doi:10.1006/jmbi.2000.3784 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **299**, 1289–1302

JMB

Reconstruction of Protein Form with X-ray Solution Scattering and a Genetic Algorithm

Pablo Chacón[1], J. Fernando Díaz[1], Federico Morán[2] and José M. Andreu[1]*

[1]*Centro de Investigaciones Biológicas, CSIC, Velázquez 144, 28006 Madrid, Spain*

[2]*Departamento de Bioquímica y Biología Molecular I Universidad Complutense 28040 Madrid, Spain*

**Corresponding author*

We have reconstructed, from experimental ~2 nm resolution X-ray solution scattering profiles, the corresponding shapes and sizes of myoglobin, troponin C, spermadhesin PSP-I/PSP-II, chymotrypsinogen A, superoxide dismutase, ovalbumin, tubulin, nitrite reductase, catalase, the structural change of troponin C upon dissociation of the two high affinity Ca^{2+}, and the solution model structure of a tandem pair of fibronectin type III cytoplasmic domains of integrin α6β4 before determination of its crystal structure. To this purpose we have designed a new genetic algorithm which gradually explores a discrete search space and evolves convergent models made of several hundred beads (down to 0.3 nm radius) best fitting the scattering profile upon Debye calculation, without geometrical constraints or penalty for loose beads. This is a procedure of effective numerical transformation of the one-dimensional scattering profiles into three-dimensional model structures. The number of beads in models is correlated with the protein molecular mass (with one exception). The shape and approximate dimensions of each protein have been retrieved by a set of ten solution models, essentially superimposable with the available crystal structures.

Keywords: protein shape determination; X-ray solution scattering; genetic algorithms; bead modelling

Introduction

Determination of the low and mid resolution structures of proteins, macromolecular complexes and assemblies, is central in understanding their functional mechanisms, and bridging the resolution gap between atomic resolution structures and biomolecular organisation, interactions and function (Baker & Johnson, 1996). The most powerful approaches include atomic force microscopy, cryo-electron microscopy and image analysis, complemented by hydrodynamic analysis, spectroscopic methods such as fluorescence resonance energy transfer, and neutron and X-ray solution scattering (SAXS). SAXS can give time-resolved structural information to nanometre resolution in experiments actually performed under synchrotron X-ray beams, for example during protein folding (Kataoka *et al*., 1997) and during taxol-induced microtubule assembly and structural changes (Díaz *et al*., 1996, 1998). However, this structural information is hardly transformable into defined 3D model structures in the absence of other information. The scattering profile of any known structure can be calculated to the desired resolution, since the scattering amplitude is the Fourier transform of the electron density of the object. However, in the isotropic scattering profiles which can be measured, the phase information is lost and the diffracted intensities are averaged over all object orientations in solution. The inverse Fourier transform of the intensity profile gives simply the 1D pair distance distribution function (a radial Patterson function; Cantor & Schimmel, 1980), instead of a 3D structure. This is known as the inverse scattering problem, which has obviously no analytical solution, but it involves a search of necessarily degenerate numerical solutions. Uncertainty is avoided by employing preexisting structures or models (see Introduction in Chacón *et al*., 1998; Boehm *et al*., 1999; Krueger *et al*., 1999), or can be satisfactorily reduced in very favourable cases with extensive directed modelling of feature-rich scatter-

Present address: P. Chacón, Centro de Astrobiología, INTA-CSIC, 28830 Torrejón de Ardoz, Spain.

E-mail address of the corresponding author: j.m.andreu@cib.csic.es

0022-2836/00/051289–14 $35.00/0

ing profiles, such as in the case of tubulin double rings (Díaz *et al*., 1994). The question arises of whether it is at all possible to generally extract a reliable 3D model structure from the 1D solution scattering profile. An elegant *ab initio* method consists of the restoration of a low resolution structural envelope, based on expanding the shape function in terms of spherical harmonics described by a set of parameters (Sturhman, 1970; Svergun & Sturhman, 1991; Svergun *et al*., 1996; Grossmann & Hasnain, 1997). A method employing icosahedral harmonics has been applied to viral structures (Zheng *et al*., 1995). The applicability of these methods to complex shapes and their resolution without symmetry constraints are rather limited. On the other hand, bead modelling can be applied to any shape. A given structure can be approximated at any resolution by a set of spheres of small enough diameter, and its solution scattering profile can be calculated with efficient implementations (Pantos & Bordas, 1994; Pantos *et al*., 1996) of the Debye formula (Debye, 1915). However, calculating the Debye formula for all possible binary mass distributions within a search space consisting of N beads (2^N configurations), for more than a few ten beads, is a formidable problem beyond present computing power. Chacón *et al*. (1998) have instead devised an *ab initio* SAXS modelling approach with a genetic algorithm that efficiently searches the configurational space and evolves the best-fit bead models. It is interesting that convergent correct models made of several hundred beads were retrieved when the method was given scattering profiles to 1.7 nm resolution calculated from a set of protein crystal structures of different shapes, and the experimental SAXS profile of lysozyme to 3 nm resolution (Chacón *et al*., 1998). The feasibility of the *ab initio* bead (dummy atom) modelling approach to restore low resolution structure from solution scattering has been confirmed using simulated annealing, and extended to multiphasic ribosomal particles by Svergun (1999).

In this work we have determined, employing X-ray solution scattering to a nominal ~2 nm resolution and an improved genetic algorithm, convergent model structures of myoglobin, troponin C (with 0 and 2 Ca^{2+} bound), spermadhesin PSP-I/PSP-II, chymotrypsinogen A, superoxide dismutase, ovalbumin, tubulin, nitrite reductase, catalase and a tandem pair of fibronectin type III cytoplasmic domains of integrin α6β4. These solution models are compared with the crystal structures.

Results and Discussion

Ab initio model fitting protein solution scattering profiles with a genetic algorithm

The X-ray solution scattering profiles of eleven homogeneous proteins of quite different sizes and shapes were acquired to resolutions (Bragg spacings) comprised between 1.7 and 3.1 nm. These included SAXS profiles from proteins with known crystal structures (Figure 1; Myo, TnC-2Ca^{2+}, PSP, ChyA, Sod, Ova, Tub, NiR and Cat) and profiles of proteins of unknown crystal structures at the time of SAXS analysis (Figure 1; TnC-0Ca^{2+}, β4-int and C-LytA). A new method of bead modelling with a genetic algorithm was devised (see Methods). This non-parametric method was employed to search the configurational space and find models made of several hundred small enough spheres (beyond the

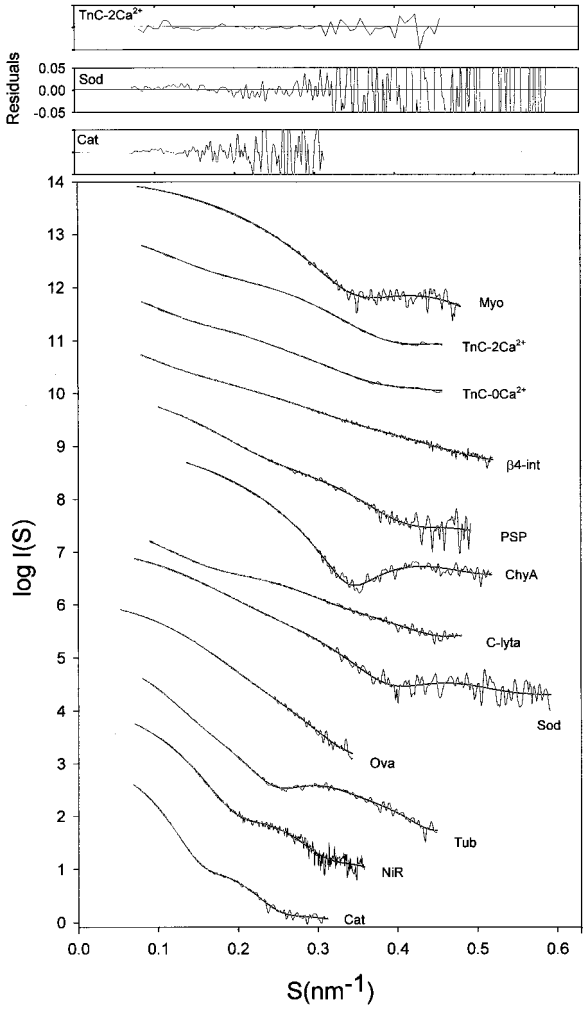

Figure 1. Modelling analysis of the X-ray solution scattering profiles of myoglobin (Myo), troponin C-2Ca^{2+} and troponin C (no Ca^{2+}) (TnC), a tandem pair of fibronectin type III cytoplasmic domains of integrin β4 (β4-int), spermadhesin PSP-I/PSP-II (PSP), chymotrypsynogen A (ChyA), pneumococcal autolysin domain C-LytA (C-LytA), superoxide dismutase (Sod), ovalbumin (Ova), tubulin (Tub), nitrite reductase (Nir) and catalase (Cat). The intensity profiles were normalized dividing by $I(0)$ and displaced one log unit in the ordinate axis for comparison (see Methods for proteins and data acquisition). Experimental data (grey traces with noise) and the scattering profiles of the best bead models obtained with a genetic algorithm search (continuous curves). The upper frames contain representative residuals.

resolution of the data) which best fit each scattering profile. Each model scattering profile obtained from ten runs of the algorithm reproduced within experimental noise the corresponding measurement (Figure 1). The R_g values of the ten final models converged within 0.2(±0.1)%. They coincided with the experimental value. The number of beads of the ten models had a larger variability than the R_g values, reaching 7.6(±2.9)% at maximal modelling resolution (Table 1).

Size of protein solution models

The number of beads (N) in the models obtained from the different SAXS profiles was linearly related with the molecular mass of the corresponding protein sequences, at every bead size (Figure 2). This indicates that the search algorithm is probing the molecular scattering volume, irrespective of molecular shape. Using calculated hydrated molecular mass values or molecular volumes computed from the crystal structures instead of the sequence M_r values did not further improve the correlations. The molecular mass of a given protein may be estimated from the inverse correlation, even with the small set of standards available ($M_r = a + b\ N \pm \sigma$; see Figure 2). This empirical method of measurement of molecular mass with SAXS uses the relative scattering profiles normalized at $S = 0$, thus it has the advantage of not requiring measurements of the absolute protein scattering intensities. However, it has limited accuracy (see M_r values in Table 1), which may improve with better buffer substraction statistics, longer camera lengths and extrapolation to zero protein concentration. The limiting error in M_r measurement inherent to this bead modelling method is about 1000, estimated by modelling synthetic scattering profiles calculated from five protein structures with M_r comprised between 15,000 and 48,000 (Chacón *et al*., 1998).

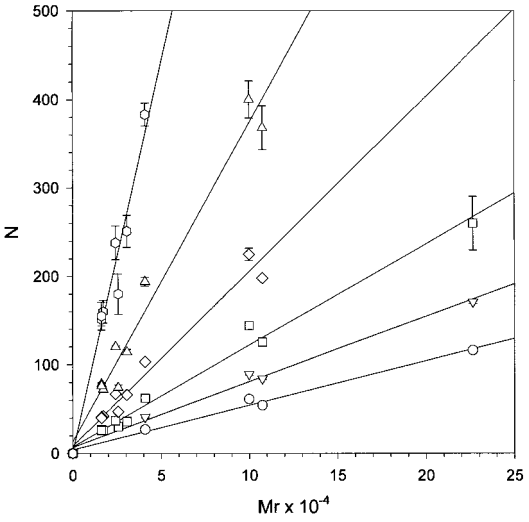

Figure 2. Size of solution scattering models. The number of beads constituting the models is plotted against the molecular mass of each protein sequence. This is made, clockwise, with bead radius 0.3 nm (correl. coeff. $r^2 = 0.962$), 0.4 nm ($r^2 = 0.971$), 0.5 nm ($r^2 = 0.969$), 0.6 nm ($r2 = 0.987$), 0.7 nm ($r2 = 0.991$) and 0.8 nm ($r^2 = 0.988$). The inverse correlations allow estimation of the molecular mass of a given protein from the number of beads in its models, $M_r = a + b\ N \pm \sigma$. The regression parameters are: 0.3 nm, $a = 1704$ $b = 106$ $\sigma = 2700$; 0.4 nm, $a = -1773$ $b = 267$ $\sigma = 6450$; 0.5 nm, $a = -2183$ $b = 487$ $\sigma = 6620$; 0.6 nm, $a = -4886$ $b = 857$ $\sigma = 8020$; 0.7 nm, $a = -8236$ $b = 1340$ $\sigma = 9450$; 0.8 nm $a = -6650$ $b = 1969$ $\sigma = 10700$.

The values of radii of gyration of the solution scattering models (Table 1, column 5) were larger than the corresponding *in vacuo* values calculated from the crystal structures (Table 1, column 4), by an average factor of 1.066(±0.043). The areas under the experimental solution scattering profiles were systematically smaller than those under their *in vacuo* scattering profiles, as expected (see Methods). The average specific volume of all the solution models with 0.3 nm bead radius was 0.81(±0.11) ml g^{-1} (0.89(±0.16) with 0.4 nm bead radius, 0.97(±0.13) with 0.5 nm radius, and 1.01(±0.14) ml g^{-1} with 0.6 nm radius). The present bead modelling algorithm underestimates molecular specific volumes by around 20% (see Methods). Nevertheless, these specific volume values in solution are larger than those obtained (with the same bead sizes) from a set of *in vacuo* synthetic scattering profiles, by a factor of ~1.35. If a similar estimation bias were assumed in both cases (in spite of the different sets of proteins employed and slight differences in S_{max}), this volume factor might be considered an overall average contribution of protein hydration and dynamics to solution scattering, which may be investigated by more extensive bead modelling of individual solution and *in vacuo* scattering profiles.

Shape of protein solution models in comparison with crystal and quaternary structures

Model analysis of the solution scattering profiles (Figure 1) rendered connected bead models which were convergent (plus or minus several surface beads) in ten different runs of the genetic algorithm, for all the profiles analysed. For nine proteins with known crystal structures, the mass distributions of the models and crystal structures were significatively correlated (Table 1) and the models were compatible with the structures, reproducing their overall dimensions and shape. In the following, several representative models are displayed, in order of decreasing size and complexity (Figure 3). In each case the model best fitting the scattering profile is shown (blue surface), superimposed with an envelope of three representative models (red

Table 1. Characteristics of twelve solution scattering protein models

Protein	M_r (sequence)	Structure used for comparison	R_g (in crystal) (nm)	R_g (exp value) (±0.04 nm)	R_g (models) (nm)[a]	Number of beads in models[a] (0.3 nm radius)	M_r (models)[a]	Correlation coefficient of model(s) with PDB[b]
Myoglobin	17,199	1mbn (Nobbs *et al.*, 1966)	1.53	1.62[c]	1.58	160(±13)	18,700(±2150)	0.864 (0.829)
Troponin C-2 Ca^{2+}	18,243	1top (Satyshur *et al.*, 1994)	2.27	2.39[d]	2.41	157(±7)	17,750(±2000)	0.863 (0.772)
Troponin C-0 Ca^{2+}	18,243	-	-	2.78[d]	2.79	155(±16)	18,150(±2250)	-
β4-integrin module	22,340	1qg3 (de Pereda *et al.*, 1999)	2.53	2.90	2.89	155(±10)	18,150(±2050)	0.847 (0.662)
Spermadhesin PSPI/PSPII	24,166	1spp (Varela *et al.*, 1997)	1.85	2.09	2.08	238(±19)	27,000(±2300)	0.819 (0.799)
Chymotrysinogen A	25,856	2cga (Wang *et al.*, 1985)	1.66	1.67	1.66	180(±23)	20,900(±2600)	0.943 (0.934)
C-LytA autolysin	32,000 (dimer)	-	-	2.60	2.56	121(±12)	14,550(±2100)	-
Superoxide dismutase	30,549 (dimer)	1xso (Bordo *et al.*, 1994)	2.00	2.05	2.07	251(±18)	28,350(±2350)	0.936 (0.780)
Ovalbumin	40,866	1ova (Stein *et al.*, 1991)	2.18	2.38	2.38	383(±13)	42,350(±2150)	0.882 (0.855)
αβ-tubulin	99,891	1tub (Nogales *et al.*, 1998)	2.90	3.17[d]	3.19	400(±21) (0.4 nm)	105,200(±6050)	0.829 (0.891)
Nitrito-reductase	107,373 (trimer)	2nrd (Godden *et al.*, 1991)	2.75	2.83[d]	2.83	368(±25) (0.4 nm)	96,600(±6550)	0.882 (0.843)
Catalase	226,625 (tetramer)	7cat (Fita & Rossman,1985)	3.60	3.68	3.70	260(±30) (0.6 nm)	217,900(±18,650)	0.892 (0.919)

[a] The values given are the average and standard error from ten runs of the modelling algorithm.

[b] This real space correlation coefficient is a minimal estimate. The correlation of three solution models manually superimposed with the crystal structure was calculated by placing them into a larger 0.1 nm mesh grid (the values in parenthesis correspond to the correlation of the best fitting model). A point in the grid is labelled with the class ''PDB'', ''model'', or ''common'' if the center of an atom and/or bead is within 0.23 nm (the van der Waals radius of a methyl group, 0.2 nm, divided by cos30° to avoid packing effects). Common: if there are atoms of the crystal structure and model bead(s) within the point region; model: if there are exclusively bead(s); PDB: if there are exclusively atoms of the crystal structure. The correlation coefficient employed was based on those used for digital images (Chapra & Canale, 1998) with the density equal to one or zero, $r = N_{common} / (N_{common}^2 + N_{PDB}^2 + N_{model}^2)^{1/2}$, where N is the number of points in each class.

[c] This R_g value is 0.13 nm smaller than a reference value (Kataoka *et al.*, 1995).

[d] These values are coincident with previously reported ones (Fujisawa *et al.*, 1989; Andreu *et al.*, 1989; Grossmann *et al.*, 1993).

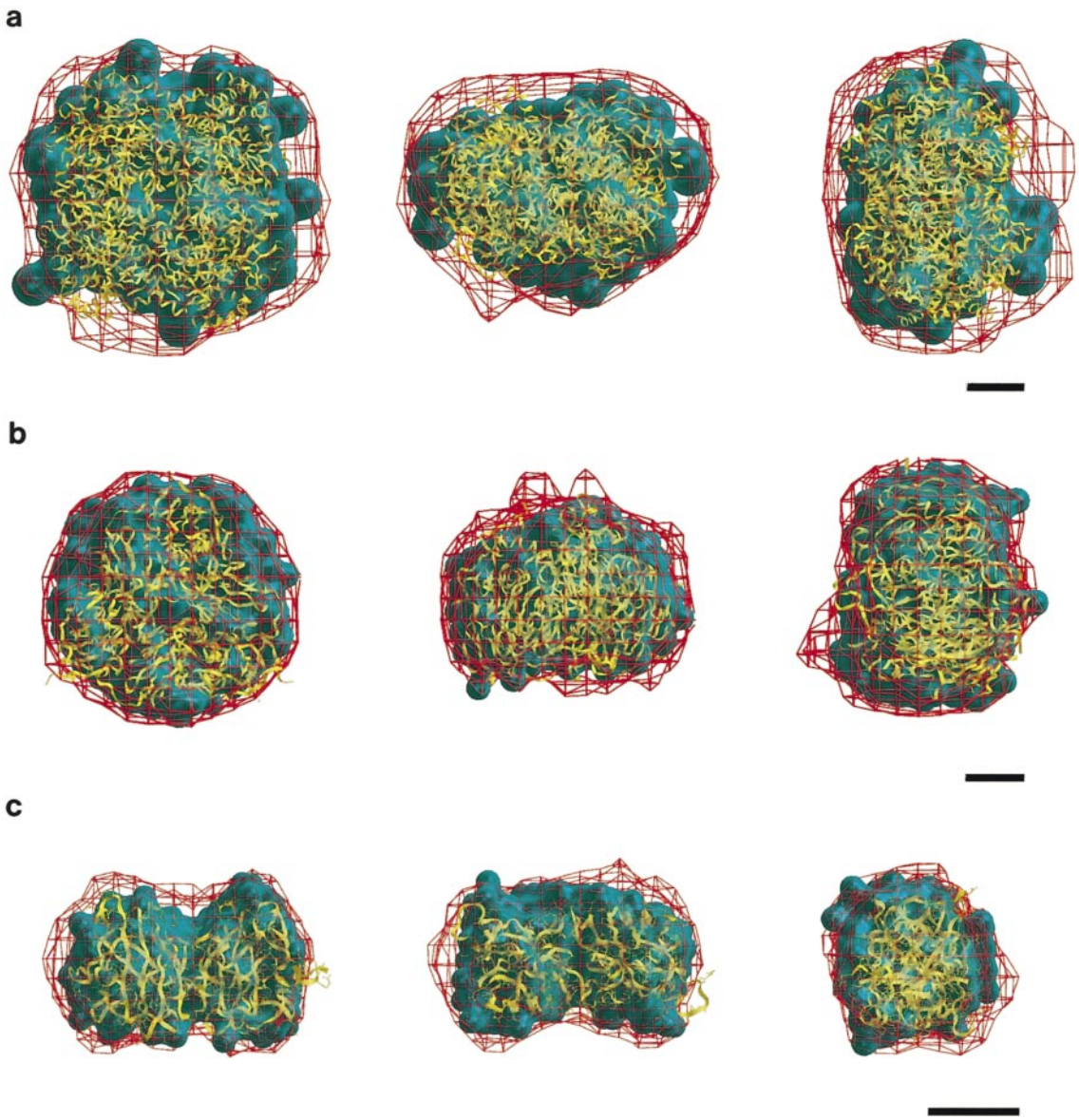

Figure 3. Solution scattering model structures of: (a) catalase (with bead radius 0.6 nm); (b) nitrite reductase (bead radius 0.4 nm); and (c) superoxide dismutase (bead radius 0.3 nm) (bars represent 2 nm). The model best fitting the scattering profile (blue surface) and an envelope of the best, medium and worst fitting model (red wire mesh) are superimposed onto the crystal structures (in yellow ribbon diagrams of the peptide chain, including side-chains in (c). Three orthogonal views are presented in each row. The models were manually aligned with InsightII 95 (MSI). The envelope representations were created with the programs pdblur and volcube, part of the Situs docking package (Wriggers *et al*., 1999), version 1.3. Bead models were convoluted with a Gaussian kernel whose half-maximal radius corresponded to the bead radius. Subsequently, isocontours of the resulting 3D density maps were computed at a threshold identical to the half-maximal density of the Gaussian kernel. This approach is advantageous compared to a solid rendering of the beads, since the resulting isocontours outline the models at the proper bead radius, thereby providing a reduced representation of their shape. The isocontours were visualized as a wire mesh with VMD (Humphrey *et al*., 1996). Molecular surfaces and the graphics were made with Grasp (Nichols *et al*., 1991) and Raster3D (Merrit & Bacon, 1997).

wire mesh: best, medium and worst fits; note that the difference in fitness value was very small) and the crystal structure peptide chain tracing (yellow ribbon diagrams).

The lower resolution SAXS models were from catalase, the largest protein employed (M_r 226,600; Table 1). The 0.6 nm radius bead models (Figure 3(a)) approach the dimensions and overall 10 nm × 10 nm × 8 nm parallelepiped shape of the catalase tetramer (Maté *et al*., 1999) with not well defined contours. Next in size was nitrite reductase (NiR) from *Achromobacter faecalis*. The 0.4 nm bead

radius models (Figure 3(b)) are fully compatible with the trimer crystal structure of NiR from *Achromobacter cycloclastes* (Godden *et al*., 1991). Although the models did not have an evident 3-fold symmetry, other dispositions of crystallographic monomers did not dock into them. The M_r of NiR models was 96,600(±6600) (theor. 107,400 for a trimer). The bead model analysis unequivocally identifies the quaternary structure of *A. faecalis* NiR as a trimer, in agreement with spherical harmonics modelling of the same SAXS data (Grossmann & Hasnain, 1997; Grossmann *et al*., 1993).

The SAXS models of tubulin were compatible with the electron crystallography model structure of the αβ-dimer in Zn^{2+}-induced polymers, which consists of two homologous subunits related by a translation (Nogales *et al*., 1998), with a variable mass excess in one projection (not shown). Possible differences between the solution and polymer crystal structures of tubulin are preliminary supported by a small shift of the scattering shoulder at ~0.31 nm (Figure 1). The relatively elongated shapes and dimensions of superoxide dismutase (Sod) (Figure 3(c)), ovalbumin and spermadhesin PSPI/PSPII (not shown) were satisfactorily retrieved from their SAXS profiles. The crystal structure of Sod is a dimer of two identical subunits related by a non-crystallographic axis (Bordo *et al*., 1994). The carefully acquired SAXS profile of Sod (Figure 1; G. Grossmann, unpublished observations) extends to a nominal resolution of 1.7 nm. The models obtained essentially reproduce the dimer crystal structure at lower resolution (Figure 3(c)). There is a weak trend of one of the monomers in the best model being larger than the other, but this is lost in the models envelope. The experimental and calculated SAXS profiles of Sod were compared after model analysis and were very similar.

The models of chymotrypsinogen A satisfactorily superimposed with the globular crystal structure (not shown). Lower resolution models of myoglobin, made of 26 beads of 0.6 nm radius, reproduced the ellipsoidal shape of its crystal structure (it is interesting that some of these simple models were related by symmetry operations). Myoglobin models made of 0.3 nm beads had correct mass estimates, however, they had nearly spherical shape (data not shown; inspection of the myoglobin scattering profile employed evidenced a relatively low intensity around 0.4 nm^{-1} in comparison with the calculated profile). Note that optimizing the alignment of lower resolution (SAXS and electron microscopy) models with high resolution (X-ray diffraction and NMR) structures is not trivial. A neural network method for modelling crystal structures into electron microscopy image reconstructions (Wriggers *et al*., 1998) can be applied to this problem, as well as possibly docking (Sternberg *et al*., 1998) the surface envelopes of the SAXS models and crystal structures, although ambiguous superpositions may be obtained for globular proteins.

Ab initio modelling a Ca^{2+}-induced protein structural change with solution scattering

The crystal structure of the muscle Ca^{2+}-binding protein troponin C (troponin-2Ca^{2+}, Herzberg & James, 1985; Satyshur *et al*., 1994) has a dumbbell shape, consisting of two globular domains connected by a helical linker. Modelling the low noise scattering profile of the troponin-2Ca^{2+} complex in the absence of Mg^{2+} (Figure 1; Ca_2Ca_0 Tn-C, Fujisawa *et al*., 1989, 1990) rendered bead models all of which described the structural organization and approximate dimensions of the crystal structure (Figure 4(a)). The models retrieved from the scattering profile of Ca^{2+}-less troponin (Tn-C, Fujisawa *et al*., 1989, 1990) (note the difference in profile shape, Figure 1) consistently exhibited a large expansion, encompassing the opening of one of the model domains while the other showed much less change (Figure 4(b)). The distance between the centroids of the domains was the same in the crystal structure and the first model (4.2(±0.2) nm), however, the protein model expanded by ~1 nm upon Ca^{2+} dissociation (Figure 4). The increment in the R_g value was 0.38(±0.01) nm, whereas the estimated molecular mass of both models was coincident with that of troponin (Table 1). We propose that the large structural change clearly imaged by the model analysis (Figure 4) is induced by the dissociation of the two Ca^{2+} from their high affinity structural sites at the C-terminal domain of intact troponin C. A related expansion was proposed by Fujisawa (1989, 1990), modelling each domain with an ellipsoid and employing tryptic fragments of the protein. Fluorescence anisotropy of single tryptophan mutants of troponin C and sedimentation velocity measurements suggest as well an elongated shape of the metal free form (Moncrieffe *et al*., 1999), similarly to the rotational dynamics of Ca^{2+}-free Xenopus calmodulin studied by ^{15}N-NMR relaxation measurements, which was interpreted in terms of a well folded relatively flexible dumbbell model (Tjandra *et al*., 1995). This is different from the structural change responsible for muscle contraction, triggered by the binding of two Ca^{2+} to the lower affinity regulatory sites of the N-domain (troponin-4Ca^{2+}; Herzberg *et al*., 1996; Slupsky & Sykes, 1995), which is noticed faintly in the scattering profile of troponin at this resolution (not shown). Solution scattering has been extensively employed to characterize protein structural changes by directed modelling, including troponin C-troponin I complexes (Olah & Trewhella, 1994; Stone *et al*., 1998). The present analysis constitutes to our knowledge a first *ab initio* model determination of a solution structural change of a non-globular protein employing SAXS. We cannot at present ascertain whether the model structures

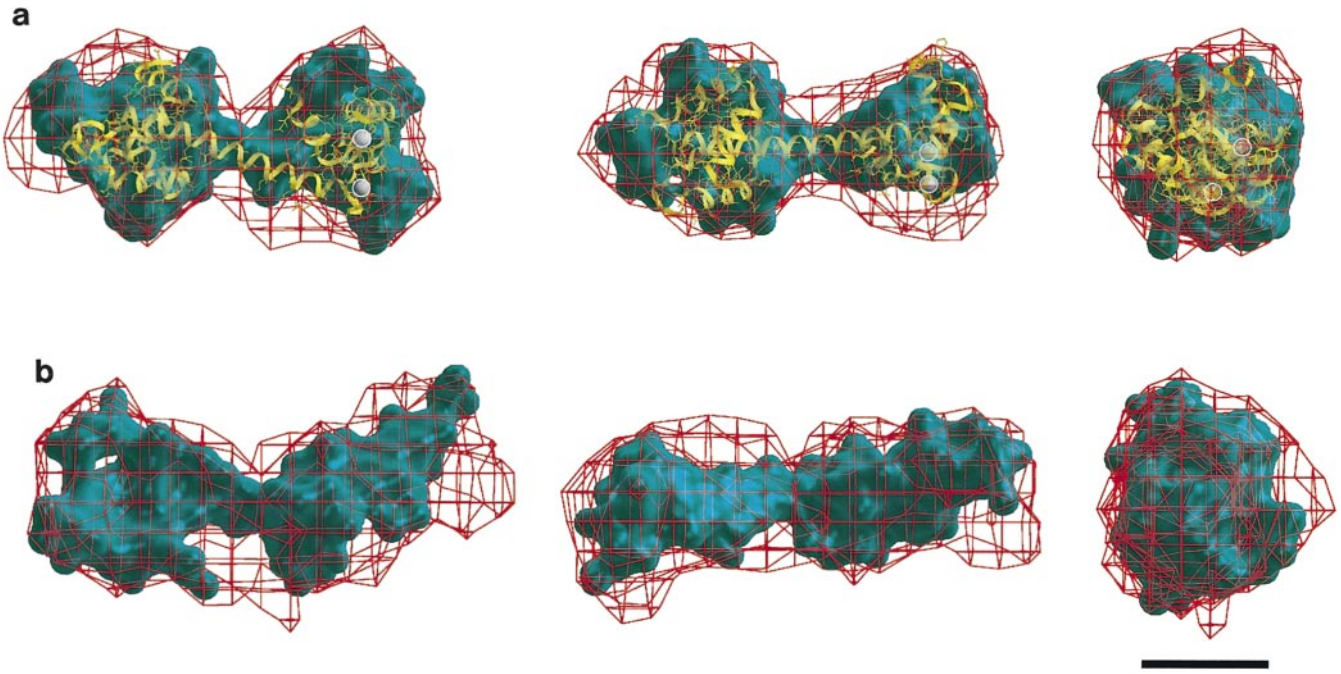

Figure 4. Solution scattering model structures of (a) troponin C with 2 Ca^{2+} ions bound (single model (blue) and three-model wire envelope as in Figure 3) in comparison with the crystal structure (yellow ribbon and side-chains, metal ions (grey circles)); and (b) troponin from which Ca^{2+} has dissociated. Note that the irregular surface of the individual models could have been made smoother by arbitrarily subtracting a small constant (0.005) from the normalized intensity values (not shown). Bar represents 2 nm.

(Figure 4) represent unique species or mid-resolution conformational averages.

Finding unknown shapes and dimensions from SAXS profiles to 2 nm resolution

The above analysis of troponin SAXS can also be considered an exercise in the prediction of the form of Ca^{2+}-less troponin. The corresponding high resolution structure was not available for comparison. Therefore, the scattering profiles of other proteins whose crystal structures were being pursued were acquired and model-analyzed. The SAXS models of C-lytA had a V shape (made of ~5 nm arms forming a 100° angle, not shown) compatible with the organization of the C-lytA dimer (M_r 32,000) (Medrano *et al*., 1996; Usobiaga *et al*., 1996), however, the model surfaces were quite irregular and their molecular masses (14,550(±2100)) corresponded to a monomer. This case was preliminary rated as a size failure, since the crystal structure of C-lytA is not yet available.

Integrin α6β4 is an essential component of hemidesmosomes, implied in cell adhesion and also in invasiveness of carcinoma cells (Suzuki & Naito, 1990; O'Connor *et al*., 1998). The relatively low noise SAXS profile of the β4 integrin subunit cytoplasmic fragment (Figure 1) yielded convergent model structures with a 10.5 nm elongated shape, made of two related domains ca. 3.9 nm × 2.4 nm thick each, forming an angle of ~140° (Figure 5). The molecular mass estimate of the models was 18,200(±2100) (theor. 22,340). After these SAXS models were obtained (P. Chacón, 1999), the X-ray crystal structure of this tandem pair of fibronectin type III cytoplasmic domains of the β4 integrin subunit was presented (de Pereda *et al*., 1999; i.e. we had made a blind model analysis, since we did not know the crystal structure from the authors until publication and coordinate release). The superposition of the *ab initio* SAXS models and crystal structure shows a remarkable compatibility at 2 nm nominal solution scattering resolution (Figure 5). The molecule is 8.9 nm long and the domains are practically extended (170° angle) in the crystal structure. The network of hydrogen bonds of the linker residues and van der Waals contacts at the interface between domains in the structure would suggest a similar conformation of the two-domains in solution. However, in the SAXS models the domains are about 30° closer and ~1 nm longer. It is difficult to ascertain whether these differences are the result of model inaccuracy or of limited flexibility in solution.

Uniqueness *versus* degeneracy of shape determination. On the information content of solution scattering. Prospects

How can exploration of the inverse scattering problem (which is analytically unsolvable) by

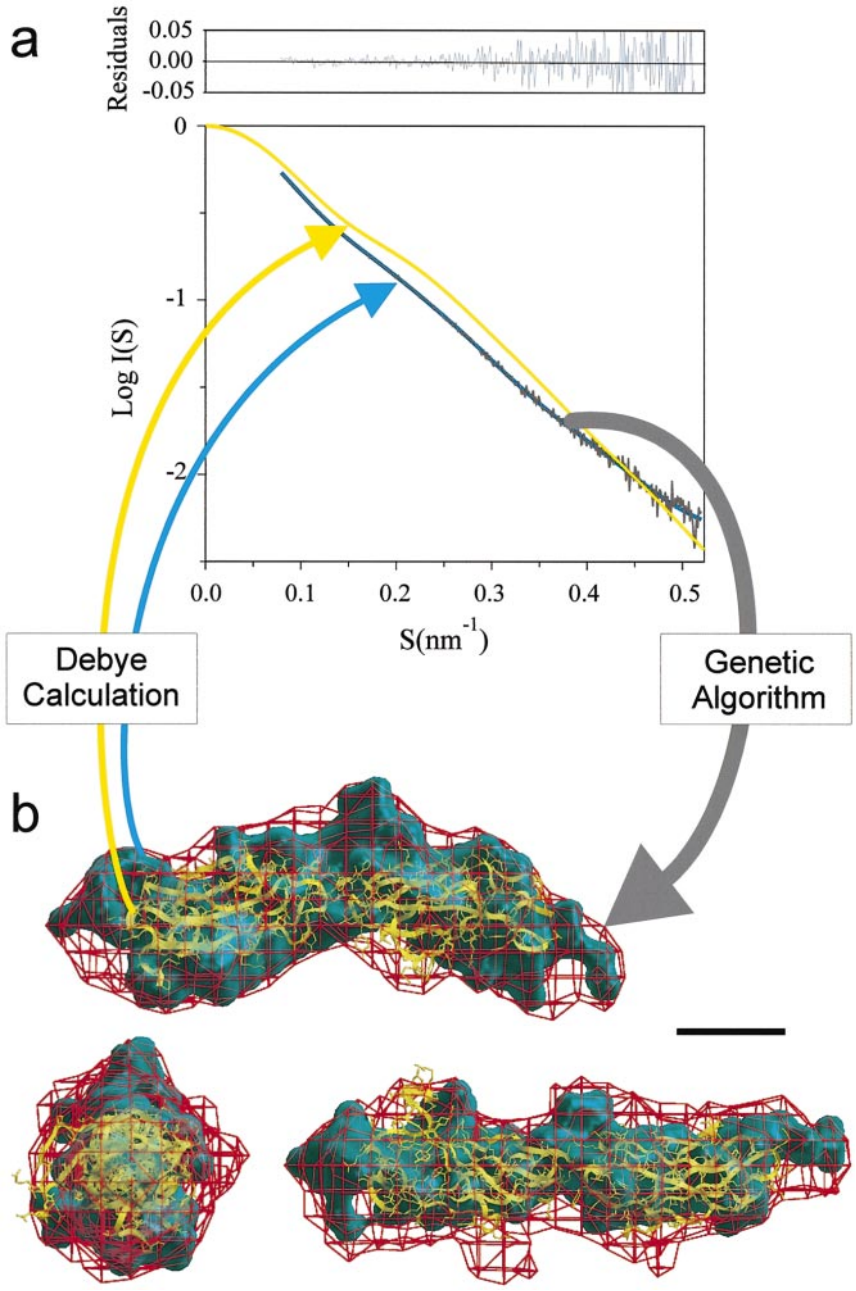

Figure 5. (a) Comparison of the scattering profiles of the cytoplasmic domains of integrin β4; experimental solution profile (grey); model profiles (blue); *in vacuo* profile calculated from the crystal structure after modelling was completed (yellow). (b) Comparison of the crystal structure (yellow ribbon and side-chain representation) with the solution scattering model structures (a single model is shown in blue and a three-model envelope in red wire mesh, as in Figure 3). Three orthogonal views are presented (bar represents 2 nm). The arrows going from (b) to (a) indicate the scattering profiles which have been calculated from the model structures. The arrow from (a) to (b) indicates that a genetic algorithm has been employed to numerically retrieve the model structures from the experimental scattering profile of this protein.

numerical modelling of a linear profile render meaningful three dimensional information? Here, a dozen protein solution scattering profiles have been model-analysed, conveying an amazing structural information content. A first issue to be addressed is the existence of degenerate solutions, i.e. any models having completely different shapes but identical distances between its elements, or the same pair distance distribution function, will obviously render the same scattering profile. As an example, when modelling the calculated scattering profile of six beads in an hexagon, the genetic algorithm found in addition to the hexagon two tetrahedral arrangements having equivalent pairs of distances. In addition, more complex models related by symmetry operations will be indistinguishable by their scattering profiles. This includes models related by mirror reflection or by inversion through a point. The genetic algorithm bead modelling method is free from any geometric constraints. On the other hand, imposing model symmetries may generate divergent solutions with a higher probability. Quite different model envelopes can be generated having nearly identical profiles (Svergun *et al.*, 1996), only distinguishable beyond S_{max}. Suppose that an unrealistic bead model appears with symmetric features whose contributions cancel each other in the model scattering profile. Such a bead model would hardly be preserved in evolution, since its characteristics

will not survive a recombination event in which the favoured compensating features are separated. A similar situation was observed when modelling some scattering profiles with the larger 0.6 nm radius beads, which rendered spermadhesin enantiomorphic models and other symmetry related myoglobin models (data not shown). However, during refinement with smaller bead sizes the degenerate models were replaced by convergent models. This may be due to a decreasing probability of randomly generating symmetric bead arrangements with larger numbers of beads. A minimal bead size commensurate with the resolution of the scattering profile was employed. A 0.1 nm smaller bead radius resulted in overfitting, detected by a larger variability in the number of beads in the models.

In this work, ten runs with ten different protein profiles (100 algorithm runs) gave model shapes compatible with the corresponding crystal structures. In one case (myoglobin) the model axial ratio was different than in the crystal structure, and molecular mass was subestimated in another (C-lytA) possibly due to incorrect data or modelling failure. Therefore correct numerical solutions were found in a minimum of 80 or 90 % of the cases analysed. The quality of the experimental SAXS profiles is determinant to the results, accuracy in buffer substraction and profile normalization allowing the distinction among similar sizes and shapes.

The number of independent parameters which describes the Fourier transform p(r) of a scattering profile I(S) can be classically estimated as $n_{max} \leqslant S_{max}/S_{min}$ (Glatter & Kratky, 1982), which is in the order of 10 for a typical SAXS profile measurement. However, the actual number of degrees of freedom in fitting a data set depends on the signal to noise ratio and may be considerably larger than the number of Shannon channels (Svergun, 1999). Note that even this lower estimate is one order of magnitude larger than that in other solution methods frequently employed to investigate protein shape. For instance, translational and rotational diffusion measurements with sedimentation velocity and fluorescence anisotropy respectively, yield single form factors (frictional coefficients) (Waxman *et al*, 1993).

Ab initio bead modelling of solution scattering with genetic algorithm search is essentially a non-parametric method. Given the nature of the problem, it can not be claimed that results obtained without additional information are the unique family of solutions fitting the scattering profile, however, other possible solutions appear of low probability, since they have not been found in a large number of cases. The procedure of shape reconstruction from SAXS can be described as a numerical transform of the one-dimensional scattering profile into a discrete three-dimensional model structure (see Figure 5). In this sense, this approach might be useful for any similar numerical problems other than the inverse scattering problem, such as direct structural determination with X-ray diffraction (Miller *et al*.,1996; Harris *et al*.,1998). In addition, the molecular envelope of a protein from solution scattering can also be employed in *ab initio* phasing X-ray diffraction (Hao *et al*., 1999).

Shape and size *ab initio* determination with solution scattering emerges as a powerful mid-resolution approach, complementary to molecular microscopy and image analysis. It is applicable to monodisperse randomly oriented species, from relatively small to large proteins. Subjects of further study should include the use of crystal structure information to automatically determine functional structural changes from solution scattering, time resolution and the analysis of ordered macromolecular assemblies.

Methods

Protein SAXS data acquisition

Synchrotron X-ray solution scattering profiles of proteins and their buffers, were measured employing a scanning cell at 4 °C at station 2.1 of the Daresbury Laboratory SRS (UK) as described (Andreu *et al*., 1989, 1992, 1994; Díaz *et al*.,1996), except where noted. Data were collected for several cell scans and the time frames were inspected before averaging for any signs of radiation damage or aggregation effects. A 2 m camera length was routinely employed, covering a S range $0.05 \leqslant S \leqslant 0.5$ nm^{-1}, defining the scattering vector modulus as $S = 2(\sin\theta)/\lambda$, where 2θ is the scattering angle and λ the X-ray wavelength. Myoglobin (from horse muscle; Sigma cat. no. M-0630), chymotrypsinogen A, ovalbumin (Pharmacia cat. no. 17-0442-01), catalase (Pharmacia cat no. 17-00441-01), and spermadhesin PSPI/PSPSII (Varela *et al*., 1997), were dialysed into 20 mM sodium phosphate, 0.1 M NaCl buffer (pH 7.0), and centrifuged at 80,000 *g* five minutes in a TL100.3 rotor (Beckman) or filtered through 0.45 μm filters (Millipore) before use. Their concentrations were measured spectrophotometrically and ranged between 4 and 14 mg/ml. Aldolase and RNAse scattering profiles acquired did not produce a correct pair distribution function (see below) and so were not employed. The expressed module of the β4 integrin subunit (de Pereda at al., 1999) was measured in 20 mM Tris (pH 7.5), 150 mM NaCl. C-lytA (Usobiaga *et al*., 1996) was dialysed into 20 mM phosphate, 140 mM choline (pH 8.0). Calf brain tubulin (4-10 mg/ml) was equilibrated in 10 mM sodium phosphate, 1 mM GTP (pH 7.0) and measurements (Andreu *et al*., 1994; Díaz *et al*., 1994) were extended up to 0.6 nm^{-1}. The homogeneous quaternary structure and absence of aggregates in these proteins was checked by sedimentation equilibrium under similar solution conditions in a Beckman XLA analytical ultracentrifuge.

The SAXS profiles of nitrite reductase (Grossmann & Hasnain, 1997) and bovine superoxide dismutase were measured by Dr G. Grossmann at the Daresbury Laboratory SRS (UK). The scattering profiles of troponin C were measured by Dr T. Fujisawa at the Photon Factory (Japan) and extrapolated to zero protein concentration (Fujisawa *et al*., 1989, 1990). Troponin C and tubulin SAXS were modeled for this work in a region

($S \leqslant 0.45\ \mathrm{nm}^{-1}$) easily compatible with the criteria applied to the rest of the profiles (see below).

SAXS profile treatment. Solution scattering theory

Scattering profiles were divided by the intensity of the transmitted beam and the detector response, employing OTOKO (Boulin *et al.*, 1986). Buffer was subtracted from each protein sample. The buffer profile was multiplied by a progressively larger factor close to unity (typically 0.98 to 1.01), before many negative values would distort the tail of the resulting profile. Note that the difference between protein and buffer at higher angles tends to have values within the noise of the measurements, and that the rectangular scanning cell used to prevent radiation damage has a slightly more variable path length than smaller fixed cells. Buffer subtraction was refined within 0.5 % (which roughly coincides with the volume fraction occupied by the proteins) according to the criterion that Kratky plots [$I(S)S^2$ *versus* S] should tend to zero at sufficiently high S values, which was confirmed with scattering profiles calculated from the protein crystal structures. Note that the relationship between the forward scattered intensity I(0) and the scattering volume V, is $I(0)/V = 4\pi \int_0^\infty I(S)S^2 dS$. The second term is known as the invariant Q, which is directly related to the mean square fluctuation of electron density irrespective of particular features of the structure (Porod, 1982). The fourth power law ($I(S)S^4 \rightarrow$ constant; Porod, 1982) was not employed to correct the data, since neither scattering profiles calculated from the crystal structures (with or without an hydration layer) nor those of the bead models had a significant region following such a trend within the measurable S range (note that the scattering of the models becomes dominated by the form factor of the spherical bead slightly beyond the end of the experimental profile). In addition, bead modelling should not necessarily be restricted to modelling an envelope unless compactness is imposed. Subtracting 1 % more or less buffer than the optimal from the protein scattering modified insignificantly the scattering intensities at lower angles, the radius of gyration values and the shape and length of the models tested (C-lytA and integrin modules; Results) but resulted in a noticeably smoother or rougher surface and about 25 % more or less beads in the respective individual model. Model structures (Results) were compact (with none to ~1 % internal cavities), except myoglobin and chymotrypsinogen models which had central cavities representing on average 7 % of their total number of beads (the myoglobin crystal structure has a similar cavity, chymotrypsinogen has none). However, subtracting a small arbitrary constant or slightly more buffer from the protein scattering profiles (both procedures gave numerically very similar profiles) produced fully compact myoglobin and chymotrypsinogen models.

The pair distribution function p(r) (the radial Patterson function), $p(r) = (1/\pi) \int_0^\infty I(S) S r \sin(2\pi S r) dS$ (Glatter & Kratky, 1982) was obtained by indirect Fourier transformation of I(S) employing the program GNOM (Semenyuk & Svergun, 1991; http://www.srs.dl.ac.uk/ncd/computing/manual.gnom.html). The p(r) function tends to zero at $r = D_{max}$, the maximal dimension of the particle. Scattering profiles were edited by eliminating any small lower angle portion which prevented p(r) from approaching zero or gave subsidiary maxima at physically meaningless distances, probably as a result of imperfect measurement or sample aggregation. The modelling method is quite sensitive to aggregation, which made loose beads appear. The forward scattering I(0) was obtained from p(r), as $I(0) = 2\int_0^\infty p(r) dr$, instead of using Guinier's extrapolation, and the radius of gyration from $R_g^2 = \int_0^\infty p(r) r^2 dr / 2\int_0^\infty p(r) dr$ (Glatter & Kratky, 1982) with GNOM. Finally, in order to be able to quantify the deviation of model from experimental scattering profiles during the fitting process, they were normalized as $I(S)_r = I(S)/I(0)$, facilitating as well comparisons among different proteins. In terms of relative intensities the scattering volume becomes $V = (4\pi \int_0^\infty I(S)_r S^2 dS)^{-1}$, whose value has the uncertainties derived from the approximations of the forward scattering I(0) and the tail end of I(S) [$S \rightarrow \infty$], beyond the experimentally measurable S range.

A new genetic algorithm for *ab initio* bead modelling of SAXS

Application of a classical genetic algorithm (Chacón *et al.*, 1998) to the set of protein SAXS data acquired generally rendered models compatible with the structures investigated, however, in some cases a tendency to fall into relatively low fitness values was noticed, and some models presented beads unconnected with the main structure. In that approach it was chosen not to introduce a penalty for loose beads in the genetic algorithm method, different to the simulated annealing method (Svergun, 1999). An improved genetic algorithm has been constructed (DALAI_GA2) to model experimental SAXS data, which differs from the previous one in automatization, parallel run strategy, and the use of three-dimensionally constrained genetic operators (pseudo 3D operators), resulting in a better model exploration. These differences in the new algorithm are summarized in the brief description which follows.

Modelling starts with the definition of an initial search space consisting of a hexagonal packing of several hundred beads, containing the maximal dimension of the object (estimated from p(r)). A few incomplete runs are rapidly performed to better define the search volume by trial and error, so that it is not much larger than ten times that of the resulting models. The SAXS profile is truncated at $S_{model} = 1/(4R)$, the reciprocal of four times the bead radius (R), in order to avoid strong interference from the bead form factor into the model scattering. The search process starts with a random population of 1000 different models within the search space, each of them codified by a linear binary sequence. The scattering profile of each model is computed through simplified Debye calculation, in which all bead pair distances are sorted into a distance histogram (Pantos & Bordas, 1994; Pantos *et al.*, 1996). The fitness value of the model is defined as $F = (1/N_p \Sigma_i [\log I_{exp}(S_i) - \log I_{model}(S_i)]^2)^{-1/2}$, where N_p is the number of points of the profile. The population is sorted according to the fitness values. The best 25 % models survive during the next generation and the rest are regenerated from this subset using genetic operators. Several pseudo-3D genetic operators were designed, including model restricted recombination, model-restricted mutation and model rearrangement (bead flipping), which were manually optimized. In each of these cases the first 12 neighbours of each given bead in the hexagonal packing employed were stored in a matrix, identifying them in the linear sequence. For model-restricted recombination, two binary sequences are taken, several positions identified

within each corresponding model and its 3D neighbour positions are randomly chosen and exchanged with a given probability (probability of model recombination 0.75, fraction of beads from each parent 0.6 and 0.4). For model-restricted mutation of a sequence, mutation is restricted to the 3D model and its neighbouring beads (probability of model mutation 0.20, probability of bead mutation within a selected model 0.01). For model rearrangement one or more beads of a model are flipped with empty positions in the neighbourhood (model probability 0.05, bead probability 0.2). These operators favoured the smooth evolution of new models related to the previous ones, and the conservation of 3D model features relevant to the scattering profile. The new genetic algorithm evolved through a spatially connected model search, without penalizing loose models in their fitness parameter.

The process of selection and reproduction is repeated for 2000 generations until there are no changes in the best ten models. After convergence is achieved a mask is automatically generated around the best model, constituting the new search space, with a 0.1 nm smaller bead radius, and the procedure repeated until reaching a 0.3 nm radius. With this mask strategy the modelling resolution is automatically increased by the algorithm. The whole procedure is repeated in ten independent parallel executions, ensuing model diversity (instead of selecting the best model out of ten runs at each resolution step; Chacón *et al.*, 1998).

Tests of the modelling method. Volume. Resolution

The method was tested with a set of five synthetic SAXS profiles calculated from protein crystal structures, with added noise. The models obtained reproduced the shape and dimensions of each problem protein, practically containing the alpha-carbon chain as with the previous algorithm (Chacón *et al.*, 1998). The improvements consisted in slightly better fits at each resolution, more local variability (indicating a better exploration), and practical absence of loose beads. The modelling method underestimated absolute molecular volumes. The volume of each model, V_{mod}, was calculated as the number of beads times the bead volume, divided by the packing factor of the compact hexagonal lattice employed (0.75). The specific volume of a protein model is $V_{mod}N_AM_r^{-1}$, where N_A is Avogadro's number and the molecular mass M_r is calculated from the corresponding sequence. The values of the average specific volume of all models were (in ml g^{-1} units) 0.73(±0.08) with 0.6 nm radius beads and 0.76(±0.05) with 0.5 nm radius, decreasing to 0.65(±0.05) with 0.4 nm radius and 0.60(±0.05) with 0.3 nm radius. This suggested an effect of the progressively smaller surface beads, towards an unknown asymptotic volume value. The average specific volume with 0.3 nm radius beads is ~20% smaller than typical protein partial specific volumes (~0.74 ml g^{-1}). The ratios of model volume to the volume of the corresponding crystal structure (V_{mod}/V_{pdb}) decreased similarly with the bead size, previously giving the impression that they converge towards unity (Chacón *et al.*, 1998). This was a wrong appreciation, since the V_{mod}/V_{pdb} ratios actually reach values smaller than one and depend on the choice of reference volume calculation. One possible explanation for the underestimation of volume may be as follows. The volume V of a scattering particle is inversely related to the area under the normalized scattering profile (see above). The scattering profiles of the bead models faithfully reproduce calculated problem profiles slightly beyond the S_{max} employed in their modelling, but they fail to fit the scattering profile at higher angles. In the models examined, there was a trend to I(S) values larger than those of problem above S_{max}, thus making the estimated volume smaller. Since the bead density is a uniform arbitrary constant, only their volume may be constrained, employing the protein partial specific volume and molecular mass, but then models would fit the shape only and not the molecular size. Other molecular information may be introduced in combination with the search method (such as symmetry, sedimentation coefficient or approximate axial ratio). However, it was preferred to test the method without any such restrictions for a better appreciation of its outcome. Further analysis will be required to determine to which extent the volume bias is due to the interfering form factor of the beads beyond the modelling resolution or to the model search procedure, and whether systematic volume errors may affect other bead or envelope modelling procedures (for examples, see Svergun, 1999; Svergun *et al.*, 1996; Zheng *et al.*, 1995).

The smallest bead radius which can be used to model a given SAXS profile is limited by the number of beads that can be easily computed (typically < 1000 with the present algorithm, which sets a practical resolution limit of ~1/10 the linear dimension of the particle) as well as by the resolution of the data. Although the resolution is routinely quoted as the smallest Bragg spacing, i.e. $1/S_{max}$, the smallest distance of a particle which can be measured by SAXS can be classically estimated as $D_{min} = 1/2\ S_{max}$ (Glatter & Kratky, 1982). That is, for typical values of $S_{max} \approx 0.5$ nm^{-1} (Figure 1) spherical beads of 0.5 nm radius would be small enough. Note that the form factor of a uniform sphere of 1 nm radius strongly interferes at $S = 0.5$ nm^{-1} ($I(S)/I(0) = 0.09$), whereas it shows up much less for spheres of radii 0.5 nm ($I(S)/I(0) = 0.60$), 0.4 nm (0.72), 0.3 nm (0.84) and 0.2 nm (0.92); that is, for a ~10% reduction in intensity in the sphere form factor at S_{max}, the bead size required is $D_{sphere} \approx 1/4\ S_{max}$. In practice, the smallest beads to be employed in the modelling of the experimental scattering profiles ($S_{max} \leqslant 0.6$ nm^{-1}) had 0.3 nm radius. This was determined by modelling the set of synthetic profiles (Chacón *et al.* 1998; $S_{max} = 0.6$ nm^{-1}) in 0.1 nm decreasing bead radius steps, and watching for overfitting, detected by nearly constant model profile fitness values with a sudden rise of variability in the number of beads in the ten models of each profile. This was observed when changing from 0.3 nm bead radius (average variability 4.9(±2.1)%) to 0.2 nm bead radius (10.9(±3.1)%).

The modelling program DALAI_GA2, written in C, took a few minutes to evolve approximate lower resolution models, and about 20 h to complete model refinement steps with 0.6, 0.5, 0.4 and 0.3 nm radius beads in a SGI Indigo workstation with a R4440 processor or in a Pentium II PC. Potential improvements of the algorithm include volume and symmetry constraints for a better shape determination, direct recombination and mutation of bead models specified in the form of cubic matrixes instead of linear sequences, simplification of the search by extracting bead clusters common to several models, as well as parallelization or parallel network execution of a number of runs of the algorithm. The present program compiled in Unix (SGI) and Linux (PC) and practi-

cal examples are available from the authors upon request (see http://akilonia.cib.csic.es/DALAI_GA2).

Acknowledgements

We thank G. Grossmann for the gift of the scattering profiles of Sod and NiR and helpful advice, E. Pantos for discussion, T. Fujisawa for the troponin C scattering profiles, G. Diakun for help at the Daresbury SRS, D. Svergun for GNOM, J.M. de Pereda for providing the β4 integrin module, A. Romero for spermadhesin, M. Menendez for C-lytA, W. Wriggers for help generating the envelope mesh display, and J. Bordas and F. Montero for encouraging approaches to the inverse scattering problem. This work was supported in part by grants from DGES PB95-0116 and CICYT BIO99-0859-C03-02 (JMA), a fellowship from MEC (PC) and the EU Large Installations Program.

References

Andreu, J. M., García de Ancos, J., Starling, D., Hodgkinson, J. L. & Bordas, J. (1989). A synchrotron X-ray scattering characterization of purified tubulin and of its expansion induced by mild detergent binding. *Biochemistry,* **28**, 4036-4040.

Andreu, J. M., Bordas, J., Díaz, J. F., García de Ancos, J., Gil, R., Medrano, F. J., Nogales, E., Pantos, E. & Towns-Andrews, E. (1992). Low resolution structure of microtubules in solution. Synchrotron X-ray scattering and electron microscopy of taxol-induced microtubules assembled from purified tubulin in comparison with glycerol and MAP- induced microtubules. *J. Mol. Biol.* **226**, 169-184.

Andreu, J. M., Díaz, J. F., Gil, R., de Pereda, J. M., García de Lacoba, M., Peyrot, V., Briand, C., Towns-Andrews, E. & Bordas, J. (1994). Solution structure of Taxotere-induced microtubules to 3-nm resolution. The change in protofilament number is linked to the binding of the taxol side-chain. *J. Biol. Chem.* **269**, 31785-31792.

Baker, T. S. & Johnson, J. E. (1996). Low resolution meets high: towards a resolution continuum from cells to atoms. *Curr. Opin. Struct. Biol.* **6**, 585-594.

Boehm, M. K., Woof, J. M., Kerr, M. A. & Perkins, S. J. (1999). The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. *J. Mol. Biol.* **286**, 1421-1447.

Bordo, D., Djinovic, K. & Bolognesi, M. (1994). Conserved patterns in the Cu,Zn superoxide dismutase family. *J. Mol. Biol.* **238**, 366-386.

Boulin, B., Kemptf, R., Koch, M. H. & McLaughlin, S. (1986). Data appraisal, evaluation and display for synchrotron radiation experiments: hardware and software. *Nucl. Instr. Methods,* **A246**, 399-407.

Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry. Part II. Techniques for the Study of Biological Structure and Function*, pp. 817-818, W. H. Freeman, New York.

Chacón, P. (June, 1999). PhD thesis, Universidad. Complutense, Madrid.

Chacón, P., Moran, F., Díaz, J. F., Pantos, E. & Andreu, J. M. (1998). Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J.* **74**, 2760-2774.

Chapra, S. C. & Canale, R. P. (1998). *Numerical Methods for Engineers with Programming and Software Applications*, McGraw-Hill, New York.

Debye, P. (1915). Zerstreung von röntgenstrahlen. *Ann. Phys.* **46**, 809-823.

de Pereda, J. M., Wiche, G. & Liddington, R. C. (1999). Crystal structure of a tandem pair of fibronectin type III domains from the cytoplasmic tail of integrin α6β4. *EMBO J.* **18**, 4087-4095.

Díaz, J. F., Pantos, E., Bordas, J. & Andreu, J. M. (1994). Solution structure of GDP-tubulin double rings to 3 nm resolution and comparison with microtubules. *J. Mol. Biol.* **238**, 214-225.

Díaz, J. F., Andreu, J. M., Diakun, G., Towns-Andrews, E. & Bordas, J. (1996). Structural intermediates in the assembly of taxoid-induced microtubules and GDP-tubulin double rings: time-resolved X-ray scattering. *Biophys J.* **70**, 2408-2420.

Díaz, J. F., Valpuesta, J. M., Chacón, P., Diakun, G. & Andreu, J. M. (1998). Changes in microtubule protofilament number induced by Taxol binding to an easily accessible site. Internal microtubule dynamics. *J. Biol. Chem.* **273**, 33803-33810.

Fita, I. & Rossmann, M. G. (1985). The active center of catalase. *J. Mol. Biol.* **185**, 21-37.

Fujisawa, T., Ueki, T. & Iida, S. (1989). Structural change of troponin C molecule upon Ca^{2+} measured in solution by the X-ray scattering technique. *J. Biochem.* **105**, 377-383.

Fujisawa, T., Ueki, T. & Iida, S. (1990). Structural change of troponin C molecule and its domains upon Ca2+ binding in the presence of Mg^{2+} ions measured by a solution X-ray scattering technique. *J. Biochem.* **107**, 343-351.

Glatter, O. & Kratky, O. (1982). *Small Angle X-ray Scattering*, pp. 118-165, Academic Press, London.

Godden, J. W., Turley, S., Teller, D. C., Adman, E. T., Liu, M. Y., Payne, W. J. & LeGall, J. (1991). The 2.3 Å X-ray structure of nitrite reductase from Achromobacter cycloclastes. *Science,* **253**, 438-442.

Grossmann, J. G. & Hasnain, S. S. (1997). X-ray scattering studies of metalloproteins in solution: a quantitative approach for studying molecular conformations. *J. Appl. Crystallog.* **30**, 770-775.

Grossmann, J. G., Abraham, Z. H., Adman, E. T., Neu, M., Eady, R. R., Smith, B. E. & Hasnain, S. S. (1993). X-ray scattering using synchrotron radiation shows nitrite reductase from Achromobacter xylosoxidans to be a trimer in solution. *Biochemistry,* **32**, 7360-7366.

Hao, Q., Dodd, F. E., Grossmann, J. G. & Hasnain, S. S. (1999). Ab initio phasing using molecular envelope from solution X-ray scattering. *Acta Crystallog. sect. D,* **55**, 243-246.

Herzberg, O. & James, M. N. (1985). Structure of the calcium regulatory muscle protein troponin-C at 2.8 Å resolution. *Nature,* **313**, 653-659.

Herzberg, O., Moult, J. & James, M. N. G. (1986). A model for the Ca^{2+}-induced conformational transition of troponin C. *J. Biol. Chem.* **261**, 2638-2644.

Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. (1998). The genetic algorithm: foundations and applications in structure solution from powder diffraction data. *Acta Crystallog. sect. A,* **54**, 632-645.

Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics,* **14**, 33-38.

Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F. & Goto, Y. (1995). Structural characterization of

the molten globule and native states of apomyoglobin by solution X-ray scattering. *J. Mol. Biol.* **249**, 215-228.

Kataoka, M., Kuwajima, K., Tokunaga, F. & Goto, Y. (1997). Structural characterization of the molten globule of alpha-lactalbumin by solution X-ray scattering. *Protein Sci.* **6**, 422-430.

Krueger, J. K., McCrary, B. S., Wang, A. H. J., Shriver, J. W., Trewhella, J. & Edmonson, S. P. (1999). The solution structure of the Sac7d/DNA complex: a small angle X-ray scattering study. *Biochemistry,* **38**, 10247-10255.

Maté, M. J., Zamocky, M., Nykyri, L. M., Herzog, C., Alzari, P. M., Betzel, C., Koller, F. & Fita, I. (1999). Structure of catalase-A from *Saccharomyces cerevisiae*. *J. Mol. Biol.* **286**, 135-149.

Medrano, F. J., Gasset, M., Lopez-Zumel, C., Usobiaga, P., Garcia, J. L. & Menendez, M. (1996). Structural characterization of the unligated and choline-bound forms of the major pneumococcal autolysin LytA amidase. Conformational transitions induced by temperature. *J. Biol. Chem.* **271**, 29152-29161.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.

Miller, S. T., Hogel, J. M. & Filman, D. J. (1996). A genetic algorithm for the ab initio phasing of icosahedral viruses. *Acta Crystallog. sect. D,* **52**, 235-251.

Moncrieffe, M. C., Eaton, S., Bajzer, Z., Haydock, C., Potter, J. D., Laue, T. M. & Prendergast, F. G. (1999). Rotational and translational motion of troponin C. *J. Biol. Chem.* **274**, 17464-17470.

Nobbs, C. L., Watson, H. C. & Kendrew, J. C. (1966). Structure of deoxymyoglobin: a crystallographic study. *Nature,* **209**, 339-341.

Nogales, E., Wolf, S. G. & Downing, K. H. (1998). Structure of the alpha beta tubulin dimer by electron crystallography. *Nature,* **391**, 199-203.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

O'Connor, K. L., Shaw, L. M. & Mercurio, A. M. (1998). Release of cAMP gating by the α6β4 integrin stimulates lamellae formation and the chemotactic migration of invasive carcinoma cells. *J. Cell. Biol.* **143**, 1749-1760.

Olah, G. A. & Trewhella, J. (1994). A model structure of the muscle protein complex $4Ca^{2+}$-troponinC-troponin I derived from small-angle scattering data: implications for regulation. *Biochemistry,* **33**, 12800-12806.

Pantos, E. & Bordas, J. (1994). Supercomputer simulation of small angle X-ray scattering, electron micrographs and X-ray diffraction patterns of macromolecular structures. *J. Pure Appl. Chem.* **66**, 77-82.

Pantos, E., van Garderen, H. F., Hilbers, P. A. J., Beelen, T. P. M. & van Santen, R. A. (1996). Simulation of small angle scattering from large assemblies of multi-type scatterer particles. *J. Mol. Struct.* **383**, 303-308.

Porod, G. (1982). General theory. In *Small Angle X-ray Scattering* (Glatter, O. & Kratky, O., eds), pp. 17-51, Academic Press, London.

Satyshur, K. A., Pyzalska, D., Greaser, M., Rao, S. T. & Sundaralingam, M. (1994). Structure of skeletal muscle troponin-C at 1.78 angstrom resolution. *Acta Crystallog. sect. D,* **50**, 40-49.

Semenyuk, A. V. & Svergun, D. I. (1991). GNOM- a program package for small angle scattering data processing. *J. Appl. Crystallog.* **24**, 537-540.

Slupsky, C. M. & Sykes, B. D. (1995). NMR solution structure of calcium-saturated skeletal muscle troponin C. *Biochemistry,* **34**, 15953-15964.

Stein, P. E., Leslie, A. G., Finch, J. T. & Carrell, R. W. (1991). Crystal structure of uncleaved ovalbumin at 1.95 Å resolution. *J. Mol. Biol.* **221**, 941-959.

Sternberg, M. J., Gabb, H. A. & Jackson, R. M. (1998). Predictive docking of protein-protein and protein-DNA complexes. *Curr. Opin. Struct. Biol.* **8**, 250-256.

Stone, D. B., Timmins, P. A., Schneider, D. K., Krylova, I., Ramos, C. H. I., Reinach, F. C. & Mendelson, R. A. (1998). The effect of regulatory Ca^{2+} on the *in situ* structures of troponin C and troponin I: a neutron scattering study. *J. Mol. Biol.* **281**, 689-704.

Stuhrmann, H. B. (1970). Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particule-scattering function. *Acta Crystallog. sect. A,* **26**, 297-306.

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879-2886.

Svergun, D. I. & Stuhrmann, H. B. (1991). New developments in direct shape determination from small-angle scattering. 1. Theory and model calculations. *Acta Crystallog. sect. A,* **47**, 736-744.

Svergun, D. I., Volkov, V. V., Kozin, M. B. & Sturhrmann, H. B. (1996). New developments in direct shape determination from small angle scattering. 2. Uniqueness. *Acta Crystallog. sect. A,* **52**, 419-426.

Suzuki, S. & Naito, Y. (1990). Amino acid sequence of a novel integrin beta 4 subunit and primary expression of the mRNA in epithelial cells. *EMBO J.* **9**, 757-763.

Tjandra, N., Kuboniwa, H., Ren, H. & Bax, A. (1995). Rotational dynamics of calcium-free calmodulin studied by ^{15}N-NMR relaxation measurements. *Eur. J. Biochem.* **230**, 1014-1024.

Usobiaga, P., Medrano, F. J., Gasset, M., Garcia, J. L., Saiz, J. L., Rivas, G., Laynez, J. & Menendez, M. (1996). Structural organization of the major autolysin from *Streptococcus pneumoniae*. *J. Biol. Chem.* **271**, 6832-6838.

Varela, P. F., Romero, A., Sanz, L., Romao, M. J., Topfer-Petersen, E. & Calvete, J. J. (1997). The 2.4 Å resolution crystal structure of boar seminal plasma PSP-I/PSP- II: a zona pellucida-binding glycoprotein heterodimer of the spermadhesin family built by a CUB domain architecture. *J. Mol. Biol.* **274**, 635-649.

Wang, D., Bode, W. & Huber, R. (1985). Bovine chymotrypsinogen a X-ray crystal structure analysis and refinement of a new crystal form at 1.8 Å resolution. *J. Mol. Biol.* **185**, 595-624.

Waxman, E., Laws, W. R., Laue, T. M., Nemerson, Y. & Ross, J. B. A. (1993). Human factor VIIa and its complex with soluble tissue factor: evaluation of asymetry and conformational dynamics by ultracentrifugation and fluorescence anisotropy decay methods. *Biochemistry,* **32**, 3005-3012.

Wriggers, W., Milligan, R. A., Schulten, K. & McCammon, J. A. (1998). Self-organizing neural networks bridge the biomolecular resolution gap. *J. Mol. Biol.* **284**, 1247-1254.

Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185-195.

Zheng, Y., Doerschuck, P. C. & Johnson, J. E. (1995). Determination of three-dimensional low-resolution viral structure from solution X-ray scattering data. *Biophys. J.* **69**, 619-639.

Edited by R. Huber

(Received 4 January 2000; received in revised form 5 April 2000; accepted 5 April 2000)

http://www.academicpress.com/jmb

Supplementary material for this paper comprising one Figure describing the SAXS solution models of two proteins of unknown crystal structure as of March 2000 is available from JMB Online