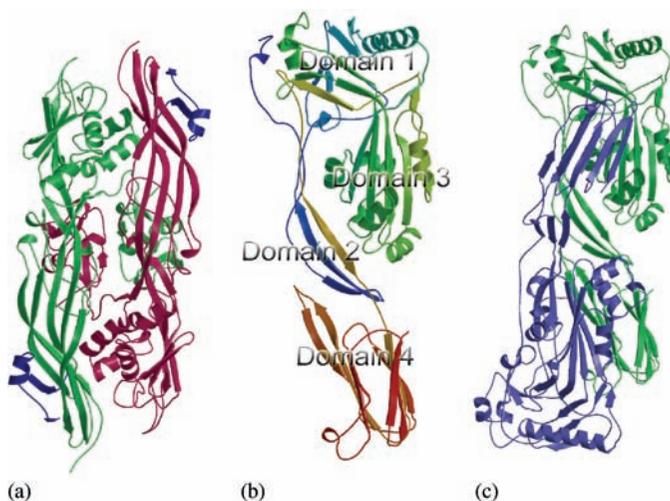


firstly recognition of and binding to a membrane receptor, and secondly pore formation. In practice, pore formation itself breaks down into a further set of steps; often proteins will oligomerise into ring-shaped structures consisting of 4–50 subunits (depending on the toxin) that define the pore, and this oligomerisation is coupled to the refolding and membrane-insertion necessary for pore formation to occur. A key question in working out the mechanism of action of these pore forming proteins is: how do they start out in solution, *i.e.*, from what state do they bind to the membrane they will permeabilise? This is related to another question: if pore forming toxins have the innate ability to refold into an alternate, oligomeric structure, what prevents them doing so in solution? AUC has proved central to working this out.

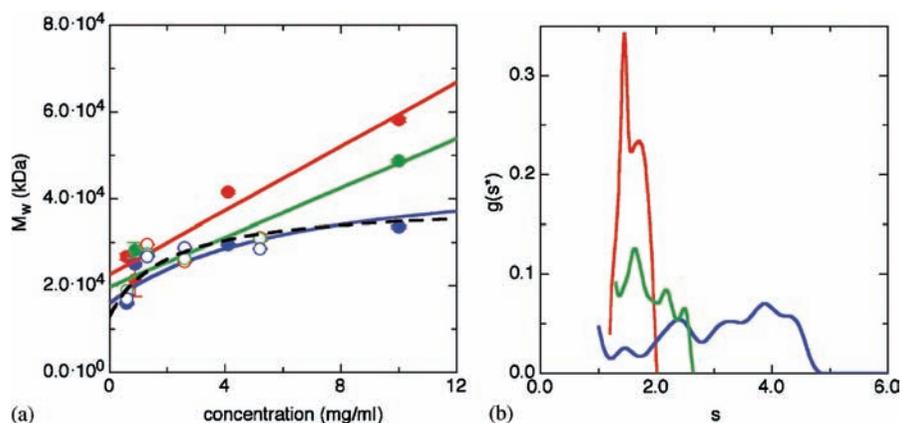
The X-ray crystal structure of aerolysin and a model of its pore-forming heptameric oligomer were published in 1994.<sup>52</sup> Aerolysin (or rather its pro-toxin, requiring proteolytic activation at the membrane surface by furin, whereby a C-terminal peptide is removed) was revealed to be a dimeric structure [Figure 1(a)]. Intriguingly, the dimer is in a head-to-tail (antiparallel) topology; since the toxin forms a parallel heptameric pore in its target membrane, doing so must involve separation of the dimer subunits, reorientation of the subunits and then oligomerisation, before membrane insertion. Alternatively, the dimer could dissociate before binding to the membrane, which might be a simpler scheme. Nevertheless, as shown by AUC, proaerolysin remains a dimer even at  $5 \mu\text{g mL}^{-1}$  concentration – and in that dimeric state is competent to bind its receptor.<sup>53</sup> Enterolysin, too, is a dimer in an antiparallel arrangement.<sup>50</sup> So, in the case of aerolysin the more complicated of two conceivable



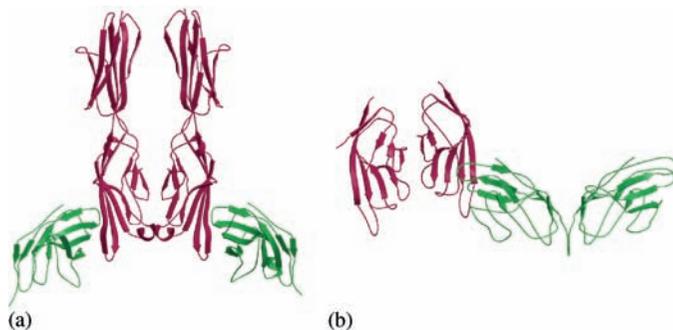
**Figure 1** (a) Crystallographic structure of the proaerolysin dimer.<sup>52</sup> The two antiparallel protomers are coloured aquamarine and cerise, respectively. The propeptide is coloured blue in both cases. (b) Crystallographic structure of perfringolysin.<sup>58</sup> The structure is coloured from blue at its N-terminus to red at its C-terminus. (c) The dimer of perfringolysin found in protein crystals.<sup>58</sup> The antiparallel protomers are colour aquamarine and blue; the aquamarine molecule is in the same orientation as in (b)

dimerised and did not otherwise non-specifically aggregate [Figure 2(a)]. This is apparent in the linearly increasing observed mass at lower speeds where aggregates have yet to be removed by the increased gravitational field at high speed, and the plateau-forming rectangular hyperbola that describes nsp9 dimerisation at high speed and when denuded of the His<sub>6</sub> tag.<sup>7</sup>

Having determined the structure and oligomerisation state of nsp9, it was necessary to try and assign a function to it on the basis of its interactions with other binding partners – nucleic acid, for example, or other proteins that derive from the replicase polyprotein. The binding partners investigated were nsp8 and nsp5 (which is the 3C-like protease of SARS CoV), in binary mixtures and when all three were combined. Nsp8 behaved in a non-ideal manner in sedimentation equilibrium experiments but did appear to be dimeric, with a solution molecular weight of ~50 kDa. Nsp5 had a molecular weight consonant with it being monomeric, and behaved in an ideal manner. Here, ideality means simply that one can observe one or more apparently independently sedimenting species; non-ideality means that some interparticle interactions are present – perhaps a charge:charge interaction, aggregation due to misfolding or a crowding effect. It was found that nsp8 behaved ideally in combination with nsp9, which suggests that nsp9 alters the behaviour of nsp8 and thus that they can interact. Use of  $g(s^*)$  analysis, to probe this insight further, seemed to confirm it – nsp9 showed two peaks, equal to the monomeric and dimeric species; nsp8 showed a broad  $g(s^*)$  spectrum, consistent with its non-ideality; a mixture of the two gave three discrete peaks, indicating a conversion of the behaviour of nsp8 in solution from a disordered melée to interacting with nsp9 in some ordered way to produce ~3 differentially sedimenting species [Figure 2(b)]. AUC here provides a



**Figure 2** (a) Plots of apparent molecular weight against concentration for tagged SARS-CoV nsp9 (closed symbols, solid lines) at 12 000 rpm (red symbols), 15 000 rpm (green) and 22 000 rpm (blue) derived from AUC experiments. For untagged nsp9 (open symbols, colours as for tagged), only the plot at 22 000 rpm is shown (blue broken line) for clarity as the measured values were similar at all three speeds. (b)  $g(s^*)$  profiles of nsp9 (red), SARS-CoV nsp8 (green), and an equimolar mixture of the two (blue) showing a change in the behaviour of nsp8 on addition of nsp9

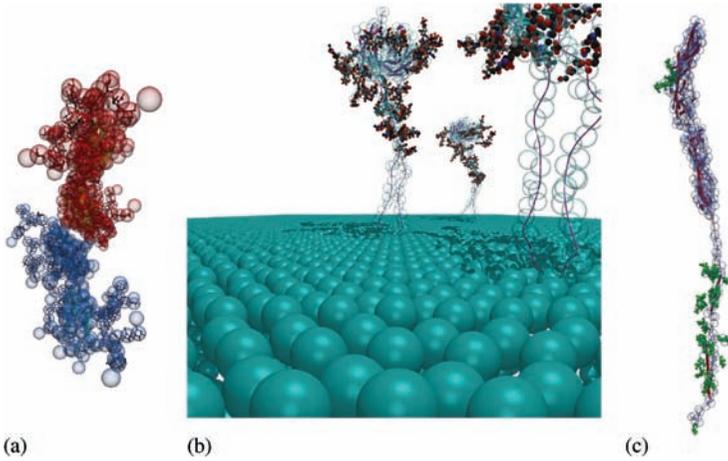


**Figure 3** (a) The crystallographic complex formed between B7-1 and CTLA-4 complex. B7-1 is coloured cerise and monomers of CTLA-4 aquamarine. CTLA-4 is normally a constitutive dimer via a disulfide bridge between monomers, which was mutated away in the construct used for crystallography.<sup>68</sup> (b) The crystallographic complex formed between B7-2 (receptor-binding domain only) and CTLA-4, which in this case was a dimer. B7-2 is cerise, CTLA-4 aquamarine: note the asymmetry in the B7-2 dimerisation interface<sup>69</sup>

Another molecule involved in enhancing the stimulation of B- and T-cell activity is known as SLAM, for signalling lymphocytic activation molecule. SLAM is part of the CD2 subset of the immunoglobulin superfamily and had been thought to be homophilic – that is, to form dimers through its receptor-binding surfaces, head-to-head. To show this it was necessary to use sedimentation velocity with data analysis by  $g(s^*)$  because the interaction was too weak to be detected at available concentrations in equilibrium mode.<sup>8</sup> The  $K_d$  of the self-interaction of SLAM was estimated, using the relative areas of the Gaussian distributions of monomer and dimer into which the total apparent sedimentation coefficient distribution in the sample ( $g(s^*)$ ) could be deconvoluted, as 0.1–1 mM. Furthermore, it could be shown using the frictional ratio of the monomeric and dimeric species (expressed as Perrin functions) in combination with their sedimentation coefficients and bead modelling that the interaction produced an elongated structure – that is, that SLAM is indeed a self-ligand interacting head to head.<sup>8</sup> The frictional coefficient can be calculated directly<sup>8</sup> or, for example, within software such as SEDFIT,<sup>9,74,75</sup> using the equation<sup>76</sup>

$$\frac{f}{f_0} = \frac{M(1 - \bar{v}\rho_0)}{N_A 6\pi\eta_0 s_{20,w}^0 \left( \frac{3\bar{v}M}{4\pi N_A} \right)^{1/3}} \quad (3)$$

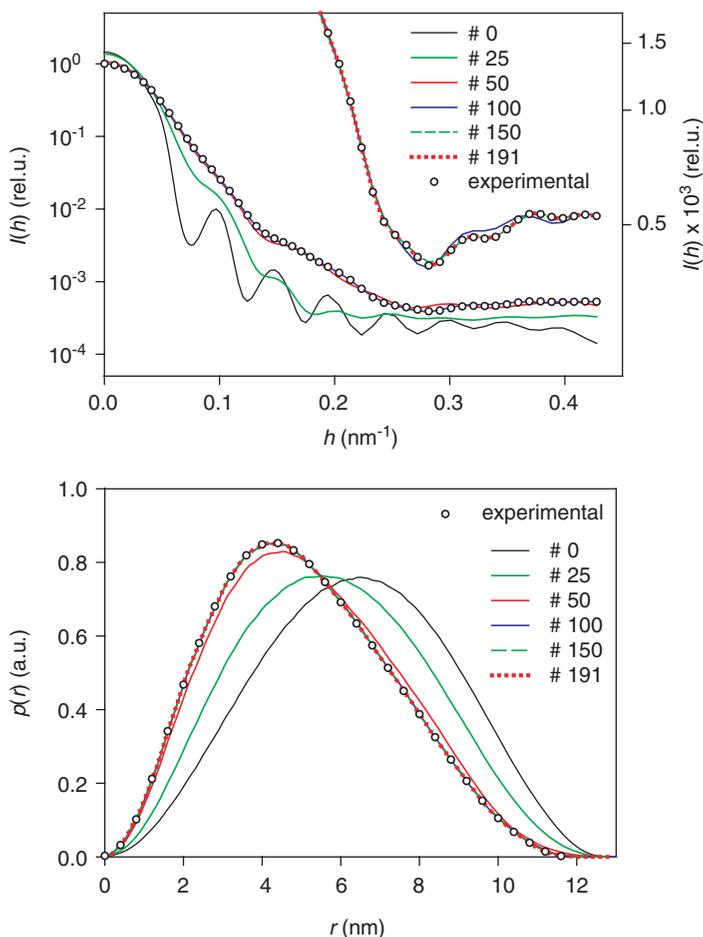
where  $M$  is molecular weight,  $N_A$  the Avogadro's number and the  $\eta$  viscosity. A Perrin function can then be calculated as described in Equation (1): this is a measure of particle shape (particularly elongation) independent of weight. Then for bead models the same value can be computed theoretically using SOLPRO, for example,<sup>77</sup> and comparison will indicate to what extent the bead model is consistent with the behaviour of the protein in hand [Figure 4(a)].



**Figure 4** (a) An atomic model for a SLAM dimer with modelled glycan sidechains within a semi-transparent representation of a bead model used to compute the expected behaviour of this low-affinity homodimer in solution.<sup>8</sup> (b) A depiction of CD8 atop its extended linker regions with modelled glycans; AUC was used to show that the level of extension shown by these linkers is maintained even in the presence of merely core sugar sidechains. Lipid headgroups defining the membrane surface: cyan spheres; CD8 is encased in its hydrodynamic bead model as in (a).<sup>9</sup> (c) A depiction of DAF, which consists of four domains in an extended arrangement supported by an extended linker, like CD8. AUC was used to show that the folded domains adopt this arrangement in solution and to probe the role of the linker regions in supporting them. Sugars are shown in green, the polypeptide backbone in red.<sup>78</sup>

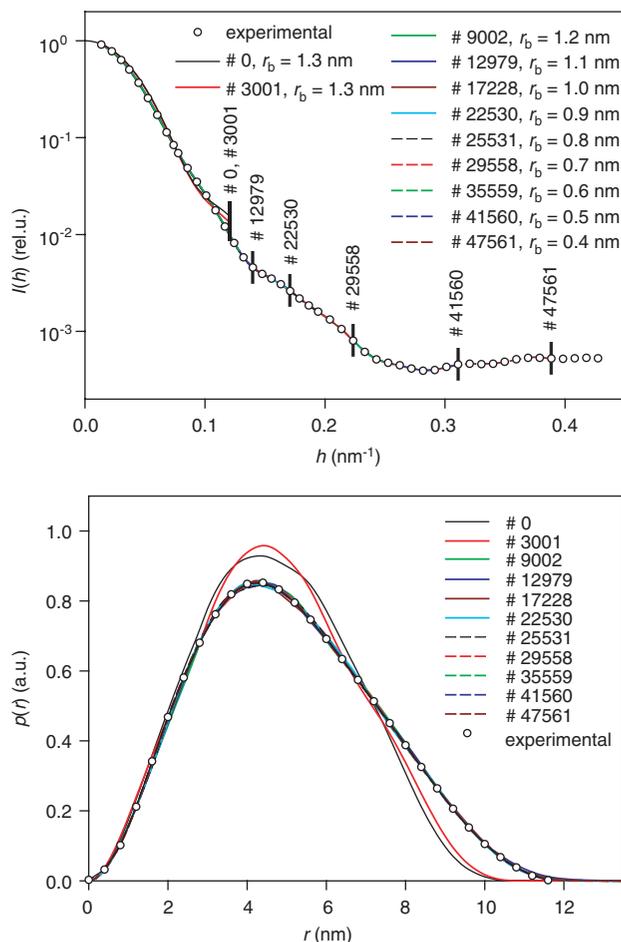
A similar approach was used to address the role of glycosylation in the structure of the linker regions by which CD8, in assisting T-cell receptors to recognise MHC molecules, reaches up to interact with the MHC close to the cell membrane of the APC. CD8 has extensive N-linked glycosylation on its head regions and O-linked glycosylation on its extended linkers by which it stands above the T-cell membrane. Since these linkers have no secondary structure it was thought that the branched O-glycans would have a scaffolding role. To investigate whether that is indeed the case, forms of CD8 were generated in which only the core sugars were added during synthesis at the O-glycan sequons; surprisingly, using the computation of Perrin functions from sedimentation velocity data compared to bead models it could be shown that in fact even with relatively small sugar sidegroups the linker regions remain extended.<sup>9</sup> Furthermore, by generating a series of entirely theoretical conformations for the stalk regions it could be shown that in this case the Perrin function and sedimentation coefficient indicated the extension of the stalks with a fair degree of discrimination compared to other possibilities<sup>9</sup> [Figure 4(b)].

Sedimentation velocity has also been used to show that the molecule decay accelerating factor (DAF; also known as CD55), a four-domain protein, possesses a conformation in solution similar to that seen in its crystal. DAF is an elongated molecule [Figure 4(c)] and a natural concern is that the domain arrangement derives from



**Figure 3** Evolution of a DAMMIN model of StP. The scattering curves  $I(h)$  (top) and pair-distance distribution functions  $p(r)$  (bottom) calculated for selected snapshots of the bead models of Figure 2 at different steps of the SA process are compared with the corresponding experimental functions

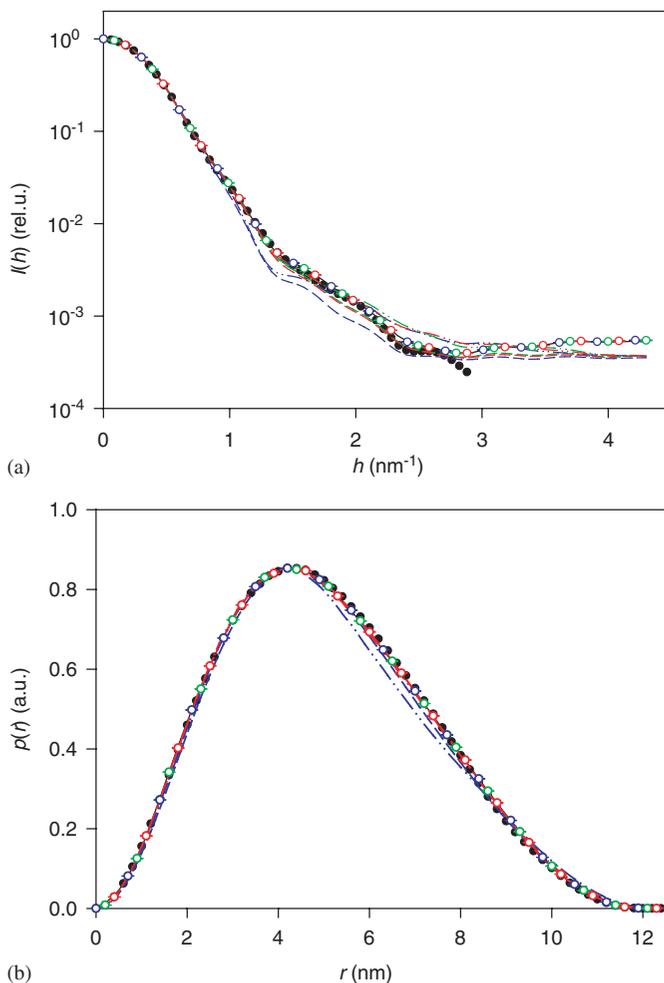
Summarising the most relevant models of a detailed application of DAMMIN calculations furnishes a gallery of similar models, irrespective of the constraints or averaging procedures applied (Figure 9: individual models; Figure 10: averaged models), provided that the applied constraints are based on realistic assumptions concerning symmetry ( $C_2$  corresponding to DAMMIN code P2). All models reveal an oblate particle shape, similar to the SAXS trial-and-error model<sup>80</sup> or the 3D image of related glycogen phosphorylases derived just recently from crystallography.<sup>81,82,84,85</sup> By contrast, the application of unrealistic constraints such as P3–P6 may reveal models that are irrelevant from the biological point of view. Surface representations of several models under analysis add weight to the previous considerations concerning the shape of the protein. The overall impression is that of an oblate spheroid without



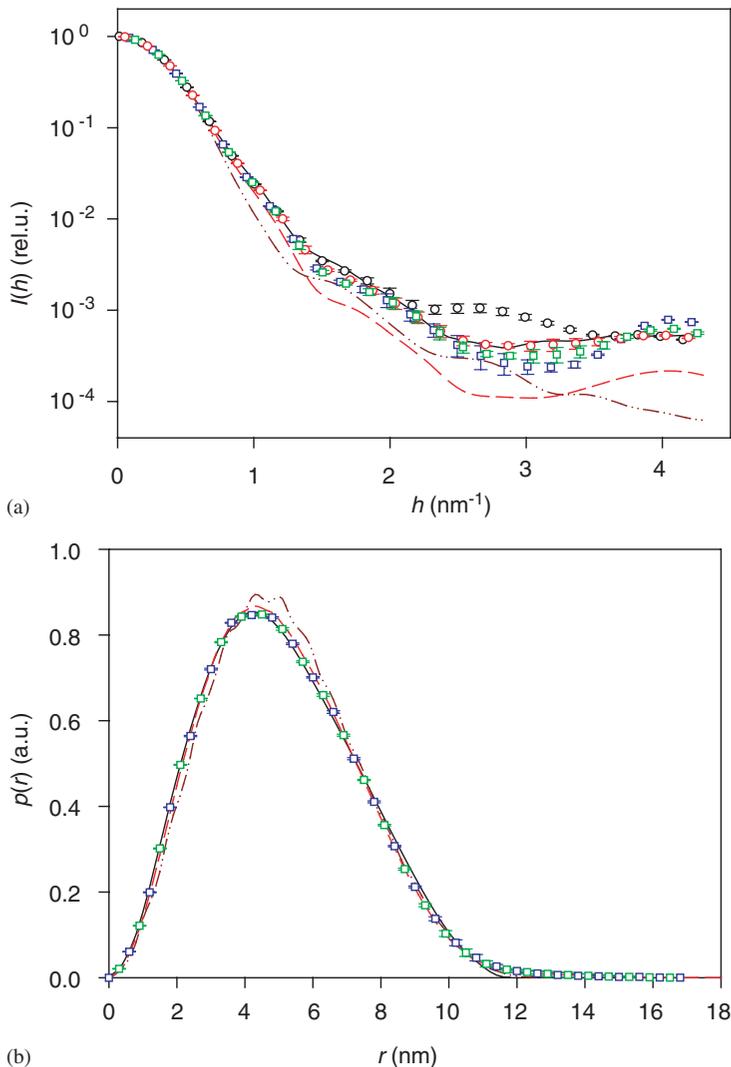
**Figure 6** Evolution of DALAI\_GA models of StP. The scattering curves  $I(h)$  (top) and pair-distance distribution functions  $p(r)$  (bottom) calculated for selected snapshots of bead models of Figure 5 at different steps of the GA process are compared with the corresponding experimental functions. The vertical bars (top) indicate the respective upper limit of the portion of the experimental  $I(h)$  curve that is analysed in the GA process at the given level of resolution

models of the isolated haeme and FAD domains with a hypothetical linker to provide an elongate model for the entire protein. Since, however, *ab initio* DAMMIN calculations<sup>2,49</sup> yielded models for all major components under consideration [holoprotein, haeme and FAD components; Figure (15a)], it was tempting to combine the two isolated domains and thus obtain a model for the complete entity. This model was compared with the crystallographic data of the isolated domains and used to derive properties of hydrodynamic interest.

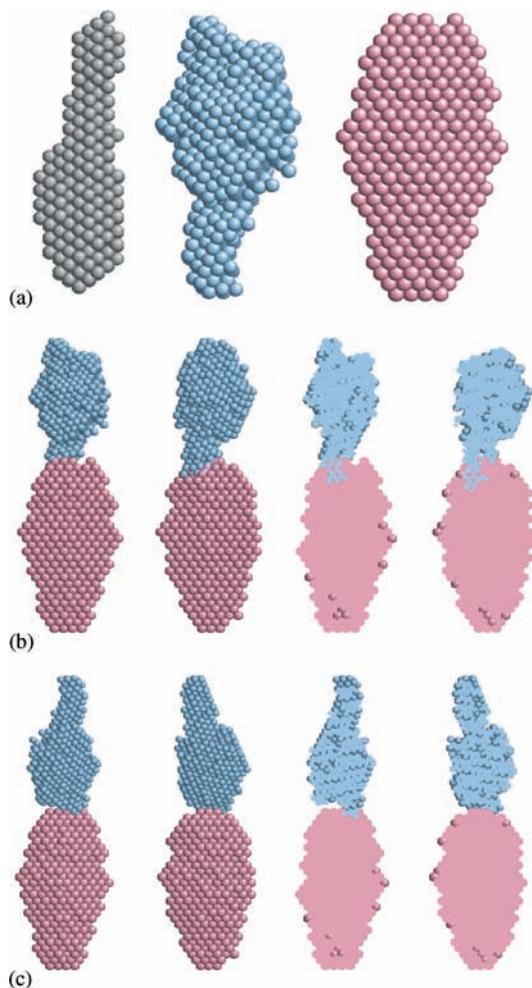
As opposed to the modus operandi used in the trial-and-error approaches, we abandoned modelling the linker, but made allowance for appropriate alignment and



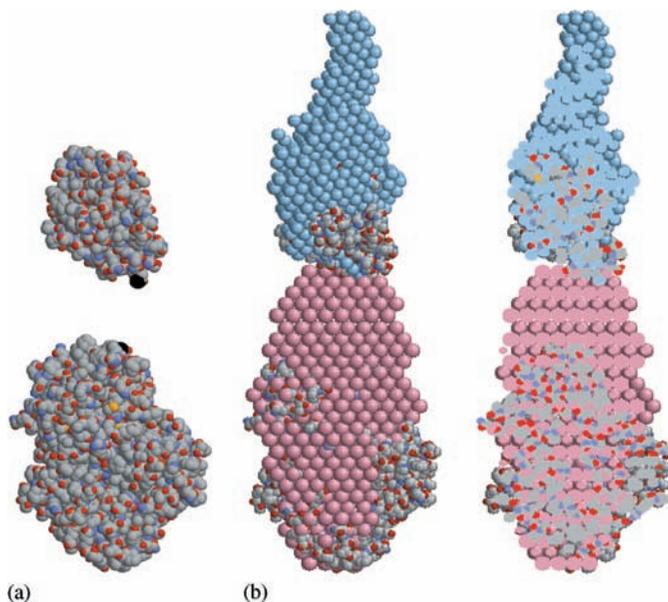
**Figure 12** Comparison of experimental and model  $I(h)$  (a) and  $p(r)$  (b) curves of StP. The experimental functions (solid black lines) are nearly completely overlaid by functions calculated for various models: the black circles represent the functions obtained for the previously established trial-and-error model;<sup>80</sup> the open circles show the averaged functions for DAMMIN models, which were generated on the assumption of  $C_2$  symmetry without any shape constraint (blue circles) or with preference to oblate shape with the symmetry axis along the shortest dimension (green circles) or across to it (red circles). In general, the averages were calculated in each case from six independent models; the corresponding standard deviations are indicated by error bars which are, however, very small and poorly visible. The SAXS functions obtained for averaged structures are shown as dashed and dot-dashed lines (the colour codes are the same as for the circles): the dashed lines refer to averaged models composed of unequal beads, the bead size reflecting the probability of occupation; averaged structures obtained by means of the DAMAVER suite consist of equal-sized beads, their scattering functions are presented as dot-dashed lines. (b) The green lines have been omitted because of their close coincidence with the red lines



**Figure 14** Comparison of the experimental  $I(h)$  and  $p(r)$  curves (solid black lines in (a) and (b)) of StP with the averaged functions of models established by different versions of SAXS3D. In (a), the black and blue symbols represent averaged intensities of four models created by means of the original version of SAXS3D; a refinement of these initial models by using a modified program version finally resulted in the averaged intensities shown by the red and green symbols. The error bars symbolise the standard deviation of the averaged intensities. The circles are based on the intensities provided by SAXS3D itself, whereas the squares represent intensities that were calculated afterwards from the bead coordinates by exact usage of Debye's formula. (b) The symbols represent the averaged  $p(r)$  functions of the initial (blue squares) and refined (green squares) models. The scattering intensities (a) and  $p(r)$  functions (b) computed for averaged models are shown as dashed (averaging by DAMHEXMIX) and dot-dashed (averaging by DAMAVER) coloured lines



**Figure 15** Reconstructions of the structure of the CDH holoprotein and its isolated domains. (a) Aligned averaged models for the entire holoprotein (grey;  $N_b=431$ ,  $r_b=0.4$  nm), the haeme domain (blue;  $N_b=909$ ,  $r_b=0.2$  nm), and the FAD domain (pink;  $N_b=1324$ ,  $r_b=0.25$  nm); the images of the models are not to scale. The initial models are based on experimental scattering profiles and *ab initio* DAMMIN calculations in the 'expert mode'. Averaging and filtering of the DAMMIN models was performed by means of the DAMAVER program suite. The models for the domains were aligned along the  $y$  direction and rotated around  $z$  or  $x$ , to make the domains ready for the docking manoeuvres to be executed. (b) and (c) Two different views ( $90^\circ$  rotation around the  $y$  axis) of models for the docked haeme and FAD domains, together with the corresponding central slabs. The required distance (about 7.4 nm) between the centres of gravity of the two averaged *ab initio* models of the domains has been computed by considering their values for  $R_G$  and  $V$ , in order to achieve the  $R_G$  value of the averaged model for the holoprotein. Model (b) shows the results of the first trial of a docking procedure (docking based on the initial orientation of the domains). The best-fitting model (c) is characterised by a minimum of overlapping of the domains and a maximum value for  $d_{max}$ .

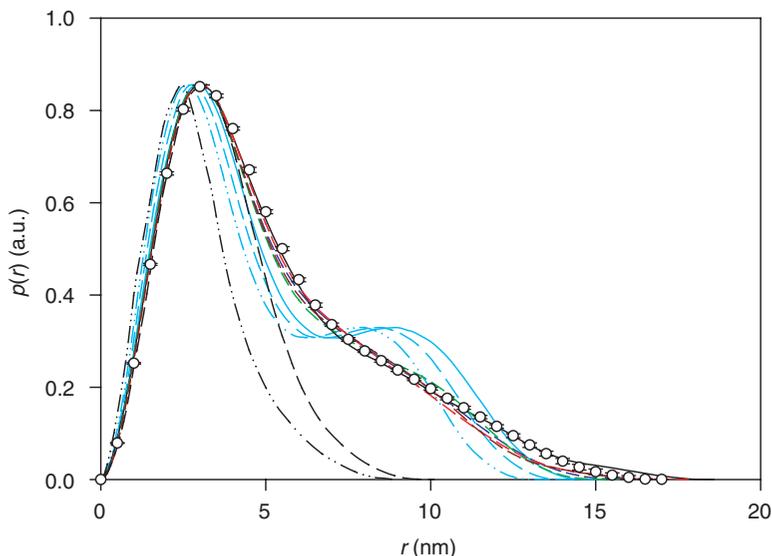


**Figure 16** Space-filling models of the crystal data of the haeme and FAD domains of CDH in usual CPK colours (C in light grey, O in red, N in light blue, and S in yellow). (a) Aligned models of the haeme and FAD domains. The coordinates of the basic protein atoms were taken from the available PDB files (1D7C, 1KDG); hetero atoms of the prosthetic groups, glycan chains and water molecules were neglected. The domains were arranged similar to the docked DAMAVER models described in Figure 15, i.e. at the required distance and facing each other. The FAD domain of the crystal structure contains a few residues of the linker. The terminal atoms of the crystal structures are marked in black (enlarged spheres); the gap between the two domains is caused by the distance specifications between the respective centres of gravity of the two domains. (b) Comparison of the aligned models obtained from crystal data (in CPK colours) with DAMMIN reconstructions (blue and pink spheres), together with the corresponding central slabs

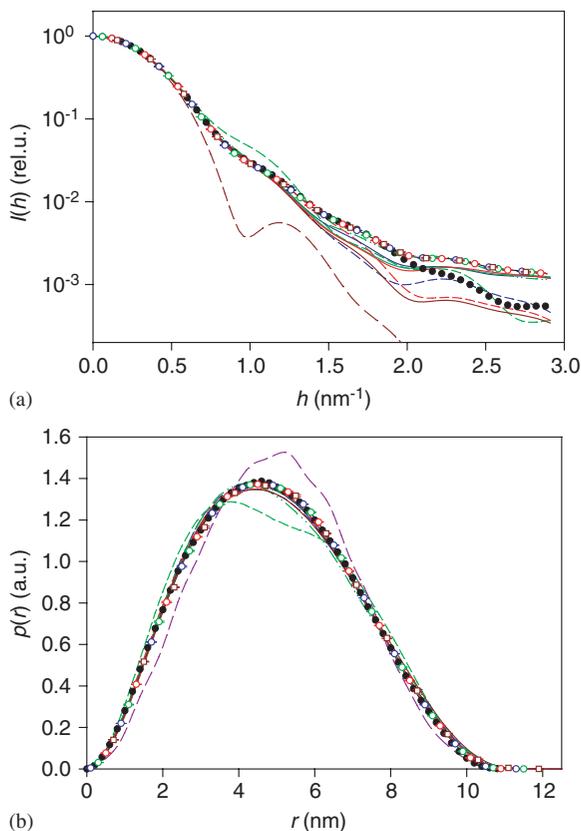
models, and the disagreement of the other parameters to the lack of mass in the centre and other regions of the molecule. It is plausible to attribute the missing mass to the linker chains forming a kind of scaffold in the native *L.t.* Hb architecture. As may be expected, the WB approach of the HBL complex supplies acceptable values for all parameters mentioned, since it does not use the subunit data; by contrast, use of the molar mass corresponding to the sum of the subunit masses yields an  $s$  value as low as with the HBL models constructed from DAMMIN models of the subunits.

#### 4.4 RNA from the Bacteriophage MS2: Modelling the Solution Structure of a Free Nucleic Acid

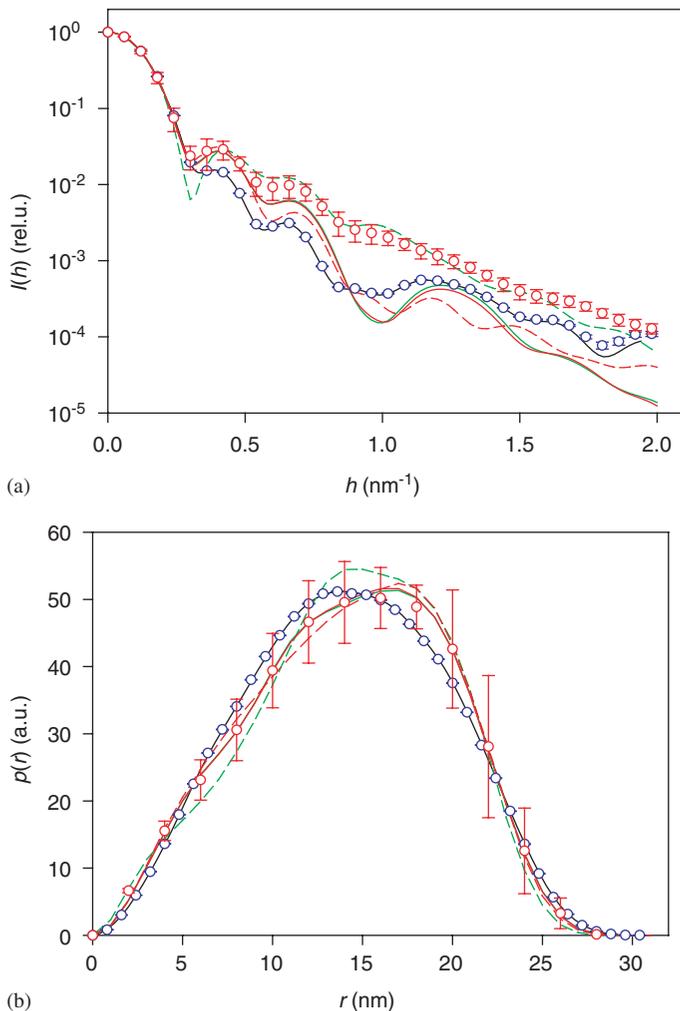
Previous SAXS data<sup>105</sup> of MS2 RNA were exploited for tentatively modelling the structure of a free RNA in solution by *ab initio* approaches, similar to the procedures adopted for proteins.



**Figure 17** Comparison of the experimental pair-distance distribution,  $p(r)$ , (solid black line) of the CDH holoprotein with the  $p(r)$  functions of various models. The black circles with error bars show the mean  $p(r)$  function and the corresponding standard deviations obtained by averaging the functions of 30 different *ab initio* DAMMIN models of the holoprotein. The dashed blue, red, green and brown lines represent the  $p(r)$  functions calculated for models composed of *ab initio* DAMMIN models of the two isolated domains (haeme domain, FAD domain). The domains were aligned along the  $y$  direction, with a constant distance of about 7.4 nm between their centres of gravity; the mutual orientation of the domains was systematically changed by rotations around  $y$  and  $z$ , respectively. The original arrangement of the domains yielded the  $p(r)$  function drawn in blue, rotations of the haeme domain around  $y$  by  $\pm 30^\circ$  did not lead to any noticeable change in the  $p(r)$  function. Red symbolises the  $p(r)$  function obtained after rotating the haeme domain by  $180^\circ$  around the  $z$  direction; the green line was obtained after rotating the FAD domain by  $180^\circ$  around  $z$ ; rotations of both domains by  $180^\circ$  around  $z$  yielded the brown line. The lines coloured in cyan represent  $p(r)$  functions calculated for models derived from the crystal structures of both domains; for these models only the coordinates of protein atoms were evaluated, and atoms belonging to hetero groups (FAD, haeme, carbohydrates) were neglected. Again at a distance of about 7.4 nm between the centres of gravity, the domains were arranged in such a way that the C-terminus of the haeme domain and the N-terminus of the FAD domain were facing each other. The dot-dashed cyan line is the  $p(r)$  function obtained without rescaling the coordinates of the crystal structure. Since the values for both  $R_G$  and  $d_{max}$  of this model are quite low if compared to the experimental data for CDH or the parameters corresponding to the DAMMIN models, the coordinates of the crystal structure were upscaled to match the mean  $R_G$  value of the DAMMIN models (dashed cyan line) or the experimental  $R_G$  of CDH (solid cyan line). For convenience, the mean  $p(r)$  function of the DAMMIN models for the isolated domains (dot-dashed black line: haeme domain; dashed black line: FAD domain) are additionally shown in the figure. To facilitate the comparison, all  $p(r)$  functions presented have been normalised to the same height of the maximum

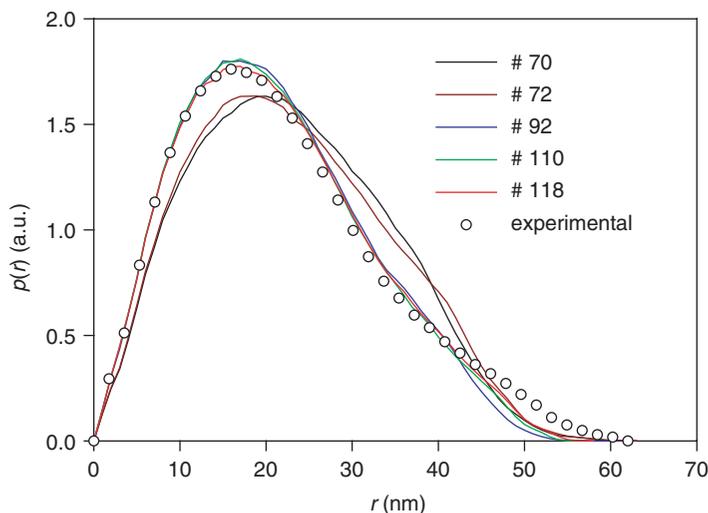


**Figure 20** SAXS curves  $I(h)$  (a) and pair-distance distribution functions  $p(r)$  (b) of the dodecameric subunit of *L.t. Hb*. The experimental functions (solid black lines) are nearly completely overlaid by functions calculated for various models: the black circles represent the functions obtained for the previously established trial-and-error model;<sup>103</sup> the open circles show the averaged functions for DAMMIN models which were generated without any a priori information (blue circles) or with the assumption of  $C_3$  symmetry, without shape constraint (green circles) or with preference to oblate shape (red circles). The brown squares represent the averaged functions obtained from indirectly EM-biased DAMMIN models. The averages were calculated from 7 to 48 independent models; the corresponding standard deviations are indicated by error bars which are, however, very small and poorly visible. The SAXS functions obtained for averaged structures are shown as coloured lines (the colour codes are the same as for the circles). The blue, green and red lines refer to DAMAVERaged models (obtained by means of the DAMAVER suite) consisting of equal-sized beads; the dashed lines are based on models which were generated by DAMMIN without the automatic subtraction of a constant background from the experimental scattering curve; the solid and dot-dashed lines refer to models where this background subtraction was applied (for the presentation, the background was readded to the  $I(h)$  curves); the green solid lines refer to models of oblate shape and the dot-dashed lines to models of prolate shape. The solid brown lines represent the functions obtained for DAMAVERaged biased models, the dashed brown lines refer to averaged structures (obtained by means of DAMMIX) composed of unequal beads, where the bead size reflects the probability of occupation



**Figure 22** Comparison of  $I(h)$  (a) and  $p(r)$  (b) functions of various HBL models based on DAMMIN models for the dodecameric subunit of *L.t. Hb* with the experimental functions (black lines) and the functions for EM-biased DAMMIN models for the native HBL complex (blue circles). The red circles represent the averaged functions of the HBL models built from the 48 different subunit models; the functions of the corresponding averaged models are shown as red lines: the solid lines refer to averaging by DAMMIX, the dashed lines to averaging by the DAMAVER suite. The green lines represent the functions of HBL models constructed from averaged models for the subunit; again the averaging procedure applied is indicated by solid (DAMMIX) and dashed lines (DAMVER), respectively

in other types of RNA (L-form of tRNA or rRNA<sup>112</sup>), and the surface peculiarities are in accord with secondary structure details in terms of “flower” or “bouquet” models.<sup>104</sup> Owing to the modus operandi of the averaging procedures applied, the DAMHEXMIX procedure accentuates the peculiar features of nucleic acid solution structure, whereas

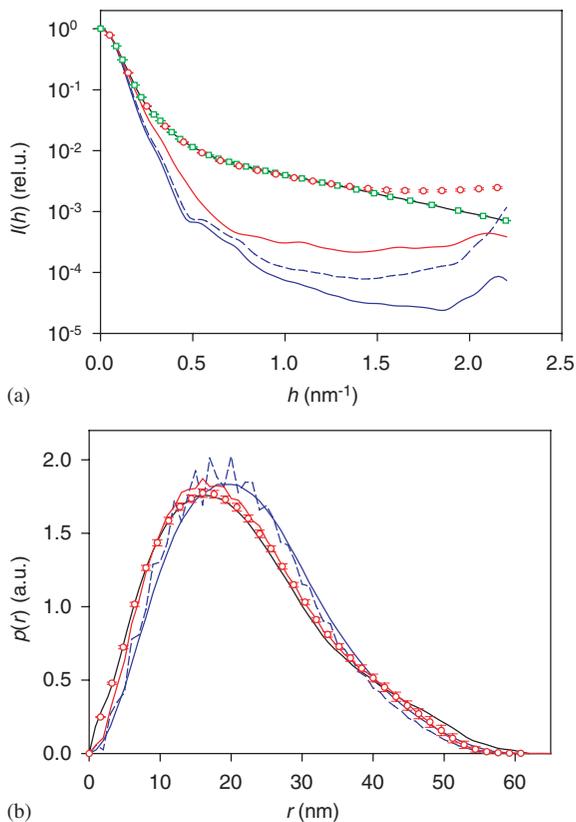


**Figure 24** Evolution of the PDDFs of the DAMMIN models for MS2 RNA shown in Figure 23

application of the DAMAVER suite rather polishes the particle surface, thus eliminating characteristic details of the secondary structure of RNA in solution.

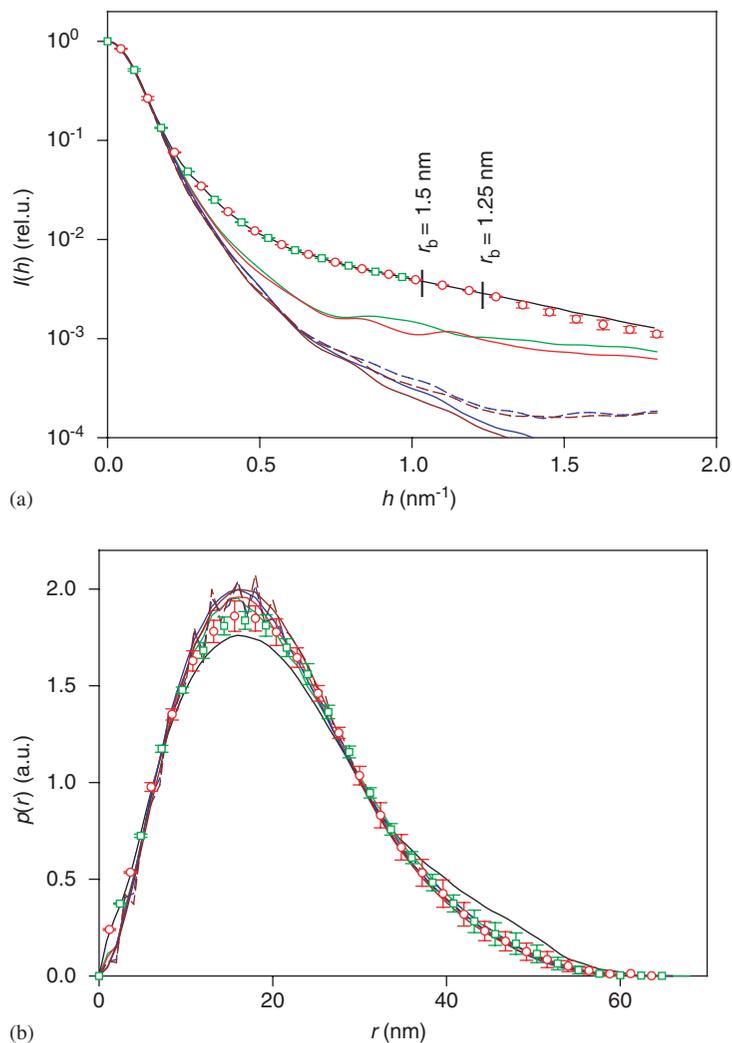
Surface representation techniques, such as those inherent in the CCP4 package, mirror the most essential features of the surface characteristics of RNA in solution. The surface renderings (Figure 26) corroborate the impression of the structural peculiarities (uneven surface, bent structure) already gained from the inspection of the models depicted in Figure 25. The less-detailed molecular features of these representations, however, characterise and govern the hydrodynamic behaviour more purposively.

Figures 27–29 illustrate the SAXS profiles of experimental and model data created by the applied *ab initio* approaches and the subsequent averaging procedures. The DAMMIN approach (Figure 27) convincingly demonstrates excellent agreement between observed  $I(h)$  profiles and DAMMIN calculations up to high angles, and fair agreement between experimental data and model calculations based on bead coordinates by means of Debye's formula. The models obtained by averaging the restored structures exhibit marked deviations from the experimental data at rather small angles ( $h < 0.2 \text{ nm}^{-1}$ ). Averaging of the models by DAMHEXMIX was superior to that of the DAMAVER suite; in the latter case, however, discrepancies can be reduced slightly by rescaling procedures. Although less pronounced, similar findings are encountered when looking at the corresponding PDDFs. A comparison of the scattering curves and PDDFs of DALAI\_GA models with the experimental data (Figure 28) reveals a behaviour similar to that found for DAMMIN: excellent agreement between experimental data and DALAI\_GA calculations, and marked deviations, even at small angles, between the experiment and the calculations applying the averaged models. The agreement between experimental PDDFs and those obtained by DALAI\_GA is inferior to that produced by DAMMIN. Application of the SAXS3D

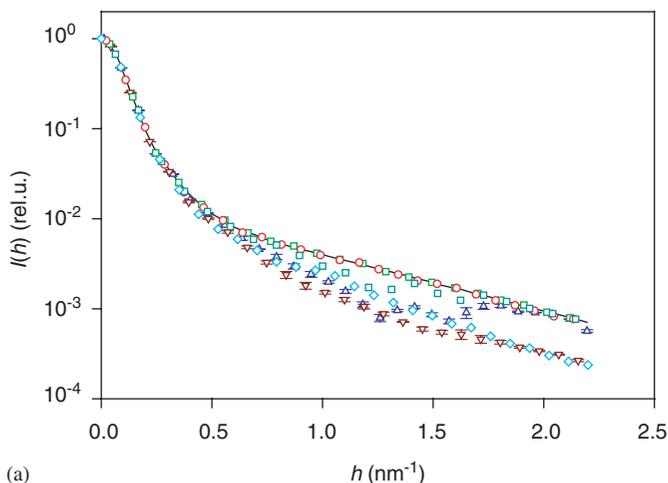


**Figure 27** Comparison of experimental and DAMMIN model scattering curves (a) and PDDFs (b) of MS2 RNA. The experimental functions are shown as solid black lines. The symbols with error bars represent the mean values and standard deviations obtained from evaluating a set of 19 *ab initio* DAMMIN models. (a) The green squares reflect the mean values of the  $I(h)$  curves calculated directly by DAMMIN (by using spherical harmonics) and the red circles symbolise the averaged  $I(h)$  curves calculated from the coordinates of the beads (by means of Debye's formula); to achieve maximum coincidence of the  $I(h)$  curves, the nominal bead radius  $r_b$  had to be reduced by a factor 0.5 in the Debye calculations. The coloured lines represent the  $I(h)$  (a) and  $p(r)$  (b) curves of averaged models; these models were obtained by averaging the structures of the 19 *ab initio* DAMMIN models by using our program DAMHEXMIX (red lines) or Svergun's DAMAVER suite (blue lines). Since the volume of the initial averaged model generated by DAMAVER was too big, this model was rescaled by applying two empirical factors to the bead radii and coordinates, respectively; the SAXS functions of the rescaled model are shown as dashed lines

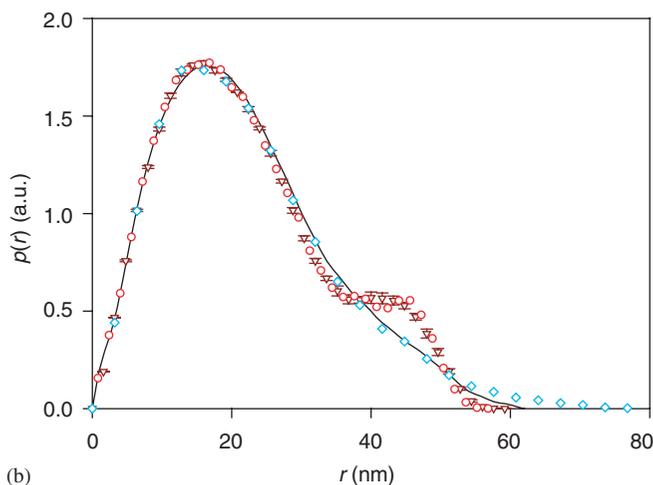
Quantification of the results of the *ab initio* modelling approaches in terms of structural and hydrodynamic parameters (Table 5) shows fair agreement between observables and predicted parameters for all procedures applied. Irrespective of the constraints selected, the DAMMIN calculations yield reasonable anticipations of the



**Figure 28** Comparison of experimental and DALAI\_GA model scattering curves (a) and PDDFs (b) of MS2 RNA. The experimental functions are shown as solid black lines. The symbols with error bars represent the mean values and standard deviations obtained from evaluating two related sets of *ab initio* DALAI\_GA models. The coloured lines reflect the functions calculated for averaged models: these models were obtained by averaging the structures of five *ab initio* models by means of our program DAMHEXMIX (green and red lines) or Svergun's DAMAVER suite (blue and brown lines). The functions drawn in green or blue refer to models based on a nominal bead radius of 1.5 nm, and those drawn in red or brown to models based on a nominal  $r_b$  of 1.25 nm. The vertical bars (a) indicate the corresponding upper limits of the portions of experimental  $I(h)$  used in the DALAI\_GA analysis. Since the volumes of the initial averaged models generated by DAMAVER were too big, these models were rescaled by applying empirical factors to the bead radii and coordinates, respectively; the SAXS functions of the rescaled models are shown as dashed lines



(a)

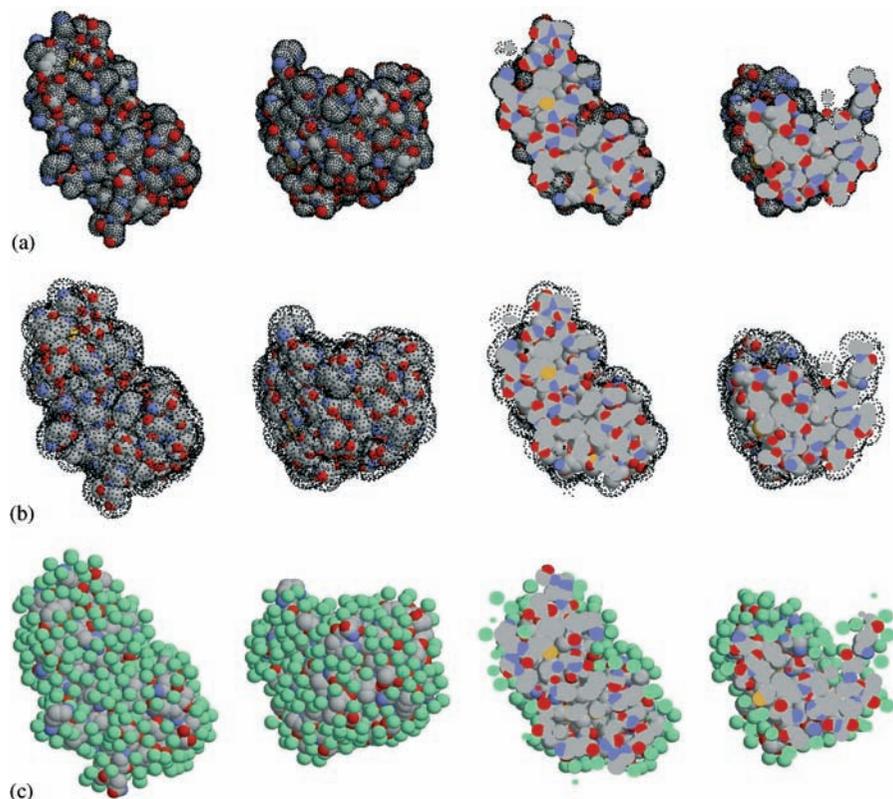


(b)

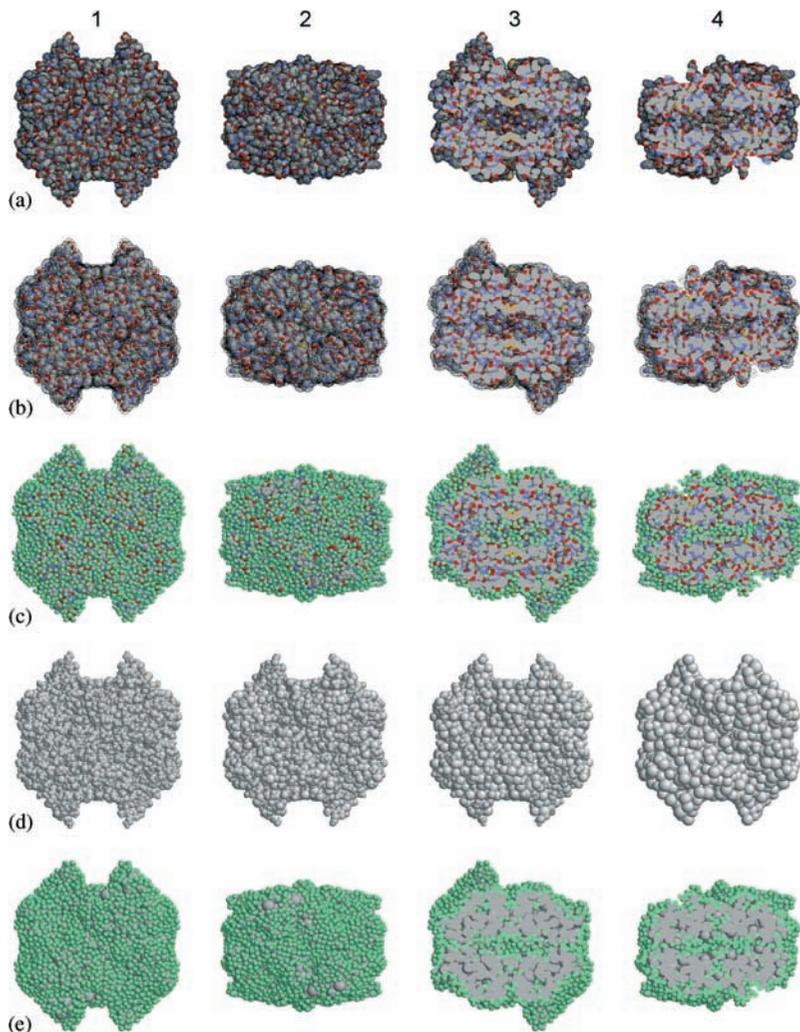
**Figure 29** Comparison of experimental and SAXS3D model scattering curves (a) and PDDFs (b) of MS2 RNA. The experimental functions are shown as solid black lines. The symbols with error bars represent the mean values and standard deviations obtained from evaluating three similar *ab initio* models generated by using the original version of SAXS3D; symbols without error bars refer to single models. (a) The blue triangles reflect the mean values of the  $I(h)$  curves calculated directly by the original version of SAXS3D (using Debye's formula in combination with an approximate distance histogram) and the brown triangles symbolise the average of the  $I(h)$  curves that were calculated from the coordinates of the beads by an exact application of Debye's formula; similarly, the dark cyan squares and the cyan diamonds symbolise the SAXS3D and the Debye scattering curve, respectively, of an alternative model, also created by means of the original program version. On the contrary, the green squares represent the SAXS3D scattering curve of another model that was established by using a modified program version, and the red circles correspond to the exact Debye scattering curve for that model. (b) The brown triangles, red circles and cyan diamonds refer to the same models as in (a)

about 200–300 water molecules is preferentially bound to the protein consisting of 124 AA residues. The comparison of predicted and observed  $s$ ,  $D$  and  $[\eta]$  values, again, proves that the anhydrous model is not suitable for correct parameter predictions, whereas a hydrated model ( $V_w=0.0270$  or  $0.0284$  nm<sup>3</sup> and  $f_K=1.3$ – $1.5$ ) is capable of an appropriate prediction.

The results obtained for LDH also demonstrate the necessity of applying hydrated models. The anhydrous models were not able to simulate scattering or hydrodynamic



**Figure 6** Space-filling models for anhydrous and hydrated RNase in two different views ( $90^\circ$  rotation around the  $x$  axis), together with the corresponding central slabs. (a) Molecular dot surface points created by the program SIMS for the anhydrous contour ( $d_{\text{dot}}=300$  nm<sup>-2</sup>, and applying  $r_{\text{probe}}=0.1525$  nm). All dot surface points were calculated from the atomic coordinates of the enzyme. The basic protein atoms derived from the Protein Data Bank (file IRBX.pdb) are shown in space-filling format and in the usual CPK colours (C in light grey, O in red, N in light blue, and S in yellow) and the dot surface points are coloured black (in a few cases they are located slightly underneath the surface of the spheres). (b) Molecular dot surface points created by HYDCRYST for the contour of potential water points; these water points are located at a distance  $r_{\text{probe}}$  from the initially obtained surface points. (c) Model for the hydrated enzyme as obtained by HYDCRYST (applying  $V_w=0.0284$  nm<sup>3</sup>, Kuntz hydration numbers,  $f_K=1.5$ ); bound water molecules are displayed in green. Graphics were made with the program RASMOL<sup>93</sup>



**Figure 7** Space-filling models for anhydrous and hydrated LDH in two different views ( $90^\circ$  rotation around the  $x$  axis), together with the corresponding central slabs. (a) Molecular dot surface points created by the program SIMS for the anhydrous contour ( $d_{\text{dot}}=300 \text{ nm}^{-2}$ , and applying  $r_{\text{probe}}=0.140 \text{ nm}$ ). All dot surface points were calculated from the atomic coordinates of the enzyme. The basic protein atoms derived from the Protein Data Bank (file 6LDH.pdb1) are shown in space-filling format and in the usual CPK colours and the dot surface points are coloured black. (b) Molecular dot surface points created by HYDCRYST for the contour of potential water points; these water points are located at a distance  $r_{\text{probe}}$  from the initially obtained surface points. (c) Model for the hydrated enzyme as obtained by HYDCRYST (applying  $V_w=0.0220 \text{ nm}^3$ , Kuntz hydration numbers,  $f_K=4.0$ ); bound water molecules are displayed in green. (d) Space-filling models for anhydrous LDH at different stages of model reduction (for details see Table 9):  $N_b=8732$  (1), 4537 (2), 2161 (3), and 1316 (4). (e) Model for the hydrated enzyme, based on the reduced structure shown in (d) (4)