Z_{eff} is proposed to be given as (Kadima et al. 1993):

$$Z_{eff} \approx \frac{Z_0}{1 + [Na^+]/K_{Na}}, \quad (14)$$

where $[Na^+]$ is molar concentration of Na^+ ions and K_{Na} is the binding constant of Na to the insulin molecule. With this equation the effective charges for the solution can be calculated using the results of Kaarsholm et al. (1990) for Z_0. The effective charges are given in Tables 1 and 2.

Virial effects on both the static and the dynamic light scattering results were assumed to be due to the Donnan effect from charge repulsion. For the light scattering experiments a model was used to analyze the data owing to the difficulty in separating virial effects from equilibrium properties. The difficulty is caused by the near-isotropy of the light scattering. For neutron scattering the data include a q range which gives structural information about the aggregates. Therefore a comparison of the forward scattering ($S(q = 0)$) and the aggregation number for the two types of experiments test the reliability of the model used in Kadima et al. (1993).

Qualitatively, the forward scattering (Tables 1 and 2) agrees with that of light scattering, i.e. decreasing pH or increasing ionic strength leads to a decrease in the forward scattering for both light and neutron scattering. Also the aggregation numbers from neutron scattering agree with those of light scattering showing an increase with decreasing pH or decreasing ionic strength. This supports the weak binding of the sodium ions and insulin suggested by Kadima et al. (1993).

In Figs. 7 and 8 the mass distributions from the small-angle scattering experiments are compared to those obtained with the equilibrium model. The broken curves are the results calculated with the equilibrium model. In order to make a direct comparison possible the mass fraction of the oligomers with odd aggregation numbers (larger than one), have been divided by two and added to the mass fractions of the two closest oligomers with even aggregation number. For the mass fraction both the neutron scattering results and those of light scattering are model dependent. In general, broader distributions are observed for neutron scattering. This could be due to the systematic error from the constraints (10) used for the analysis of the neutron data (and the assumption of the cubic insulin structure) or it could indicate limitations in the model used for the analysis of the light scattering data. However the trend with respect to variations in concentration, pH and ionic strength is the same for neutron and light scattering.

The ratio $2R_{g,av}/D_h$ of twice the radius of gyration to the hydrodynamic diameter determined by light scattering (Kadima et al. 1993) is also given in Tabe 1. It varies from 0.7 to 1.0, with a tendency for the value to increase with increasing size or mass of the aggregates. This indicates that the larger aggregates are relatively branched structures in accordance with conclusions drawn from the scaling relation between the radius of gyration and the mass (Eq. 13).

A discussion of the agreement of the insulin aggregation model with the previous equilibrium sedimentation experiments (Jeffrey and Coates 1966 a, b), NMR experiments (Kadima et al. 1992), and osmotic pressure measurements (Hansen 1991) is given by Kadima et al. (1993). The agreement is good when appropriate corrections for virial effects are made. Note finally that the equilibrium model was tested by the light scattering experiments described in Kadima et al. (1993) over a wider range of experimental conditions than used for the present neutron scattering experiments.

6. Summary and conclusions

The present work has demonstrated that detailed information on the aggregation behavior of insulin can be obtained from small-angle scattering experiments. The change in the structure and in aggregation state has been studied as a function of protein concentration, pH, and ionic strength. It has been shown that the indirect Fourier transformation of the scattering spectra can be used for eliminating particle interaction effects and that the distance distribution functions of the independent particles can be determined. From the distance distribution function the radius of gyration and the weight-average molecular weight were obtained. It was further shown that for highly charged, and thus highly interacting insulin (high pH, low ionic strength) the insulin structure is independent of protein concentration. Mass distribution were obtained by assuming structures of the oligomers based on the crystal structure of zinc-free insulin. The results from indirect Fourier transformation and from the determination of the mass distributions support the aggregation model of Kadima et al. (1993). It should be noted that the results derived from the small-angle scattering data are derived independently of the light-scattering results and model.

Acknowledgements. The insulin powders were kindly provided by Novo Nordisk A/S. We thank K. Brodersen for the skilled technical assistance with the sample preparations. The financial support from the Danish Natural Science, Technical, and Veterinary and Agricultural Research Councils, and the Swedish Natural Science Research Council is gratefully acknowledged.

References

Badgar J, Caspar DLD (1991) Water structure in cubic insulin crystals. Proc Natl Acad Sci, USA 88:622–626

Badger J, Harris MR, Reynolds CD, Evans AC, Dodson EJ, Dodson GG, North ACT (1992) Structure of the pig insulin dimer in the cubic crystal. Acta Crystallogr B47:127–136

Baker EN, Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, Crowfoot Hodgkin DM, Hubbard RE, Isaacs NW, Reynolds CD, Sakabe K, Sakabe N (1988) The structure of 2 Zn pig insulin crystals at 1.5 Å resolution. Philos Trans R Soc London B319:369–456

Bevington BR (1969) Data reduction and error analysis for the physical sciences. McGraw-Hill, New York

Bohidar HB, Geissler E (1984) Static and dynamic light scattering from insulin solutions. Biopolymers 23:2407–2417

Coffman FD, Dunn MF (1988) Insulin-metal ion interactions: The binding of divalent cations to insulin hexamers and tetramers and the assembly of the insulin hexamers. Biochemistry 27: 6179–6187

Dodson EJ, Harding MM, Crowfoot Hodgkin (1966) The crystal structure of insulin: III. Evidence for 2-fold axis in rhombohedral zinc insulin. J Mol Biol 16:227–241

Dodson EJ, Dodson GG, Lewitowa A, Sabesan M (1978) Zinc-free pig insulin: crystallization and structure determination. J Mol Biol 125:387–396

Doty P, Gellert M, Rabinovitch B (1952) The association of insulin. I. Preliminary investigations. J Am Chem Soc 74:2065–2069

Glatter O (1977) A new method for the evaluation of small-angle scattering data. J Appl Crystallogr 10:415–421

Glatter O (1979) The interpretation of real-space information from small-angle scattering experiments. J Appl Crystallogr 12:166–175

Goldman J, Carpenter FH (1974) Zinc binding, circular dichroism, and equilibrium sedimentation studies of insulin (bovine) and several of its derivative. Biochemistry 13:4566–4574

Guinier A, Fournet G (1955) Small-angle scattering of X-rays. Wiley, New York

Hansen JF (1991) The self-association of zinc-free human insulin and insulin analogue B13-glutamine. Biophys Chem 39:107–110

Hansen S, Pedersen J Skov (1991) A comparison of three different methods for analyzing small-angle scattering data. J Appl Crystallogr 24:514–518

Hill CP, Dauter Z, Dodson EJ, Dodson GG, Dunn MF (1991) X-ray structure of an unusual Ca^{2+} site and the roles of Zn^{2+} and Ca^{2+} in the assembly, stability, and storage of the insulin hexamer. Biochemistry 30:917–924

Jacrot B (1976) The study of biological structures by neutron scattering from solutions. Rep Progr Phys 39:911–953

Jeffrey PD, Coates JH (1966a) An equilibrium ultracentrifuge study of self-association of bovine insulin. Biochemistry 5:489–498

Jeffrey PD, Coates JH (1966b) An equilibrium ultracentrifuge study of the effect of ionic strength on the self-association of bovine insulin. Biochemistry 5:3820–3824

Kaarsholm N, Havelund S, Hougaard P (1990) Ionization behavior of native and mutant insulins: pK perturbation of B13-Glu in aggregated species. Arch Biochem Biophys 283:496–502

Kadima W, Roy M, Lee RW-K, Kaarsholm NC, Dunn MF (1992) Studies of the association and conformational properties of metal-free insulin in alkaline sodium chlorine solutions by one- and two-dimensional ^1H NMR. J Biol Chem 267:8963–8970

Kadima W, Øgendal L, Bauer R, Kaarsholm N, Brodersen K, Hansen JF, Porting P (1993) The influence of ionic strength and pH on the aggregation properties of zinc-free insulin studied by static and dynamic laser light scattering. Biopolymers (in press)

Lawson CL, Hanson RJ (1974) Solving least squares problems. Prentice-Hall, Englewood Cliffs, NJ

Lord RS, Gubensek F, Rupley JA (1973) Insulin self-association. Spectrum changes and thermodynamics. Biochemistry 12: 4385–4392

Palmieri R, Lee RW-K, Dunn MF (1988) ^1H Fourier transform NMR studies of insulin. Coordination of Ca^{2+} to Glu(B13) site drives hexamer assembly and induces a conformation change. Biochemistry 27:3387–3397

Pedersen J Skov, Posselt D, Mortensen K (1990) Analytical treatment of the resolution function in small-angle scattering. J Appl Crystallogr 23:321–333

Pekar AH, Frank BH (1972) Conformation of proinsulin. A comparison of insulin an proinsulin self-association at neutral pH. Biochemistry 11:4013–4016

Pocker Y, Biswas (1981) Self-association of insulin and the role of hydrophobic bonding: a thermodynamic model of insulin dimerization. Biochemistry 20:4354–4361

Roy M, Lee RW-K, Kaarsholm NC, Thøgersen H, Brange J, Dunn MF (1990) Sequence-specific ^1H-NMR assignments for the aromatic region of several biologically active, monomeric insulins including native human insulin. Biochim Biophys Acta 1053: 63–73

Sober HA (1970) Handbook of Biochemistry 2. edn. CRC Press

Steiner RF (1951) Reversible association processes of globular proteins: Insulin. Arch Biochem Biophys 39:333–354

Strazza S, Hunter R, Walker E, Darnall DW (1985) The thermodynamics of bovine and porcine insulin determined by concentration difference spectroscopy. Arch Biochem Biophys 238:30–42

Svergun DI, Pedersen J Skov (1993) On the estimation of error propagation in small-angle data treatment. J Appl Crystallogr (in press)