Published online November 10, 2005

Nucleic Acids Research, 2005, Vol. 33, No. 19 ***6361–6371***
doi:10.1093/nar/gki936

A coarse graining approach to determine nucleic acid structures from small angle neutron scattering profiles in solution

J. Zhou, S. Krueger[1] and S. K. Gregurick*

Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, 100 Hilltop Circle, Baltimore, MD 21250, USA and [1]NIST Center for Neutron Research, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8562, Gaithersburg, MD 20899-8562, USA

Received August 2, 2005; Revised September 19, 2005; Accepted October 12, 2005

ABSTRACT

We present a theoretical method to calculate the small angle neutron scattering profile of nucleic acid structures in solution. Our approach is sensitive to the sequence and the structure of the nucleic acid. In order to test our approach, we apply this method to the calculation of the experimental scattered intensity of the decamer d(CCAACGTTGG)$_2$ in H_2O. This sequence was specifically chosen for this study as it is believed to adopt a canonical B-form structure in 0.3 M NaCl. We find that not only will our methodology reproduce the experimental scattered intensity for this sequence, but our method will also discriminate between B-, A- and Z-form DNA. By studying the scattering profile of this structure in 0.5 and 1.0 M NaCl, we are also able to identify tetraplex and other similar oligomers formation and to model the complex using the experimental scattering data in conjunction with our methodology.

INTRODUCTION

Deoxyribonucleic acid (DNA) is a charged polyelectrolyte with a sequence dependent structure that plays an important role in protein–DNA recognition (1–4). The sequence of the nucleic acid will confer a preference for the conformation of the molecule in solution (A-, B-, . . ., Z-forms), which in turn will influence the flexibility and the rigidity of the structure. However, because a simple one-to-one relationship between protein recognition sites and the DNA sequence does not exist, perhaps a meta level of genetic information is encoded into the structural properties of the nucleic acid themselves (5,6). Therefore, the hypothesis that the DNA sequence alone can not distinguish between functional and non-functional nucleic acids, has recently led to their classification based on structural properties (7,8).

However, DNA is a highly charged polymer, and thus structure, dynamics and the corresponding biological properties can not be fully understood without a consideration of solvent hydration and the ionic atmosphere surrounding the molecule. For example, when the concentration of the cation is increased, it is possible the DNA will undergo a conformation transition from B to A or from B to Z, depending on the sequence (9–14). Complicating the matter is the ability of guanine rich sequences to self associate into higher ordered structures that can be composed of one, two or four strands (15–20). The ability of DNA to change conformation is an important component in DNA–protein interactions, while the ability of DNA to form tetraplexes has been linked to the aging mechanism and the inhibition of the telomerase enzyme.

In short, the study and characterization of nucleic acid structures, and their assemblies, lies at the heart of modern biophysics. Most of our understanding of the effects of mono and divalent cations interacting with DNA comes from molecular dynamics (MD) simulations (21–28), NMR studies (29,30), atomic resolution X-ray diffraction (31–33), spectroscopic (34) and small angle scattering experiments (35–41). MD provides an opportunity to study the short to intermediate [up to 60 ns (42)] time dynamics of the nucleic acid structures with the corresponding ions, yet the application of this method to provide global, millisecond and beyond dynamical information is still under investigation. In fact, the question of convergence, stability and sampling of DNA molecular dynamics trajectories has recently been addressed by the 'Ascona B DNA Consortium (ABC)' and is well presented in reference 28. In addition, the determination of nucleic acid structures using NMR techniques is difficult due to the lack of tertiary

*To whom correspondence should be addressed. Email: greguric@umbc.edu

The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint First Authors

contacts and NOE data. Crystallographic determination of structures, although powerful, does have limitations both in errors due to mobility and/or multiple conformations of a given structure (42,43). Small angle scattering, while still a lower resolution technique, has the advantage of determining overall molecular shape as a function of the experimental conditions. Until this time, scattering studies of nucleic acids has been limited by the inability to extract high resolution information, and as such remains a lesser applied technique.

In this article we present an approach to integrate high resolution crystallographic data with small angle neutron scattering (SANS) experiments in order to determine the structure of nucleic acids in solution. Previous work on the small angle neutron scattering of nucleic acids in solution has focused on a rigid (38,41) or semi-rigid rod (35,36,38) approximation to the DNA structure or has represented the molecule as a lower resolution helical structure (44). This work aims to bridge the gap between lower resolution SANS techniques and the atomic resolution of X-ray crystallographic, NMR or MD simulations. At present our method builds coarse graining models from higher resolution structures in order to predict theoretical SANS curve in solution. However a future goal is to extend this methodology to include simplistic MD, such that more of the configurational space available to the nucleic acids can be explored. This will open up the possibility to perhaps solve related structures in solution. The source code and examples for the current Coarse Graining method and SANS calculations are freely available upon request to greguric@umbc.edu.

THEORY

In SANS experiments, neutrons are scattered and measured preferentially from atomic nuclei. If one considers a well collimated beam of neutrons as a plane wave, then the momentum transferred between the initial and scattered wave is given by the formula:

$$Q = \frac{4\pi}{\lambda}\sin\theta, \qquad \mathbf{1}$$

where 2θ is the scattering angle and λ is the neutron wavelength. At a neutron wavelength of 5 Å, the corresponding angular region is $2\theta \approx 0.3^\circ - 5^\circ$, hence the term small angle neutron scattering. The scattered intensity, $I(Q)$, can be measured by counting the number of neutrons at each value of Q within a solid volume:

$$I(Q) = \left| \int_{V_0} \rho(r)\, e^{i\bar{Q}(r)\cdot\bar{r}} d^3r \right|^2, \qquad \mathbf{2}$$

where $\rho(r)$ is the scattering length density (SLD) and V_0 is the volume of the scatterer. Van Hove showed that this scattered intensity is simply proportional to a Fourier transform of the pair-wise distance distribution function $P(r)$ between all possible pairs of two scattering points (45):

$$I(Q) = 4\pi V_0 \int_0^{D_{\max}} P(r)\frac{\sin(Qr)}{Qr} dr. \qquad \mathbf{3}$$

The integral is carried out to value of $D_{\max}$ after which there is no significant scattering mass of the sample. If one considers a molecule in a solvent with fixed density ρ_0, by first subtracting the scattering due to the solvent, the resulting scattered intensity is approximately that of the isolated molecule, for all but the smallest of angles. Because our model assumes a heterogeneous scattering, e.g. each base will carry a specific scattering length, the distance distribution function in Equation 3 is weighted by the base specific scattering length density, ρ_i as follows:

$$P(r) = \rho_i \rho_j P(r_i - r_j) \qquad \mathbf{4}$$

Previous methods for DNA analysis relied on the calculation of a structure factor (35,38–40) in order to calculate $I(Q)$ as in Equation 2. In our work, we first calculate a distance distribution function and then obtain $I(Q)$ using Equation 3. Thus, ours is a real-space approach, making use of the structural information from X-ray crystallography, NMR and MD simulations. We make a coarse grain approximation of the structure for the scattering calculation in order to include both the sequence specificity of the nucleic acid in question and the geometry that the polymer would naturally adopt. An analogous coarse graining method was developed by Gregurick and co-workers (46,47) for the study of the small angle neutron scattering of proteins in solution. This type of approach has proven highly useful when studying complexes of macromolecules.

We have also added the ability to determine the nucleic acid scattering profile in mixed solvents of D_2O/H_2O based on the work of Jacrot (48). In the case of mixed solvents, the scattering length b_i for any given base i, is determined by the following formula:

$$b_i = b_i^{H_2O} + X * N * (b_D - b_H) \qquad \mathbf{5}$$

where X is the percentage of D_2O in solution, N is the number of exchangeable protons and b_D and b_H are the scattering lengths of deuterium and hydrogen, respectively. The number of exchangeables for each base is as follows: adenine 2, guanine 3, cytosine 2 and thymine 1. A similar equation can be written if the nucleic acid is fully deuterated and studied in a mixture of D_2O/H_2O as follows:

$$b_i = b_i^{D_2O} + X * N * (b_H - b_D) \qquad \mathbf{6}$$

The values of the scattering length for fully hydrogenated and deuterated bases can be found in Table 2.

COARSE GRAIN MODEL

In our study we have developed a method to correlate the three-dimensional structure of a nucleic acid in solution to the calculated SANS scattered intensity. We do so by first developing a course grain approach to the molecular structure. The coarse model is built by stacking a number of smaller unit structures superimposed onto every nucleotide. The unit models are then filled with an adequate number of scattering points to match the scattering density for each specific base in the sequence.

Each unit model is composed of a cylindrical ellipsoid and a sphere to simulate the shape of the nucleoside and phosphate

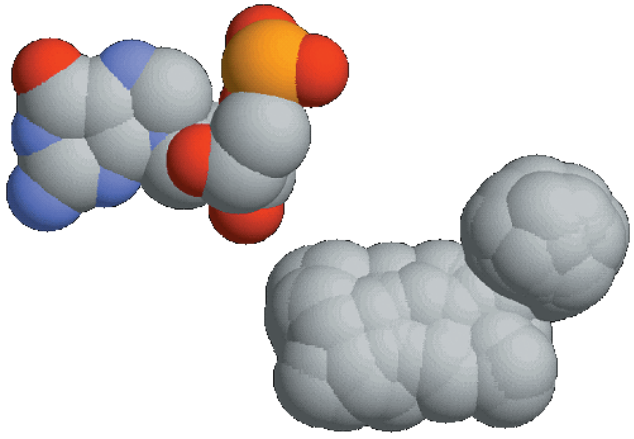

Figure 1. The crystal structure of a nucleotide represented in a space filling model (left) and the unit model of this nucleotide generated through our coarse graining program, XTAL2SAS (right).

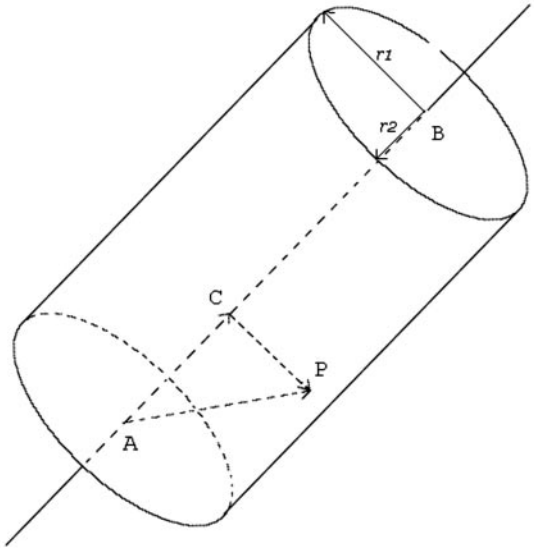

Figure 2. Geometric illustration of the cylindrical ellipsoid used to represent a course grained nucleoside. The parameters are: the position of the two end points (*A* and *B*) of the cylinder axis, length of the long (*r*1) and the short (*r*2) axis of the ellipsoidal cross section.

group, respectively (Figure 1). It is straightforward to create a sphere to represent the coarse grained phosphate group. The center of the sphere is defined by the location of the phosphate 'P' atom and the radius of the sphere is defined by the distance between the 'P' and the 'OP1' atoms. The resulting coarse grained sphere is filled with random points to satisfy the criteria that the distance to the sphere center is smaller or equal to the radius. The sphere is filled with random points until the desired point density is reached.

In order to define a unique cylindrical ellipsoid that overlaps each nucleoside in the sequence, the following parameters are required (Figure 2): the position of the two end points (*A* and *B*) of the cylinder axis, length of the long (*r*1) and the short (*r*2) axis of the ellipsoidal cross section. The length of the long radius is defined as the longest distance between an atom on the surface and the cylinder axis. We defined the length of the short axis to be the size of an atom, 1.0 Å. Table 1 is a summary of the coarse graining parameters required for each type of nucleoside.

Table 1. Parameters for our coarse graining approach

Nucleotide type	A	T	G	C	U
Cylindrical ellipsoid					
End-point 1 (*A*)	N1	C1* O2	N1	N3 C4	C4
End-point 2 (*B*)	C4* C3*	C5M O4	C4* C3*	C4* C3*	C4*
Long radius (*r*1)	N6	O2	O6	O2	O2
Sphere					
Center	P	P	P	P	P
Radius	OP1	OP1	OP1	OP1	OP1

Table 2. Nucleoside specific scattering parameters (48)

Base	Volume (Å^3)	Number of points	Scattering length ($\times 10^{-12}$ cm) H_2O	Scattering length ($\times 10^{-12}$ cm) D_2O
C	115	150	8.68	10.77
T	116	150	8.61	9.65
G	130	150	11.23	14.35
A	134	150	10.65	12.73
Phosphate nucleoside	14	50	Assigned scattering length of corresponding base	Assigned scattering length of corresponding base

After the position and the shape of the cylindrical ellipsoid has been determined, random points are generated to fill the model, until the desired point density for each specific base is reached (Table 2). The criteria for any random point *P* to lie within the volume of the cylindrical ellipsoid is (Figure 2):

(i) Point *C*, which is the projection of point *P* on the axis *AB*, has to be between point *A* and point *B*.
(ii) Given *d*1 and *d*2 as the length of the projection of vector $C\vec{P}$ onto *r*1 and *r*2 separately, the following equation has to be satisfied:

$$\frac{d1^2}{r1^2} + \frac{d2^2}{r2^2} \leqslant 1 \qquad \mathbf{7}$$

A coarse grained model for the nucleic acid is generated by building unit models for every nucleotide in the PDB file or from the MD trajectory data. To make use of dynamics trajectory data, a set of snapshots by time needs to be exported and saved in PDB format. A coarse grained model is then generated for each snapshot. Figure 3 is an illustration of this method for the DNA structure corresponding to 5DNB.pdb.

In order to judge the quality of our coarse graining method to represent the structure of the molecule in solution, we calculate a χ^2 distance between the theoretical $I(Q)$ (Equation 3) and the experimental $I(Q)$. The χ^2 distribution is defined as:

$$\chi^2 = \frac{1}{N}\sum_i w_i \left[I^{\text{expt}}(Q_i) - I^{\text{model}}(Q_i)\right]^2 \qquad \mathbf{8}$$

where N is the number of degrees of freedom when m data points are fitted with a model involving n adjustable parameters and w_i is the weight. In this case each w_i is taken to be 1.

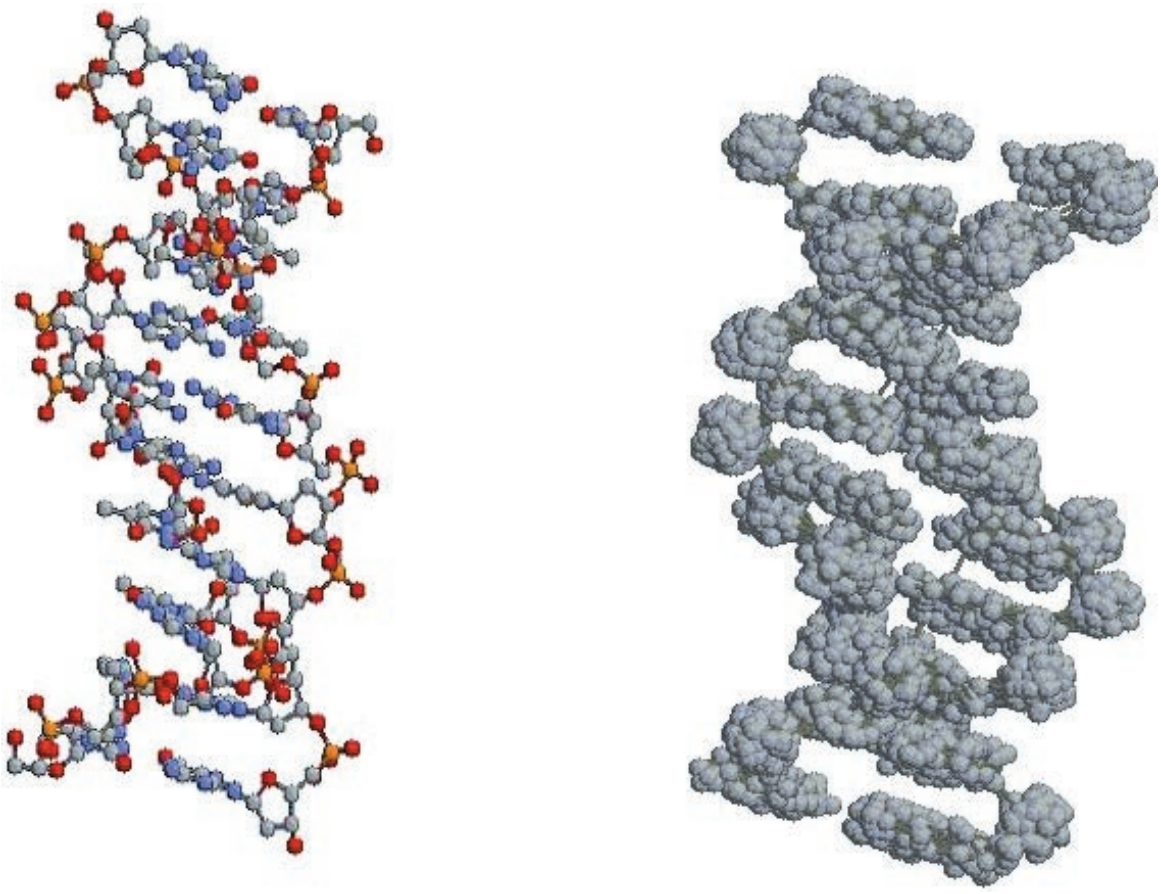

Figure 3. The crystal structure of the DNA duplex, 5DNB (PDB ID), represented in ball and stick (left) and the coarse model generated through the program, XTAL2SAS, from the original PDB structure of 5DNB (right).

The sum in Equation (68) is over all m data points. We also calculate a regression coefficient, R^2, defined as:

$$R^2 = \frac{\sum_i [I^{\text{expt}}(Q_i) - \overline{I(Q_i)}]^2}{\sum_i [I^{\text{expt}}(Q_i) - \overline{I(Q_i)}]^2 + \sum_i [I^{\text{expt}}(Q_i) - I^{\text{model}}(Q_i)]^2} \quad \mathbf{9}$$

where $\overline{I(Q_i)}$is the average experimental intensity.

MATERIALS AND METHODS

For this study we have investigated the theoretical and experimental small angle neutron scattering of the duplex DNA decamer, d(CCAACGTTGG)$_2$, in a series of salt concentrations ranging from 0.1 to 1.0 M. This particular decamer was chosen as a candidate duplex for our SANS analysis based on previous MD simulations which indicated that this sequence is stable in solution for up to 25 ns and remains in the B configuration (11,49–51). The structure of d(CCAACGTTGG)$_2$, (PDB ID: 5DNB) was first crystallized by Prive *et al*. (52). The PDB file of 5DNB.pdb contains the coordinates of the crystallographic asymmetric unit corresponding to one strand of the duplex. A second strand of the duplex was added to the unit cell by translating the initial duplex according to the guidelines published by Prive *et al*. (52). The PDB file with the coordinates of both strands can be directly downloaded from www.pdb.org.

SANS experiment

The self-complementary 10mer single-strand DNA sequence 5′-CCAACGTTGG-3′ was purchased and purified to high-performance liquid chromatography (HPLC) Level I (90–95 mol %) from Oligos, etc. Inc. (Wilsonville, OR)*. The duplex DNA decamer was formed by re-hydrating the DNA at 10 mg/ml concentration in 0.1 and 1 M NaCl solutions. The solutions were first heated to 42°C for 15 min, to minimize self-association of the DNA strands, and then allowed to cool for 3 h in order to form the double-stranded decamer. The solution was heated to 42°C in order to work above the melting temperature of this DNA sequence in 1 M NaCl (∼38°C). The samples were then placed in the refrigerator at 6°C overnight. SANS measurements were initially performed at 15°C on a 10 mg/ml DNA solution with 0.1 M NaCl, in order to remain below the melting temperature of the sample under these salt conditions. This minimized the chances of observing single-strand behavior in the data. The solutions were then diluted to 5 mg/ml DNA concentration and mixed, as necessary, to obtain 1, 0.5, 0.3 and 0.1 M NaCl solutions with the DNA. SANS measurements were then performed on these four samples at 15°C.

The SANS measurements were performed on the NG7 30 m SANS instrument at the National Institute of Standards and Technology Center for Neutron Research in Gaithersburg, MD (53). A neutron wavelength of $\lambda = 5.5$ Å with a wavelength

spread, $\Delta\lambda/\lambda$ of 0.11 was used for the measurements. The source and sample apertures were 5.0 and 1.27 cm, respectively. Neutrons were detected on a 64.0×64.0 cm two-dimensional position-sensitive detector with 0.5 cm resolution. A sample to detector distance of 1.5 m and a source to sample distance of 5.47 m were used. The center of the detector was offset by 15.0 cm to obtain a range of momentum transfer, Q, values between 0.029 $Å^{-1}$ and 0.39 $Å^{-1}$, where $Q = 4\pi\sin(\theta)/\lambda$ and 2θ is the scattering angle.

The SANS data were normalized to a common monitor count and corrected for empty cell counts, ambient room background counts and non-uniform detector response. Data were placed on an absolute scale by normalizing the scattered intensity to the incident beam flux. The two-dimensional data were then radially averaged to produce $I(Q)$ versus Q curves. The one-dimensional scattered intensities from the samples were then corrected for buffer scattering and incoherent scattering from hydrogen in the samples. After these final data correction procedures, usable data were obtained in the range 0.029 $Å^{-1} \leqslant Q \leqslant 0.3$ $Å^{-1}$.

RESULTS AND DISCUSSION

The experimental scattering curves for the DNA duplex sequence $d(CCAACGTTGG)_2$, are illustrated in Figure 4,

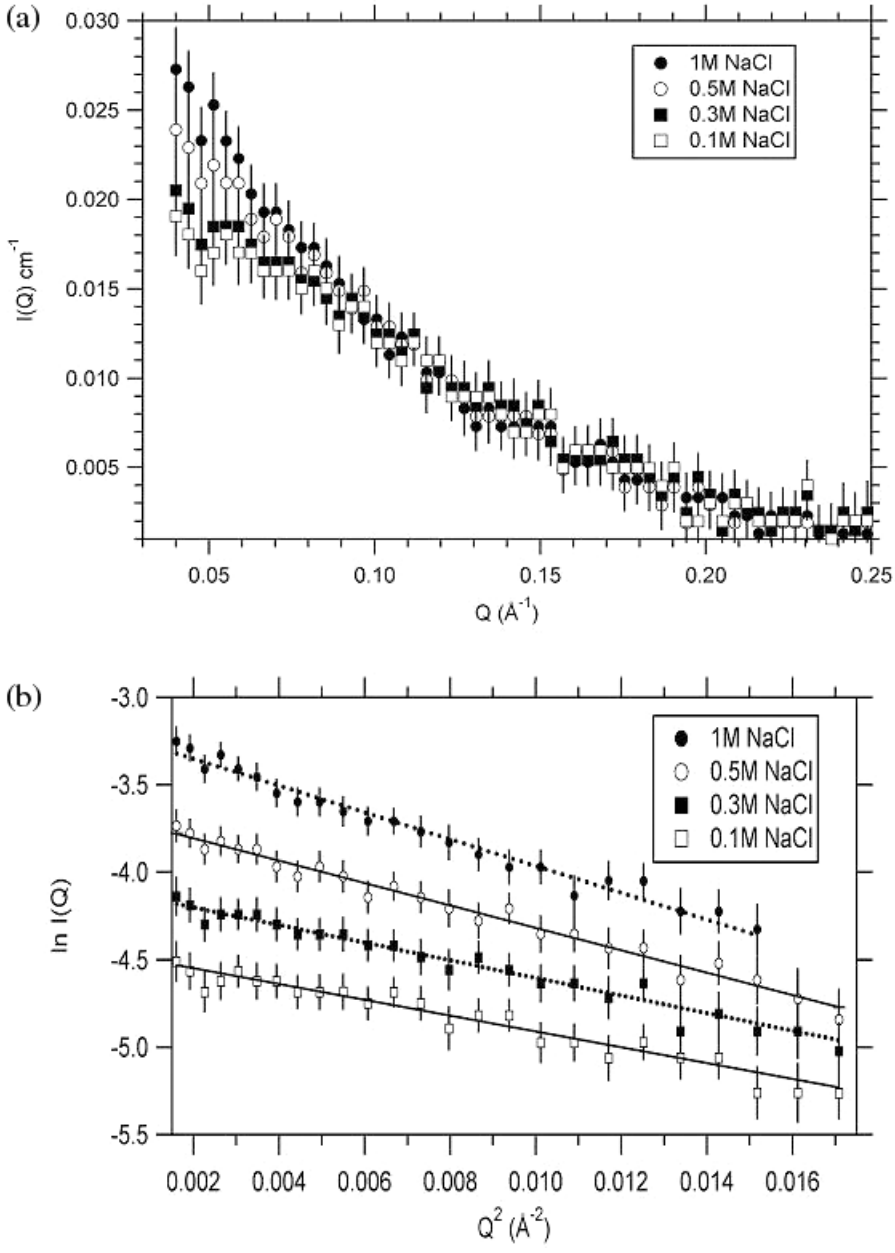

Figure 4. (**a**) The experimentally measured neutron scattered intensity, $I(Q)$, for the decamer $d(CCAACGTTGG)_2$ in H_2O, at four different Na^+ concentrations: 0.1 M (open squares), 0.3 M (closed squares), 0.5 M (open circles) and 1 M (closed circle). (**b**) Guinier analysis for the decamer $d(CCAACGTTGG)_2$ in H_2O, at four different Na^+ concentrations: 0.1 M (open squares), 0.3 M (closed squares), 0.5 M (open circles) and 1 M (closed circles).

where the open squares represent the 0.1 M solution, the filled squares represent the 0.3 M solution, the open circles represents the 0.5 M solution and the filled circles represent the 1.0 M solution. The DNA concentration was 5 mg/ml in all four cases. This sequence was specifically chosen as it is believed to adopt a canonical B structure at 0.3 M NaCl. From Figure 4a it is clear that there is a upturn in the scattered intensities at very small Q equal to 0.02 $Å^{-1}$, for all concentrations studied. Although this has been observed in other polyelectrolyte scattering experiments (36,54–58), the physical interpretation of this effect remains problematic. In the small angle neutron scattering experiments of Borsali *et al*. (36) this upturn was attributed to a close packing of the nucleic acid helices. Recently Shibano and co-workers (54) have studied the intra and intermolecular scattering functions of sodium polystyrene-sulfonates in solution. These authors also find that the neutron scattered intensity depends on the salt concentration. The addition of salt is primarily affecting the intermolecular scattering function, as at higher salts the Debye electrostatic screening is increased. Shibano *et al*. suggests to account for the intramolecular scattering function by performing the SANS experiment in both hydrogenated and deuterated solvents. By varying the deuteration in the samples studied, the intermolecular scattering function can be removed. Because the concentration of the DNA we studied is relatively low, the effect of the added salt is minimized at all but the smallest region of Q. Therefore, all of our analysis is preformed for values of Q from 0.05 to 0.3 $Å^{-1}$.

In order to assess the effect of the salt concentration on the scattering curves, we have preformed a Guinier analysis (48) of the scattering intensities at small Q (48) (Figure 4b). We find that the radius of gyration (Rg) increases as the salt concentration increases from 0.1 to 1.0 M (Table 3). Wang and Bloomfield also observed a salt dependency on the small angle X-ray scattering profiles of nucleic acid solutions and interpreted this result to indicate that the effective diameter of the DNA is dependent on the ionic atmosphere surrounding the nucleic acid (41). In the case of small angle neutron scattering, the Na^+ and Cl atoms are not believed to contribute to the scattered intensities (38). Thus, our observed increase in the Rg at the higher salt concentrations is likely due to a change in the structure, or to an assembly of the nucleic acids.

To illustrate the power behind our methodology, we plot in Figure 5 the experimental scattering profiles from the 0.3 M NaCl sample at 5 mg/ml and the 0.1 M NaCl sample at 10 mg/ml and compare these with the scattered intensities calculated from the B DNA crystal structure, 5DNB, using our coarse graining approach. On the same figure, we also plot the calculated scattered intensities of the same DNA sequence with an A and Z conformation. The A and Z structures were constructed using the program, 3DNA, by Lu and Olson (59) and are illustrated in Figure 6. We note that the calculated data are plotted on an absolute scale, with $I(0)$ set to match the concentration of the experimental duplex B DNA. The fits of the ideal B, Z and A DNA structures to the scattered intensity of the 0.3 M NaCl sample yields R^2 and χ^2 values of 0.89, 0.38 (B-form DNA), 0.57, 0.91 (Z-form DNA) and 0.47, 1.06 (A-form DNA), respectively. The R^2 and χ^2 values were calculated using Equations 8 and 9. Similar fits of the model data to the 10 mg/ml 0.1 M NaCl sample, in which the errors on the scattered intensity values are smaller, produced χ^2 values of 0.58 (B DNA), 1.71 (A DNA) and 1.93 (Z DNA). For both the 0.1 M NaCl and 0.3 M NaCl solutions, and in all regions of Q analyzed, the B-form of this nucleic acid fits to the experimental scattering data very well, whereas the Z-form fits only at the lower Q-values and the A from fits only at the higher Q-values. This is not surprising, as the sequence was chosen because it is believed to adopt a canonical B DNA conformation in 0.3 M NaCl, based on previous molecular dynamic simulations (49,50).

Another equally valid but computationally distinctive approach that is used to calculate the small angle scattering intensities of biological molecules in solution is the multipole expansion method of Svergun and co-workers (60–62). In this method the scattering intensity is calculated from the atomic structure factors, which are expanded in spherical harmonics. In this way, the molecule is represented as an envelope with an additional hydration shell of 0.3 nm in thickness. This method which is implemented in the programs CRYSON and CRYSOL (X-ray scattering), has been widely used to calculate the small angle scattering intensity of proteins in solution. Recently Nöllmann *et al*. (63) have used CRYSON in their SANS study of the oligomerization properties of Tn3R transposons. We have calculated a scattering intensity for the crystal structure of 5DNB using the CRYSON program and compare this with our coarse graining calculation in Figure 5C for 5DNB in 0.3 M NaCl. It is clear from this figure that in most regions of Q, the coarse graining method is superior to that of CRYSON, particularly at higher values. However, in our analysis we have used the default values of CRYSON and this may be inappropriate for the study of nucleic acids. We note that CRYSON adds an overall hydration layer to the DNA, whereas our method neglects hydration. We believe that an envelope of hydration may not be appropriate for nucleic acids which are thought to hydrate primarily in the minor groove (spine of hydration) (27,64–66) and around the phosphate group (67). Our current coarse graining method is an alternative to CRYSON and is strictly useful for the calculation of the small angle neutron scattering profile of nucleic acids in solution.

More challenging is the analysis of the scattering in 0.5 M and 1.0 M NaCl solutions. Neither the A-, B- nor Z-form of this DNA sequence would fit to the experimental scattering data. Moreover, the Guinier analysis suggests that the Rg and the concentration in both of these cases is not indicative of a

Table 3. Guinier analysis of the experimental and theoretical scattering intensities

Salt concentration	Rg (Å)	$I(0)$ cm^{-1}
0.1 M NaCl, 10 mg/ml	12.0 ± 0.05	0.0446 ± 0.0005
0.1 M NaCl, 5 mg/ml	12 ± 1	0.021 ± 0.001
0.3 M NaCl, 5 mg/ml	12.5 ± 1	0.022 ± 0.001
0.5 M NaCl, 5 mg/ml	13 ± 1	0.024 ± 0.001
1.0 M NaCl, 5 mg/ml	15 ± 1	0.027 ± 0.001
B DNA (fit to 0.3 M NaCl)	11.7	0.02, $\chi^2 = 0.38$
B DNA (fit to 0.1 M NaCl)	11.7	0.02, $\chi^2 = 0.58$
A DNA (fit to 0.3 M NaCl)	11.1	0.02, $\chi^2 = 1.06$
A DNA (fit to 0.1 M NaCl)	11.1	0.02, $\chi^2 = 1.71$
Z DNA (fit to 0.3 M NaCl)	12.2	0.02, $\chi^2 = 0.91$
Z DNA (fit to 0.1 M NaCl)	12.2	0.02, $\chi^2 = 1.93$
Tetraplex model (fit to 1.0 M NaCl)	13.8	0.026, $\chi^2 = 0.54$

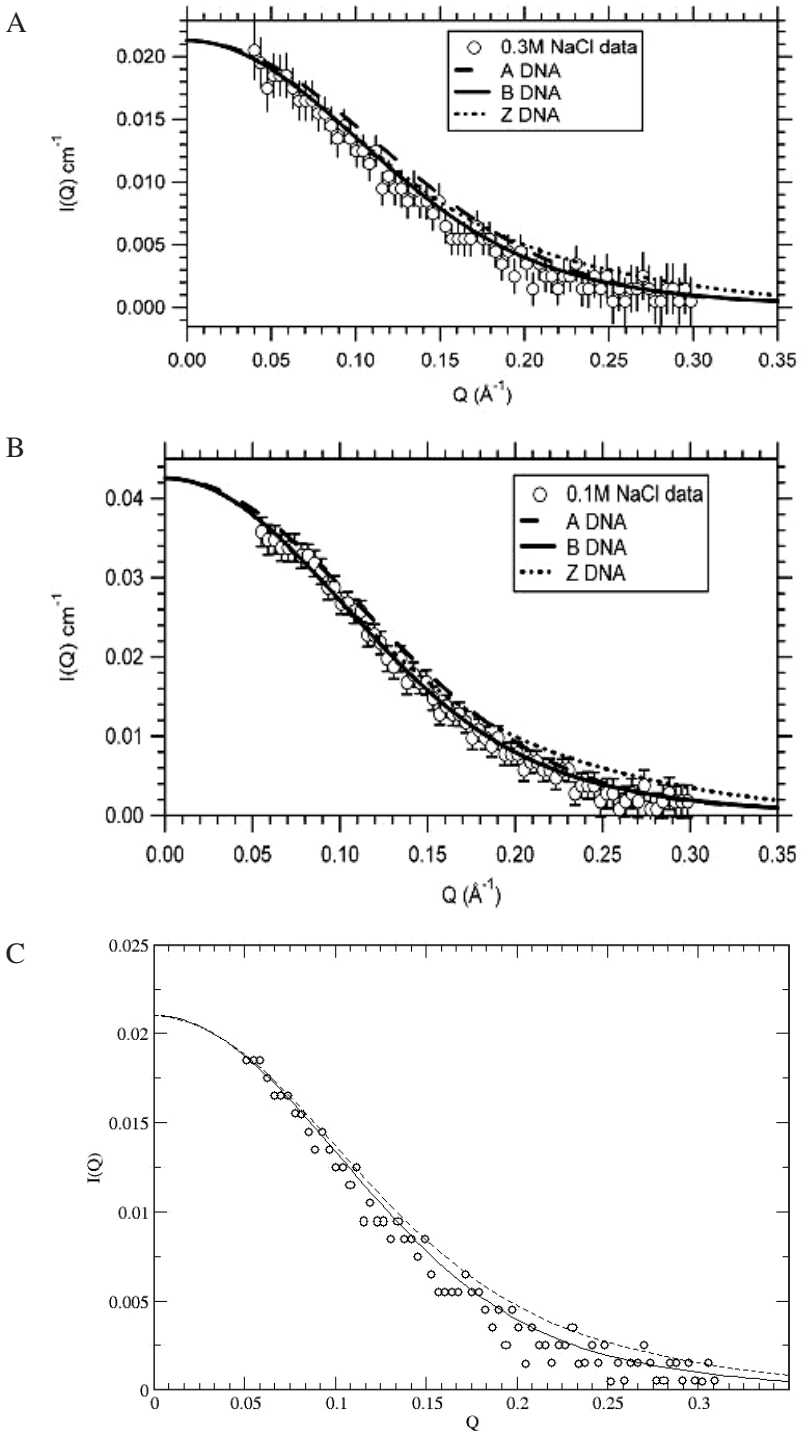

Figure 5. (**A**) A comparison between the experimentally measured $I(Q)$ at 0.3 M NaCl (circles) and the $I(Q)$ calculated for the crystal structure of 5DNB (solid line), compared to the A (solid line) and Z (dash-dotted line) forms of this nucleic acid. (**B**) A comparison between the experimentally measured $I(Q)$ at 0.1 M NaCl (circles) and the $I(Q)$ calculated for the crystal structure of 5DNB (solid line), compared to the A (solid, line) and Z (dash-dotted line) forms of this nucleic acid. (**C**) A comparison between the experimentally measured $I(Q)$ at 0.3 M NaCl (circles) and the $I(Q)$ calculated for the crystal structure of 5DNB using the current coarse graining method (solid line) and that of CRYSON (dashed line). Error bars have been omitted for clarity.

single duplex in solution (Table 3). Therefore a tetraplex (quadruplex) model was created, based on the crystal structure of the DNA quadruuplex d(TGGGGT) by Caceres *et al*. (68). Because we could not append the crystal structure of Caceres, the coarse model is based on this exact crystal structure and is illustrated in Figure 7. For this model, the theoretical scattering curve was calculated using our methodology and plotted in Figure 8. Again, the calculated data are on an absolute scale, with the $I(0)$ set to match that of the experimental scattering curve. The Rg value of this model was found to be 13.8 Å and

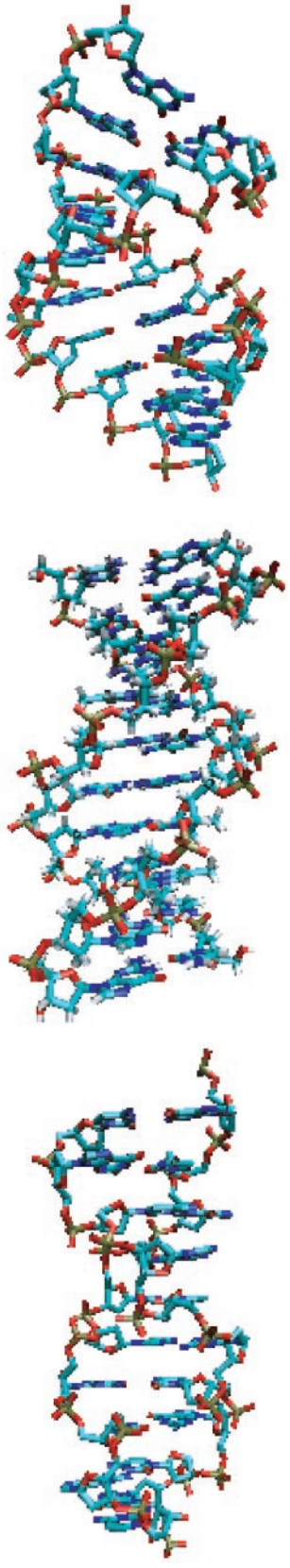

Figure 6. The A (upper), B (middle) and Z (lower) conformations of the decamer d(CCAACGTTGG)$_2$. This figure was created using the programs 3DNA (for A- and Z-forms) and illustrated using VMD.

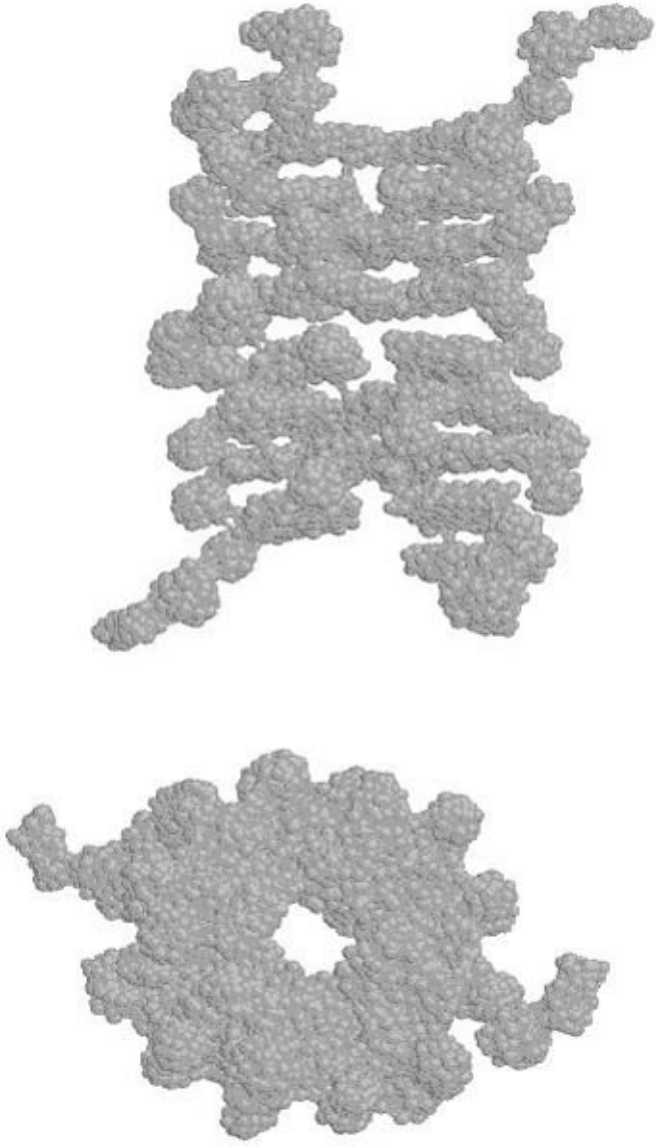

Figure 7. Coarse grained model of a tetraplex, based on the structure of 1S45.pdb (68).

the shape of the scattered intensity matched the 1 M NaCl data well, with a χ^2 value of 0.54. However, the molecular weight (Mw) of the tetraplex formed in our DNA samples would be twice that of the decamer. Since $I(0)$ is proportional to Mw (48) this implies that the measured $I(0)$ value would have to increase by a factor of two if the solution consisted of tetraplexes alone. Table 3 shows that our measured $I(0)$ values increase by a smaller amount, indicating a possible mixture of B DNA (decamer) and tetraplex DNA in solutions of 0.5 M NaCl and above.

To test this hypothesis, we modeled the scattered intensities, on an absolute scale, for comparison to the measured data (Figure 8). It was found that the 0.5 M NaCl data fit reasonably well to a model constructed assuming a mixture of B duplex DNA and tetraplex DNA in the solution in the ratio of 85:15 duplex:tetraplex. This particular ratio was chosen to match the observed change in $I(0)$ between the 0.5 M NaCl sample and

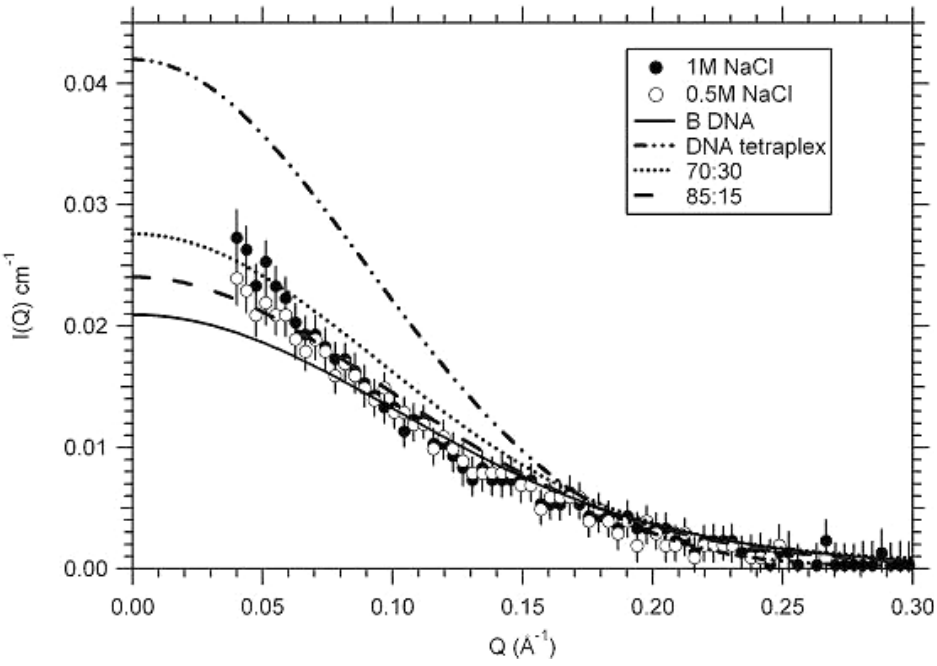

Figure 8. A comparison between the experimentally measured $I(Q)$ at 0.5 M NaCl (open circles), 1.0 M NaCl (closed circles) and the $I(Q)$ calculated from the B DNA duplex (solid line) model, the tetraplex model (dash-dotted line) and mixtures of the two models in the duplex:tetraplex ratios of 85:15 (dashed line) and 70:30 (dotted line).

the 0.3 M NaCl sample. For the 1 M NaCl sample, the change in $I(0)$ suggests a 70:30 duplex:tetraplex solution. However, the 1 M NaCl data do not fit a model of 70:30 duplex:tetraplex in solution. In fact, the data show evidence of slightly larger oligomeric structures. Because this is only a coarse grained model, we are now investigating further the stability of a tetraplex for this sequence by using molecular dynamics simulations.

DISCUSSION

We present a theoretical method to calculate the small angle neutron scattering profile of a nucleic acid structure in solution. Our method goes beyond a rigid or semi-rigid rod approximation and calculates a theoretical scattered intensity based on a high resolution structure. Our method utilizes a coarse graining approach for the calculation of the intensity, by way of a Fourier transform of the distance distribution function between all scattering points in the model. We apply this method for the calculation of the experimentally determined scattered intensity of the decamer $d(CCAACGTTGG)_2$ in H_2O with varying amounts of NaCl between 0.1 and 1 M. This sequence was specifically chosen for this study as it is believed to adopt a canonical B-form structure in 0.3 M NaCl. We find that this is the case and, moreover, that our method will reproduce the experimental scattered intensity with an R^2 value equal to 0.89. Because our methodology is valid for any structure with atomic coordinates, we also illustrate that neither the Z- ($R^2 = 0.57$) nor the A-form ($R^2 = 0.47$) structures will reproduce the experimental data. The value and usefulness of our method however, is not necessarily in just discriminating B from A and Z DNA in solution, but to model larger structures, such as Ribozymes, RNA and DNA bound to proteins. As an example of this, we modeled the structure at the higher salt concentrations. We constructed a coarse grain model of a tetraplex, based on the crystal structure of the tetrad d(TGGGGT) by Caceres *et al*. (68). It was found that the 0.5 M NaCl data fit reasonably well to our model assuming a mixture of B duplex DNA and tetraplex DNA in solution in the ratio of 85:15 duplex:tetraplex. The 1 M NaCl data show evidence of slightly larger oligomeric structures and were difficult to model using a similar mixed solution model. In conclusion, when NMR or crystallographic data are available, this method can be used to determine a model small angle neutron scattering curve. Thus, this is an advance of previous scattering calculations which relied on low resolution shapes such as a rigid rod, semi-rigid rod or a simplistic helix approximation to model nucleic acid structures in solution. We note and caution previously determined high resolution structures are required as an input for the coarse graining method. A non-trivial extension would be to extend this method to include coarse graining molecular dynamics in order to search for new configurations. We are currently undertaking this possibility.

ACKNOWLEDGEMENTS

This material is based upon activities supported by the National Science Foundation under agreement DMR-9423101. We acknowledge the support the National Institute of Standards and Technology, US Department of Commerce, in providing the neutron facilities used in this work. Financial support was also provided to S.K.G. from the University of Maryland, Baltimore County (UMBC) in the form of a Special Research Initiative Award and the ADVANCE program,

grant SBE-0244880 (NSF). S.K.G. would like to thank Dr A. MacKerell (UMB) for insightful discussions regarding nucleic acid conformations under differing salt conditions and Mr Hailiang Zhang for incorporating deuterium/hydrogen exchange into the coarse-graining software. *Brand names are stated for clarity only and their use does not imply endorsement by NIST. Funding to pay the Open Access publication charges for this article was provided by NSF.

Conflict of interest statement. None declared.

REFERENCES

1. Drew,H.R. and Travers,A.A. (1985) Structural junctions in DNA: the influence of flanking sequence on nuclease digestion specificities. *Nucleic Acids Res*., **13**, 4445–4467.
2. Burley,S.K. (1996) The TATA box binding protein. *Curr. Opin. Struct. Biol*., **6**, 69–75.
3. Sarai,A. and Kono,H. (2005) Protein–DNA recognition patterns and predictions. *Annu. Rev. Biophys. Biomol. Struct*., **34**, 379–393.
4. Sarai,A., Siebers,J., Selvaraj,S., Gromiha,M.M. and Kono,H. (2005) Integration of bioinformatics and computational biology to understand protein–DNA recognition mechanism. *J. Bioinform. Comput. Biol*., **3**, 169–183.
5. Brennan,R.G. and Matthews,B.W. (1989) Structural basis of DNA–protein recognition. *Trends Biochem. Sci*., **14**, 286–290.
6. Albright,R.A. and Matthews,B.W. (1998) How Cro and lambda-repressor distinguish between operators: the structural basis underlying a genetic switch. *Proc Natl Acad Sci USA*, **95**, 3431–3436.
7. Gardiner,E.J., Hunter,C.A., Packer,M.J., Palmer,D.S. and Willett,P. (2003) Sequence-dependent DNA structure: a database of octamer structural parameters. *J. Mol. Biol*., **332**, 1025–1035.
8. Gardiner,E.J., Hunter,C.A., Lu,X.-J. and Willett,P. (2004) A structural similarity analysis of double-helical DNA. *J. Mol. Biol*., **343**, 879–889.
9. Lim,W. and Fen,Y.P. (2005) Applying the stochastic difference equation to DNA conformational transitions: a study of B-Z adn B-Z DNA transitions. *Biopolymers*, **78**, 107–120.
10. Vargason,J.M., Henderson,K. and Ho,P.S. (2001) A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. USA*, **98**, 7265–7270.
11. Cheatham,T.E.,III and Kollman,P.A. (1996) Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol*., **259**, 434–444.
12. Pastor,N. (2005) The B- to A- transition and the reorganization of solvent at the DNA surface. *Biophys. J*., **88**, 3262–3275.
13. Ng,H.-L., Kopka,M.L. and Dickerson,R.E. (2000) The structure of a stable intermediate in the A to B DNA helix transition. *Proc. Natl. Acad. Sci. USA*, **97**, 2035–2039.
14. Jose,D. and Porschke,D. (2004) Dynamics of the B-A transition of DNA double helices. *Nucleic Acids Res*., **32**, 2251–2258.
15. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
16. Egholm,M., Nielson,P.E., Buchardt,O. and Berg,R.H. (1992) Recognition of guanine and adenine in DNA by cytosine and thymine containing peptide nucleic acids (PNA). *J. Am. Chem. Soc*., **114**, 9677–9678.
17. Poon,K. and Macgregor,R.B.,Jr (2000) Formation and structural determinants of multi-stranded guanine-rich DNA complexes. *Biophys. Chem*., 205–216.
18. Yafe,A., Etzioni,S., Weisman-Shomer,P. and Fry,M. (2005) Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Res*., **33**, 2887–2900.
19. Simonsson,T. and Sjoback,R. (1999) DNA tetraplex formation studied with fluorescence resonance. *J. Biol. Chem*., **274**, 17379–17383.
20. Lane,A.N. and Jenkins,T.C. (2001) Structures and properties of multi-stranded nucleic acids. *Curr. Org. Chem*., **5**, 845–869.
21. Varnai,P. and Zakrzewska,K. (2004) DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res*., **32**, 4269–4280.
22. Young,M.A., Jayaram,A.B. and Beveridge,D.L. (1997) Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc*., **119**, 59–69.
23. McConnell,K.J. and Beveridge,D.L. (2000) DNA structure: what's in charge? *J. Mol. Biol*., **304**, 803–820.
24. Feig,M. and Pettitt,B.M. (1999) Sodium and chlorine ions as part of the DNA solvation shell. *Biophys. J*., **77**, 1769–1781.
25. Egli,M. (2002) DNA-cation interactions: quo vadis? *Chem. Biol*., **9**, 277–286.
26. Rueda,M., Cubero,E., Laughton,C.A. and Orozco,M. (2004) Exploring the counterion atmosphere around DNA: what can be learned from molecular dynamics simulations. *Biophys. J*., **87**, 800–811.
27. Halle,B. and Denisov,V.P. (1998) Water and monovalent ions in the minor groove of B-DNA oligonucleotides as seen by NMR. *Biopolymers*, **48**, 210–233.
28. Beveridge,D.L., Barreio,G., Byun,K.S., Case,D.A., Cheatham,T.E.,III, Dixit,S.B., Giudice,E., Lankas,F., Lavery,R., Maddocks,J.H. *et al*. (2004) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results of d(CpG) steps. *Biophys. J*., **87**, 3799–3813.
29. Hud,N.V. and Polak,M. (2001) DNA–cation interactions: the major and minor grooves are flexible ionophores. *Curr. Opin. Struct. Biol*., **11**, 293–301.
30. Denisov,V.P. and Halle,B. (2000) Sequence-specific binding of counterions to B-DNA. *Proc. Natl. Acad. Sci. USA*, **97**, 629–633.
31. Soler-Lopez,M., Malinina,L., Liu,J., Huynh-Dinh,T. and Subirana,J.A. (1999) Water and ions in a high resolution structure of B-DNA. *J. Biol. Chem*., **274**, 23683–23686.
32. Chiu,T.K., Kaczor-Grzeskowiak,M. and Dickerson,R.E. (1999) Absence of minor groove monovalent cations in the crosslinked dodecamer C-G-C-G-A-A*-T-T-C-G-C-G. *J. Mol. Biol*., **292**, 589–608.
33. Tereshko,V., Wilds,C.J., Minasov,G., Prakash,T.P., Maier,M.A., Howard,A., Wawrzak,Z., Manoharan,M. and Egli,M. (2001) Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. *Nucleic Acids Res*., **29**, 1208–1215.
34. Kankia,B.J. and Marky,L.A. (1999) DNA, RNA and DNA/RNA oligomer duplexes: a comparative study of their stability, heat, hydration, and Mg^{2+} binding properties. *J. Phys. Chem. B*., **103**, 8759–8767.
35. Hammermann,M., Brun,N., Klenin,K.V., May,R., Toth,K. and Langowski,J. (1998) Salt dependent DNA superhelix diameter studied by small angle neutron scattering measurements and Monte Carlo simulations. *Biophys. J*., **75**, 3057–3063.
36. Borsali,R., Nguyen,H. and Pecora,R. (1998) Small angle neutron scattering and dynamic light scattering from a polyelectrolyte solution: DNA. *Macromolecules*, **31**, 1548–1555.
37. Bastos,M., Castro,V., Mrevlishvili,G. and Teixeira,J. (2004) Hydration of ds-DNA and ss-DNA by neutron quasielastic scattering. *Biophys. J*., **86**, 3822–3827.
38. Zakharova,S.S., Egelhaaf,S.U., Bhuiyan,L.B., Outhwaite,C.W., Bratko,D. and van der Maarel,J.R.C. (1999) Multivalent ion-DNA interaction: neutron scattering estimates of polyamine distribution. *J. Chem. Phys*., **111**, 10706–10716.
39. Zakharova,S.S., Jesse,W., Backendorf,C., Egelhaaf,S.U., Lapp,A. and van der Maarel,J.R.C. (2002) Dimensions of plectonemically supercoiled DNA. *Biophys. J*., **83**, 1106–1118.
40. Morfin,I., Horkay,F., Basser,P.J., Bley,F., Hecht,A.-M., Rochas,C. and Geissler,E. (2004) Adsorption of divalent cations on DNA. *Biophys. J*., **87**, 2897–2904.
41. Wang,L. and Bloomfield,V.A. (1991) Small angle X-ray scattering of semidilute rodlike DNA solutions: polyelectrolyte behavior. *Macromolecules*, **24**, 5791–5795.
42. Acharya,K.R. and Lloyd,M.D. (2005) The advantage and limitations of protein crystal structures. *Trends Pharmacol. Sci*., **26**, 10–14.
43. DePristo,M.A., de Bakker,P.I.W. and Blundell,T.L. (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure*, **12**, 831–838.
44. Zhou,J., Deyheim,A., Krueger,S. and Gregurick,S.K. (2005) LORES: low resolution shape program for the calculation of small-angle scattering profiles for biological macromolecules in solution. *Comput. Phys. Commun*., **170**, 186–204.
45. Van Hove,L. (1954) Correlations in space and time and born approximation scattering in systems of interacting particles. *Phys. Rev*., **95**, 249–262.
46. Krueger,S., Gregurick,S.K., Shi,Y., Wang,S., Wladkowski,D.B. and Schwarz,F.P. (2003) Entropic nature of the interaction between promoter

bound CRP mutants and RNA polymerase. *Biochemistry*, **47**, 1958.
47. Krueger,S., Gregurick,S.K., Zondlo,J. and Eisenstein,E. (2003) Interaction of GroEL and GroEL/GroES complex with a non-native subtilisin variant: a small angle neutron scattering study. *J. Struct. Biol.*, **141**, 240–258.
48. Jacrot,B. (1976) The study of biological structures by neutron scattering from solution. *Rep. Prog. Phys.*, **39**, 911–953.
49. Bevan,D.R., Leping,L., Pedersen,L.G. and Darden,T.A. (2000) Molecular dynamics simulations of the d(CCAACGTTGG)2 decamer: influence of the crystal environment. *Biophys. J.*, **78**, 668–682.
50. Baucom,J., Transue,T., Fuentes-Cabrera,M., Krahn,J.M., Darden,T.A. and Sagui,C. (2004) Molecular dynamics simulations of the d(CCAACGTTGG)2 decamer in crystal environment: comparison of atomic point-charge, extra-point, and polarizable force fields. *J. Chem. Phys.*, **121**, 6998–7008.
51. Tsui,V. and Case,D.A. (2001) Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275–291.
52. Prive,G.G., Yanagi,K. and Dickerson,R.E. (1991) Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J. Mol. Biol.*, **217**, 177–199.
53. Glinka,C.J., Barker,J.G., Hammouda,B., Krueger,S., Moyer,J.J. and Orts,W.J. (1998) The 30 m small-angle neutron scattering instruments at the national institute of standards and technology. *J. Appl. Cryst.*, **31**, 430–445.
54. Nishida,K., Kaji,K., Kanaya,T. and Shibano,T. (2002) Added salt effects on the intermolecular correlation in flexible polyelectrolyte solutions: small angle scattering study. *Macromolecules*, **35**, 4084–4089.
55. Borsali,R., Rinaudo,M. and Noirez,L. (1995) Light scattering and small-angle neutron scattering from polyelectrolyte solutions: the succinoglycan. *Macromolecules*, **28**, 1085–1088.
56. Milas,M., Rinaudo,M., Duplessix,R., Borsali,R. and Lindner,P. (1995) Small angle neutron scattering from polyelectrolyte solutions: from disordered to ordered xanthan chain conformation. *Macromolecules*, **28**, 3119–3124.
57. Ermi,B.D. and Amis,E.J. (1997) Influence of backbone solvation on small angle neutron scattering from polyelectrolyte solutions. *Macromolecules*, **30**, 6937–6942.
58. Dubois,E. and Boue,F. (2001) Conformation of poly(styrenesulfonate) polyions in the presence of multivalent ions: small-angle neutron scattering experiments. *Macromolecules*, **34**, 3684–3697.
59. Lu,X.-J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
60. Svergun,D.I. (1997) Restoring three-dimensional structure of biopolymers from solution scattering. *J. Appl. Cryst.*, **30**, 792–797.
61. Svergun,D.I., Volkov,V.V., Kozin,M.B., Stuhrmann,H.B., Barberato,C. and Koch,M.H. (1997) Shape determination from solution scattering of biopolymers. *J. Appl. Cryst.*, **30**, 798–802.
62. Svergun,D.I., Koch,M.H., Sayers,Z. and Kuprin,S. (1998) Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA*, **95**, 2267–2272.
63. Nollmann,M., Byron,O. and Stark,W.M. (2005) Behavior of Tn3 resolvase in solution and its interaction with res. *Biophys. J.*, **89**, 1920–1931.
64. Drew,H.R., Wing,R.M., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA*, **78**, 2179–2183.
65. Egli,M., Tereshko,V., Teplova,M., Minasov,G., Joachimiak,Z., Sanishvili,R., Weeks,C.M., Miller,R.L., Maier,M.A., An,H. *et al.* (2000) X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers*, **48**, 234–252.
66. Arai,S., Chatake,T., Ohhara,T., Kurihara,K., Tanaka,I., Suzuki,N., Fujimoto,Z., Mizuno,H. and Niimura,N. (2005) Complicated water orientations in the minor groove of the B-DNA decamer d(CCATTAATGG)2 observed by neutron diffraction measurements. *Nucleic Acids Res.*, **33**, 3017–3024.
67. Schneider,B., Patel,K. and Berman,H.M. (1998) Hydration of the phosphate group in double helical DNA. *Biophys. J.*, **75**, 2422–2434.
68. Caceres,C., Wright,C., Gouyette,C., Parkinson,G. and Subirana,J.A. (2004) A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Ti^{+}/Na^{+} ions. *Nucleic Acids Res.*, **32**, 1097–1102.