Supporting Material for: "Determining the locations of ions and water around DNA from X-ray scattering measurements"

Steve P. Meisburger[†], Suzette A. Pabit[†], and Lois Pollack[†*]

[†]School of Applied and Engineering Physics, Cornell University, Ithaca, New York

SUPPORTING THEORY

The following sections provide derivations of equations 1-6 in the Main Text, with supporting references and discussion.

Fundamental equations of X-ray scattering from a macromolecule in solution

The fundamental equation that describes a scattering experiment for a dilute solution of macromolecules is (1)

$$I_{sub}(q)/N = \left\langle \left| A_M(\mathbf{q}) + A_{solv}(\mathbf{q}) - \overline{B_{solv}(\mathbf{q})} \right|^2 \right\rangle_{\Omega,\Pi} - \left\langle \left| B_{solv}(\mathbf{q}) - \overline{B_{solv}(\mathbf{q})} \right|^2 \right\rangle_{\Omega,\Pi} \tag{S1}$$

where $\mathbf{q}$ is the momentum transfer, $I_{sub}(q)/N$ is the scattering per macromolecule (in excess of the bulk solvent background), $A_M(\mathbf{q})$ is the scattering amplitude of the macromolecule (in vacuum), $A_{solv}(\mathbf{q})$ is amplitude of the solvent molecules in a volume V that encloses the macromolecule and its region of influence, and $B_{solv}(\mathbf{q})$ is the amplitude of bulk-like solvent occupying a volume identical in shape to V but with no macromolecule present. The angle brackets with subscripts Ω and Π signify the averages over orientation and configuration respectively. The over-bar indicates a configuration-average with the positions of the macromolecule and solvent particles underlying $A_M(\mathbf{q})$ and $A_{solv}(\mathbf{q})$ held constant.

In Equation S1, the term $\left\langle \left| B_{solv}(\mathbf{q}) - \overline{B_{solv}(\mathbf{q})} \right|^2 \right\rangle_{\Omega,\Pi}$ is very small at low angles, and its calculation is only necessary for modeling WAXS profiles (q of order 1 Å^{-1}). Here, we set it equal to zero and approximate the total scattering intensity as

$$I_{sub}(q)/N \approx \left\langle \left| A_M(\mathbf{q}) + A_{solv}(\mathbf{q}) - \overline{B_{solv}(\mathbf{q})} \right|^2 \right\rangle_{\Omega,\Pi} \tag{S2}$$

Each amplitude term can be expressed as a sum over K different types of spherically symmetric particles,

$$A(\mathbf{q}) = \sum_{k=1}^{K} f_k(q) \sum_{j=1}^{N_k} e^{-i\mathbf{q}\cdot\mathbf{r}_j} \tag{S3}$$

where $f_k(q)$ is the scattering factor of the k^{th} type of particle, and r_j is the position of the j^{th} particle of type k. Because hydrogen atoms have weak scattering, for computational efficiency their contributions are folded into an effective scattering factor for their chemical group, as in the CRYSOL program (2). Similarly, water molecules are modeled by a spherically-symmetric, effective scattering factor $f_w(q)$ that combines the contributions from the oxygen the hydrogen atoms (3). Otherwise, $f_k(q)$ are the atomic form factors (4).

Excess scattering amplitude of of a charged macromolecule and its ion atmosphere

We consider a macromolecule with fixed atomic coordinates whose solvent atmosphere consists of water and one ion species. This case applies polyions (such as DNA) in low concentration monovalent salts, and is easily generalized to more complex

solvents. First, we derive a straightforward expression for the excess solvent amplitude, $A_{solv}(\mathbf{q}) - \overline{B_{solv}(\mathbf{q})}$. The solvent around the macromolecule, consisting of N_w water molecules and N_I ions, has a scattering amplitude given by Equation S3,

$$A_{solv}(\mathbf{q}) = f_w(q) \sum_{j=1}^{N_w} e^{-i\mathbf{q}\cdot\mathbf{r}_j} + f_I(q) \sum_{k=1}^{N_I} e^{-i\mathbf{q}\cdot\mathbf{r}_k} \tag{S4}$$

The configurational average of the excluded solvent amplitude can be expressed as a Fourier transform of the average electron density in the excluded volume V,

$$\overline{B_{solv}(\mathbf{q})} = \overline{\rho_{solv}} \int_V e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{S5}$$

where the average bulk solvent electron density is $\overline{\rho_{\text{solv}}}$. The exact volume V remains to be defined. A minimal choice includes the immediate region around the macromolecule (labeled by the superscript M) plus the region around each ion (the k^{th} ion is labeled by the superscript I_k). If the N_w water molecules are partitioned into these sub-volumes ($N_w^{(M)}$ waters in the volume $V^{(M)}$, and $N_w^{(I_k)}$ waters in the volume $V^{(I_k)}$), the excess solvent amplitude is

$$\begin{aligned} A_{solv}(\mathbf{q}) - \overline{B_{solv}(\mathbf{q})} \quad = \quad & f_w(q) \sum_{j=1}^{N_w^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}_j} - \overline{\rho_{solv}} \int_{V^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \\ & + \sum_{k=1}^{N_I} e^{-i\mathbf{q}\cdot\mathbf{r}_k} \left(f_I(q) + f_w(q) \sum_{l=1}^{N_w^{(I_k)}} e^{-i\mathbf{q}\cdot\mathbf{r}'_l} - \overline{\rho_{solv}} \int_{V^{(I_k)}} e^{-i\mathbf{q}\cdot\mathbf{r}'} d\mathbf{r}' \right) \end{aligned} \tag{S6}$$

The primed coordinates are relative to each ion's position (e.g. $\mathbf{r}'_l = \mathbf{r}_l - \mathbf{r}_k$).

If an ion's hydration shell is independent of its position around the macromolecule (i.e. ions do not dehydrate or otherwise "bind" the macromolecule), the term in parenthesis is approximately the same for all ions (independent of k), and can be replaced by an averaged effective scattering factor. The simplest choice of effective scattering factor is a constant, Z_{eff}, obtained by setting $q = 0$ inside the parentheses:

$$Z_{\text{eff}} = f_I(0) + f_w(0)\, N_w^{(I)} - \overline{\rho_{\text{solv}}}\, V^{(I)} \tag{S7}$$

This effective scattering factor is evidently valid only at low scattering angles. In the case where the bulk solvent electron density is close to that of pure water,

$$\overline{\rho_{\text{solv}}} \approx f_w(0)\, \overline{n_w} \tag{S8}$$

where $\overline{n_w}$ is the number density of water molecules. Then the approximate scattering factor of the hydrated ions is

$$Z_{\text{eff}} \approx f_I(0) - f_w(0)\, \overline{n_w}\, V_I \tag{S9}$$

where V_I is the volume change due to the presence of the ion

$$V_I = V^{(I)} - N_w^{(I)} / \overline{n_w} \tag{S10}$$

equal to the absolute limiting partial molar volume (5) divided by Avogadro's number. Equation S9 is the same as Equation 2 in the Main Text, with the notational change $Z_I = f_I(0)$ and $\bar{\rho}_e = f_w(0)\, \overline{n_w}$.

Finally, substituting Z_{eff} into Equation S6, we obtain an expression for the excess solvent scattering amplitude

$$A_{solv}(\mathbf{q}) - \overline{B_{solv}(\mathbf{q})} \approx f_w(q) \sum_{j=1}^{N_w^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}_j} - \overline{\rho_{solv}} \int_{V^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} + Z_{\text{eff}} \sum_{k=1}^{N_I} e^{-i\mathbf{q}\cdot\mathbf{r}_k} \tag{S11}$$

After substitution of Equation S11 into Equation S2, the intensity is

$$I_{sub}(q)/N \approx \left\langle \left| A_M(\mathbf{q}) + f_w(q) \sum_{j=1}^{N_w^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}_j} - \overline{\rho_{solv}} \int_{V^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} + Z_{\text{eff}} \sum_{k=1}^{N_I} e^{-i\mathbf{q}\cdot\mathbf{r}_k} \right|^2 \right\rangle_{\Omega,\Pi} \tag{S12}$$

Total scattering intensity as a two-phase system

Central to the decomposition method is a treatment of the scattering as a two-phase system: the macromolecule and associated water constitute the first phase (M), hydrated ions are the second phase (I). The intensity for this two-phase system is

$$I_{sub}(q)/N \approx \left\langle \left| F_M(q) + F_I(q) \right|^2 \right\rangle_{\Omega,\Pi} \tag{S13}$$

where $F_M(q)$ and $F_I(q)$ are the amplitudes of each phase. With reference to Equation S12, the amplitudes are

$$F_M(q) = A_M(\mathbf{q}) + f_w(q) \sum_{j=1}^{N_w^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}_j} - f_w(0)\,\overline{n_w} \int_{V^{(M)}} e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{S14}$$

and

$$F_I(q) = Z_{\text{eff}} \sum_{k=1}^{N_I} e^{-i\mathbf{q}\cdot\mathbf{r}_k} \tag{S15}$$

Next, we expand Equation S13

$$I_{sub}(q)/N \approx \left\langle \left| F_M(q) \right|^2 \right\rangle_{\Omega,\Pi} + 2\left\langle \text{Re}\left(F_M(q) F_I^*(q) \right) \right\rangle_{\Omega,\Pi} + \left\langle \left| F_I(q) \right|^2 \right\rangle_{\Omega,\Pi} \tag{S16}$$

The magnitude of each term at $q = 0$ can be separated from the q-dependent part by defining real functions $P(q)$ with $P(0) = 1$,

$$I_{\text{sub}}(q)/N \approx \left| F_M(0) \right|^2 P_M(q) + 2Re\left(F_M(0) \left\langle F_I^*(0) \right\rangle_{\Pi} \right) P_{MI}(q) + \left\langle \left| F_I(0) \right|^2 \right\rangle_{\Pi} P_I(q) \tag{S17}$$

To simplify the notation, we define the partial amplitudes

$$\delta_M = F_M(0) \tag{S18}$$

$$\delta_I N_I = F_I(0) \tag{S19}$$

If the X-ray energy is far from the absorption edges for the atoms, the scattering factors at $q = 0$ are real numbers, and δ_M and δ_I are real. Furthermore, the configurational averages may be dropped from Equation S17 provided the number fluctuations for ions and water are small (i.e. $\left\langle N^2 \right\rangle \sim \left\langle N \right\rangle^2$), leaving

$$I_{\text{sub}}(q)/N \approx \delta_M^2 P_M(q) + 2\delta_M N_I \delta_I P_{MI}(q) + \left(N_I \delta_I \right)^2 P_I(q) \tag{S20}$$

Equation S20 is identical to Equation 4 in the Main Text. From Equations S14, S15, S18 and S19, the partial amplitudes δ are equal to

$$\delta_M = A_M(0) + f_w(0) \left(N_w^{(M)} - \overline{n_w}\, V^{(M)} \right) \tag{S21}$$

$$\delta_I \quad = \quad Z_{\text{eff}} \tag{S22}$$

Equation 3 in the Main Text, $\delta_M = Z_{DNA} + 10\,(N_H - N_E)$, can be obtained from Equation S21, above, by the change of notation, $A_M(0) = Z_{\text{DNA}}$, $f_w(0) = 10$, $N_w^{(M)} = N_H$, and $\overline{n_w}\, V^{(M)} = N_E$.

Two-phase model for anomalous scattering

In an anomalous scattering experiment, the X-ray energy is tuned near the absorption edge for an atom of interest. If the ion contrast is varied in this way, its scattering factor is

$$f_I(q, E) = f_I(q) + f'(E) + i\, f''(E) \tag{S23}$$

Thus, the effective scattering amplitude of the ion is modified by

$$\delta_I(E) = Z_{\text{eff}} + f'(E) + i\, f''(E) \tag{S24}$$

The anomalous scattering experiment measures $I_{sub}(q)$ at two energies, E_1 and E_2. The anomalous difference is

$$I_{anom}(q) = I_{sub}(q, E_2) - I_{sub}(q, E_1) \tag{S25}$$

According to Equations S17, S18, S19 and S24,

$$\begin{aligned} I_{anom}(q)/N = {} & 2\langle N_I\rangle \delta_M \left(Z_{\text{eff}} + f'(E_2)\right) P_{MI}(q) + \left\langle N_I^2\right\rangle \left\{\left(Z_{\text{eff}} + f'(E_2)\right)^2 + f''(E_2)^2\right\} P_I(q) \\ & - 2\langle N_I\rangle \delta_M \left(Z_{\text{eff}} + f'(E_1)\right) P_{MI}(q) - \left\langle N_I^2\right\rangle \left\{\left(Z_{\text{eff}} + f'(E_1)\right)^2 + f''(E_1)^2\right\} P_I(q) \end{aligned} \tag{S26}$$

This simplifies to

$$\begin{aligned} I_{anom}(q)/N = {} & 2\langle N_I\rangle \delta_M \left(f'(E_2) - f'(E_1)\right) P_{MI}(q) \\ & + \left\langle N_I^2\right\rangle \left\{2\, Z_{\text{eff}} \left(f'(E_2) - f'(E_1)\right) + \left(f'(E_2) - f'(E_1)\right)\left(f'(E_2) + f'(E_1)\right) + f''(E_2)^2 - f''(E_1)^2\right\} P_I(q) \end{aligned} \tag{S27}$$

In the ASAXS experiment, we chose the energies E_1 and E_2 to be below the absorption edge, so that $f''(E_2) \approx f''(E_1)$. Thus, Equation S27 simplifies further:

$$I_{anom}(q)/N = 2\langle N_I\rangle \delta_M \left(f'(E_2) - f'(E_1)\right)\left(P_{MI}(q) + \gamma\, P_I(q)\right) \tag{S28}$$

where

$$\gamma = \frac{\left\langle N_I^2\right\rangle}{\langle N_I\rangle \delta_M}\left(Z_{\text{eff}} + \frac{1}{2}\left(f'(E_2) + f'(E_1)\right)\right) \tag{S29}$$

Again, assuming the number density fluctuations are small, so that $\left\langle N_I^2\right\rangle \sim \langle N_I\rangle^2$,

$$\gamma \approx \frac{\langle N_I\rangle}{\delta_M}\left(Z_{\text{eff}} + \frac{1}{2}\left(f'(E_2) + f'(E_1)\right)\right) \tag{S30}$$

Equations S28 and S30 are the same as Equation 6 in the Main Text.

SUPPORTING MATERIALS AND METHODS

Sample preparation

The 25 base-pair DNA duplex was composed of two single stranded oligonucleotides with sequence GCATCTGGGC-TATAAAAGGGCGTCG and its complement. The sequence has been used in many previous small-angle X-ray scattering (SAXS) studies (6–14). Oligonucleotides were synthesized and HPLC purified by Integrated DNA Technologies (Coralville, IA) and delivered as lyophilized powders. Each strand was rehydrated in aqueous buffer containing 10 mM TRIS, 50 mM NaCl, 1 mM EDTA, pH 8.0. The concentration of each strand was calculated from the UV absorption at 260 nm using extinction coefficients derived from the nearest-neighbor model (15); $\varepsilon_{\text{S1}} = 244,400\ \text{L mol}^{-1}\text{cm}^{-1}$, and $\varepsilon_{\text{S2}} = 228,500\ \text{L mol}^{-1}\text{cm}^{-1}$. Strands were mixed in an equimolar ratio, annealed at 94 C for 4 minutes, and allowed to cool on the bench.

Buffered salt solutions were prepared using 1 mM Na-MOPS pH 7.0, 100 mM of 1:1 salt (NaCl, KCl, RbCl, CsCl) and NANOpure water (Barnstead, Dubuque, IA). All reagents were purchased from Sigma-Aldrich, unless specified. DNA in each buffered salt solution was prepared by spin dialysis using an Amicon Ultra-0.5 mL with a 10 kDa cutoff (Millipore, Billerica, MA), repeated to ensure complete exchange. The DNA concentration during spin dialysis did not exceed 0.1 mM. The final flow-through was retained for SAXS background subtraction. The The UV absorbance of each duplex DNA solution at 260 nm was converted to concentration using the hypochromicity-corrected extinction coefficient (16), $\varepsilon_{\text{S1+S2}} = 397,600\ \text{L mol}^{-1}\text{cm}^{-1}$.

In previous studies, 50 μM DNA was shown to agree with the infinite dilution limit within the measurement noise (6). However, considering improvements in signal strength and extended q-range of the measurements, some residual interparticle interference is possible at this concentration, if difficult to confirm experimentally (e.g. by collecting data at even lower concentration). Therefore, we modeled the DNA structure factor theoretically using parameters determined previously for this system at high DNA concentrations (6). At 50 μM, the structure factor contribution is predicted to be less than 3% of the signal at $q = 0$, and it decays rapidly to zero by $q \sim 0.05\ \text{Å}^{-1}$ (Figure S1). When the fits shown in Figure 6 in the Main Text are repeated using data truncated at $q = 0.05\ \text{Å}^{-1}$, we find little effect on the minimum value of $N_I^{(S)}$ (Figure S3).

Absolute intensity calibration

For the ASAXS and heavy atom experiments, the scattering intensity was placed on an absolute scale using liquid water as a calibrant (17). The forward X-ray scattering of a liquid (i.e. the macroscopic scattering cross-section, $d\Sigma/d\Omega$) is proportional to the isothermal compressibility:

$$\frac{d\Sigma}{d\Omega} = n\, r_0^2\, Z^2\, k_B T\, \chi_T \tag{S31}$$

where r_0 is the classical electron radius, n is the molecular number density, Z is the number of electrons per molecule (10 for water), and χ_T is the (temperature-dependent) isothermal compressibility, which has been determined accurately using speed of sound measurements (18). For liquid water at 23C, $\chi_T = 4.55 \times 10^{-10}\text{Pa}^{-1}$ and $d\Sigma/d\Omega = 0.0164\,\text{cm}^{-1}$.

For calibrating X-ray data, the scattering curve for pure water was measured in the sample cell, and the scattering from the empty cell was measured and subtracted. The scattering curves for each macromolecular sample, $I(q)_{\text{sample}}$, were scaled by a constant:

$$\left.\frac{d\Sigma(q)}{d\Omega}\right|_{\text{sample}} = \frac{I(q)_{\text{sample}}}{I(q\to 0)_{\text{water}}} \left.\frac{d\Sigma}{d\Omega}\right|_{\text{water}} \tag{S32}$$

The absolute scattering cross section was converted to units of electron^2 per molecule by dividing by the concentration of molecules (number density n) and the square of the classical electron radius r_0:

$$I(q)\ \left[\text{electron}^2\right] = \frac{1}{n\, r_0^2} \frac{d\Sigma(q)}{d\Omega} \tag{S33}$$

Equations S32 and S33 can be combined in a convenient form

$$I(q)_{\text{sample}}\ \left[\text{electron}^2\right] = \frac{1000\,\text{cm}^3/\text{L}}{N_A\ c_{\text{sample}}} \cdot \frac{1}{r_0^2} \cdot \frac{I(q)_{\text{sample}}}{I(q\to 0)_{\text{water}}} \left.\frac{d\Sigma}{d\Omega}\right|_{\text{water}} \tag{S34}$$

where c_{sample} is the sample concentration in units of mole/L (as measured from standard techniques like UV-VIS), r_0 and $d\Sigma/d\Omega$ are in CGS units, and N_A is Avogadro's number.

Robust matrix method for decomposition of isomorphous replacement SAXS data

Equation S20 can be written in matrix form as $I = P \cdot C$, where the columns of I contain the scattering intensities, the columns of P contain the basis functions $P_M(q)$, $P_{MI}(q)$, and $P_I(q)$, and C is a matrix of coefficients. Written explicitly,

$$\begin{bmatrix} I^{(1)}(q_1) & I^{(2)}(q_1) & \cdots \\ I^{(1)}(q_2) & I^{(2)}(q_2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} P_M(q_1) & P_{MI}(q_1) & P_I(q_1) \\ P_M(q_2) & P_{MI}(q_2) & P_I(q_2) \\ \vdots & \vdots & \vdots \end{bmatrix} \cdot \begin{bmatrix} \delta_M^2 & \delta_M^2 & \cdots \\ 2\delta_M N_I \delta_I^{(1)} & 2\delta_M N_I \delta_I^{(2)} & \cdots \\ N_I^2 \left(\delta_I^{(1)}\right)^2 & N_I^2 \left(\delta_I^{(2)}\right)^2 & \cdots \end{bmatrix} \tag{S35}$$

where superscripts enumerate the different salt solutions. In the case where the matrix I contains experimental data, Equation S35 is

$$I_{exp} = P\, C + E \tag{S36}$$

where E is a matrix of errors due to measurement noise. The solution P that minimizes the least-squares error is

$$P = I_{exp}\, C^{+} \tag{S37}$$

where C^{+} is the Moore-Penrose inverse of C. However, finding P in this way will only work with perfectly-calibrated data. The solution can be found robustly if both sides of Equation S35 are normalized by the forward scattering:

$$I_{exp}\, I(0)_{exp}^{-1} = P\left(C\, I(0)^{-1}\right) \tag{S38}$$

where $I(0)_{exp}$ is a diagonal matrix containing the zero-angle scattering (determined from the experimental data, e.g. using the indirect Fourier transform),

$$I(0)_{exp} = \begin{bmatrix} I^{(1)}(0) & 0 & \cdots \\ 0 & I^{(2)}(0) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{S39}$$

and $I(0)$ is the prediction for $I(0)_{exp}$,

$$I(0) = \begin{bmatrix} \left(\delta_M + N_I\delta_I^{(1)}\right)^2 & 0 & \cdots \\ 0 & \left(\delta_M + N_I\delta_I^{(2)}\right)^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{S40}$$

The robust least-squares solution for P in Equation S38 is

$$P = I_{exp}\, I(0)_{exp}^{-1} \left(C\, I(0)^{-1}\right)^{+} \tag{S41}$$

where $\left(C\, I(0)^{-1}\right)^{+}$ is the Moore-Penrose inverse of $(C\, I(0)^{-1})$, calculated using the PINV function in MATLAB.

DNA hydration model and X-ray scattering simulation

The spherical harmonic expansion method enables efficient calculation of the orientationally-averaged intensity from a system composed of discrete, spherically-symmetric particles (2). To take advantage of this method, the electron densities of the DNA, its associated solvent molecules, and the excluded solvent were modeled using an explicit particle representation.

The atomic coordinates of a 25 base-pair, B-form DNA duplex were generated by Nucleic Acid Builder (19). A geometric hydration model described previously (20) was implemented in MATLAB: water-like particles were placed at random but non-overlapping positions >1.50 Å from the DNA's Van-der-Waals surface. A total of 384 (8 per phosphate group) were used. To model the excluded solvent, a pre-computed water box (21) was superimposed on the DNA model, and water molecules outside a cutoff distance from the DNA surface were discarded. This cutoff was chosen to match the experimentally determined δ_M (see Results in the Main Text). For averaging over solvent configurations, the excluded volume calculation was repeated for 8 different positions of the DNA within the water box and the particles were assigned spherical Gaussian scattering factors to better approximate uniform density (2, 22).

The ion atmosphere was modeled by solving the nonlinear Poisson-Boltzmann equation (NPBE) numerically on a grid using the program APBS (23). First, the DNA's atomic charges and radii were assigned by PDB2PQR with Amber parameters (24). The APBS program was configured with a 100 Å × 100 Å × 168 Å box size, 129 × 129 × 193 mesh elements, and a zero-potential boundary condition. The excess ion density was computed from the NPBE solution, and discretely sampled using Monte Carlo with 10,000 points (the probability for a point to occupy a cell was proportional to the excess ion density times the cell volume, and within each cell the coordinates were assigned at random withuniform probability).

Reduced χ^2 statistic for comparing scattering profiles

To provide a quantitative comparison between experimentally-determined terms $P_{exp}(q)$ and those predicted by models $P_{calc}(q)$, we compute the following chi-squared statistic:

$$\chi^2_{rel} = \frac{1}{\beta^2 N} \sum_{i=1}^{N} \left(\frac{P_{exp}(q_i) - P_{calc}(q_i)}{\sigma_{exp}(q_i)} \right)^2 \tag{S42}$$

where σ_{exp} is an estimate of the experimental uncertainty and β is a scale factor so that $\chi^2_{rel} = 1$ when $P_{calc}(q)$ is replaced by the regularized version of $P_{exp}(q)$ (smoothed data in Figures 2 and 3 in the Main Text). Although in the ideal case $\beta = 1$, we find $\beta \sim 0.94$ when σ_{exp} is determined by propagation of uncertainty through Equation S41.

SUPPORTING REFERENCES

1. Park, S., J. P. Bardhan, B. Roux, and L. Makowski, 2009. Simulated X-ray scattering of protein solutions using explicit-solvent models. J. Chem. Phys. 130:134114.
2. Svergun, D., C. Barberato, and M. H. J. Koch, 1995. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Crystallogr. 28:768–773.
3. Wang, J., A. N. Tripathi, and V. H. Smith Jr, 1994. Chemical binding and electron correlation effects in x-ray and high energy electron scattering. The Journal of chemical physics 101:4842–4854.
4. Cromer, D. T., and J. B. Mann, 1968. X-ray scattering factors computed from numerical Hartree-Fock wave functions. Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography 24:321–324.

5. Marcus, Y., 1985. Ion Solvation. Wiley.
6. Qiu, X., L. W. Kwok, H. Y. Park, J. S. Lamb, K. Andresen, and L. Pollack, 2006. Measuring Inter-DNA Potentials in Solution. Phys. Rev. Lett. 96:138101+.
7. Pabit, S. A., S. P. Meisburger, L. Li, J. M. Blose, C. D. Jones, and L. Pollack, 2010. Counting ions around DNA with anomalous small-angle X-ray scattering. J. Am. Chem. Soc. 132:16334–16336.
8. Qiu, X., K. Andresen, J. S. Lamb, L. W. Kwok, and L. Pollack, 2008. Abrupt transition from a free, repulsive to a condensed, attractive DNA phase, induced by multivalent polyamine cations. Phys. Rev. Lett. 101:228101.
9. Andresen, K., R. Das, H. Y. Park, H. Smith, L. W. Kwok, J. S. Lamb, E. Kirkland, D. Herschlag, K. Finkelstein, and L. Pollack, 2004. Spatial distribution of competing ions around DNA in solution. Phys. Rev. Lett. 93:248103.
10. Andresen, K., X. Qiu, S. A. Pabit, J. S. Lamb, H. Y. Park, L. W. Kwok, and L. Pollack, 2008. Mono-and trivalent ions around DNA: a small-angle scattering study of competition and interactions. Biophys. J. 95:287–295.
11. Pabit, S. A., X. Qiu, J. S. Lamb, L. Li, S. P. Meisburger, and L. Pollack, 2009. Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA. Nucleic Acids Res. 37:3887–3896.
12. Li, L., S. A. Pabit, S. P. Meisburger, and L. Pollack, 2011. Double-stranded RNA resists condensation. Phys. Rev. Lett. 106:108101.
13. Blose, J. M., S. A. Pabit, S. P. Meisburger, L. Li, C. D. Jones, and L. Pollack, 2011. Effects of a Protecting Osmolyte on the Ion Atmosphere Surrounding DNA Duplexes. Biochemistry 50:8540–8547.
14. Das, R., T. Mills, L. Kwok, G. Maskel, I. Millett, S. Doniach, K. Finkelstein, D. Herschlag, and L. Pollack, 2003. Counterion distribution around DNA probed by solution X-ray scattering. Phys. Rev. Lett. 90:188103.
15. Cantor, C. R., M. M. Warshaw, and H. Shapiro, 1970. Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleolides. Biopolymers 9:1059–1077.
16. Tataurov, A. V., Y. You, and R. Owczarzy, 2008. Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. Biophys. Chem. 133:66–70.
17. Orthaber, D., A. Bergmann, and O. Glatter, 2000. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. J. Appl. Crystallogr. 33:218–225.
18. Kell, G. S., 1970. Isothermal compressibility of liquid water at 1 atm. J. Chem. Eng. Data 15:119–122.
19. Macke, T. J., and D. A. Case, 1998. Modeling Unusual Nucleic Acid Structures. In N. B. Leontis, and J. SantaLucia, editors, Molecular Modeling of Nucleic Acids, American Chemical Society, Washington, DC, 379–393.
20. Durchschlag, H., and P. Zipper, 2003. Modeling the hydration of proteins: prediction of structural and hydrodynamic parameters from X-ray diffraction and scattering data. Eur. Biophys. J. 32:487–502.
21. Yang, S., M. Parisien, F. Major, and B. Roux, 2010. RNA structure determination using SAXS data. J. Phys. Chem. B 114:10039–10048.
22. Merzel, F., and J. C. Smith, 2002. SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. Acta Crystallogr. D Biol. Crystallogr. 58:242–249.
23. Holst, M., N. Baker, and F. Wang, 2000. Adaptive multilevel finite element solution of the Poisson–Boltzmann equation I. Algorithms and examples. J. Comput. Chem. 21:1319–1342.
24. Dolinsky, T. J., J. E. Nielsen, J. A. McCammon, and N. A. Baker, 2004. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res. 32:W665–W667.

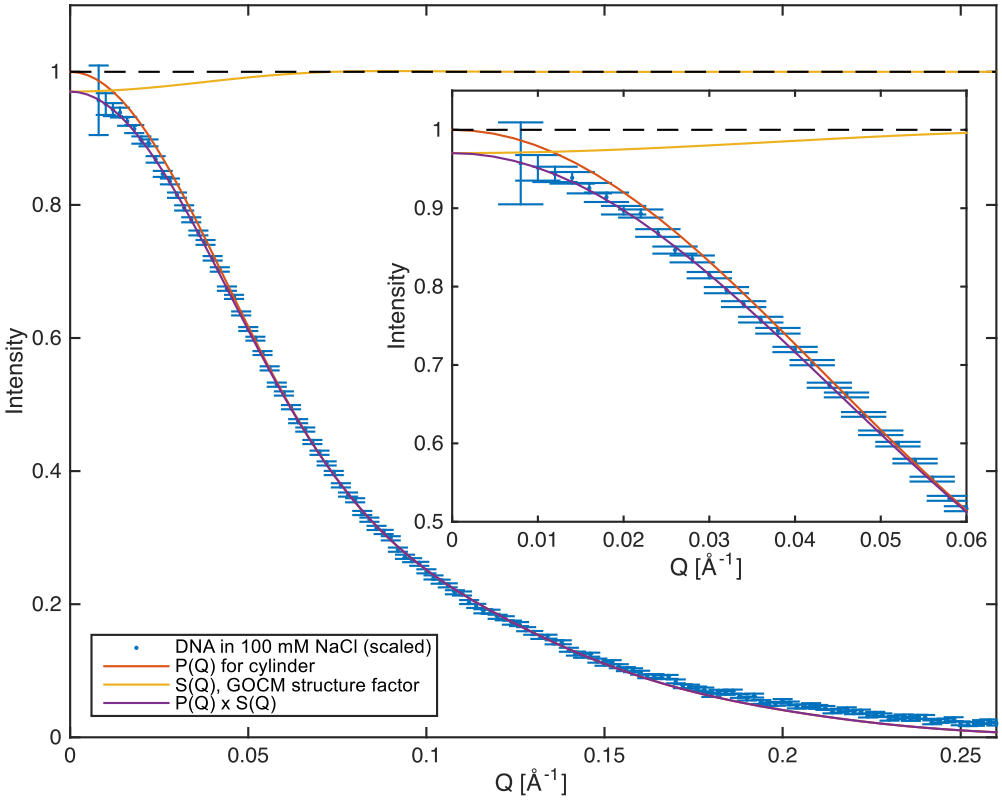

Figure S1: Modeling interparticle interference. The structure factor for a DNA duplex in 100 mM NaCl was calculated using the generalized one component macroion theory (GOCM) and decoupling approximation with a cylindrical form factor (6). Model parameters were determined from SAXS measurements of DNA solutions at high concentration, published previously (6): $\sigma = 55.0\,$Å (effective particle diameter), z_m = 9.1 (effective charge), and $I = 100\,$mM (ionic strength). The predicted concentration-dependent scattering profile is $I(q) = P_{CYL}(q, R, H) \times S(q, c, R, H)$, where $P(q) = P_{CYL}(q, R, H)$ is the form factor for a cylinder with radius R and height H, and $S(q)$ is the structure factor (note that $S(q)$ depends on R and H through the decoupling approximation). To check for consistency between the structure factor model and the observed data at low concentration, we fit the model to the SAXS data ($50\,\mu$M DNA in 100 mM NaCl, Figure 1 in the Main Text) on an arbitrary intensity scale over the q-range $0.008 < q < 0.1\,$Å^{-1}. The model with parameters $R = 11.8\,$Å and $H = 82\,$Å is shown along with the corresponding form factor $P(q)$ and structure factor $S(q)$. According to the model, the data differs from the dilute limit, $P(q)$, in the low-q region only (inset), and the magnitude of the interparticle interference effect is at most 3%.

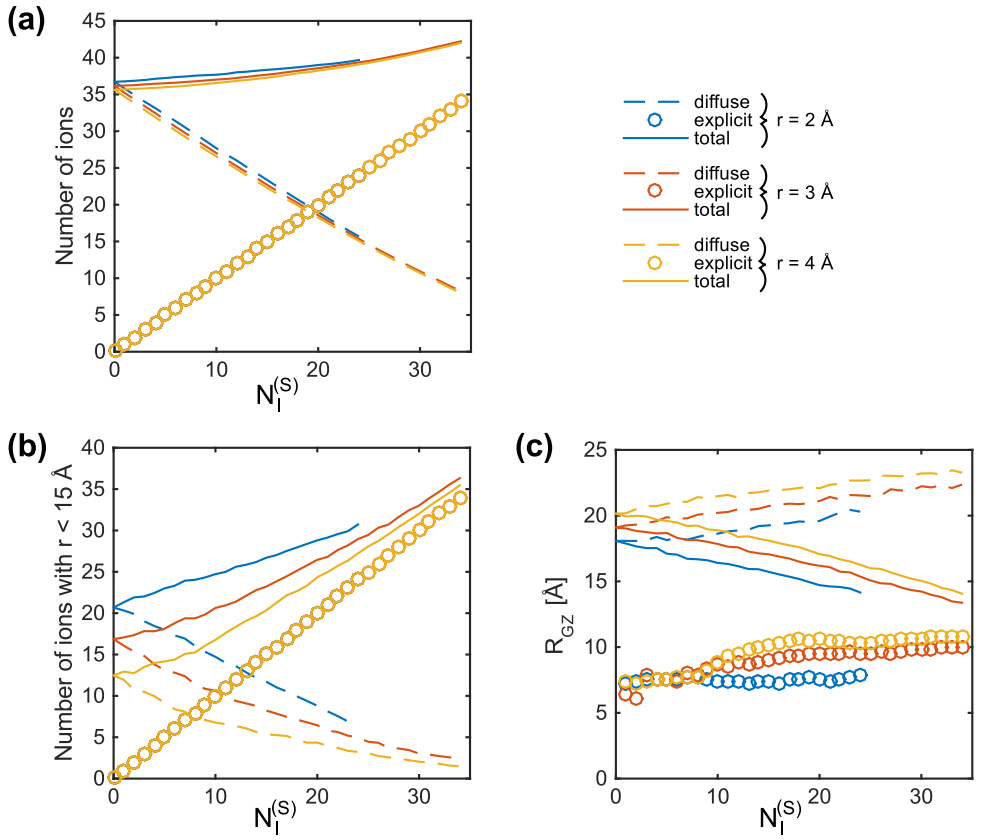

Figure S2: Statistics from ion atmosphere models generated using the explicit / nonlinear Poisson Boltzmann equation (NPBE) approach described in the Main Text. The ion atmosphere has two classes of ions, explicit ions fixed in place near the DNA surface, and diffuse ions modeled by a continuous distribution from the NPBE solution. The total number of ions is the sum of the explicit (“surface”) ions, $N_I^{(S)}$, and the number of diffuse ions $N_I^{(D)}$ (equal to the integral of the diffuse density, $N_I^{(D)} = \int_V \rho(r)\, dV$). The calculation was repeated for three different values of the Stern layer thickness, $r = 2$, 3, or 4 Å and for $N_I^{(S)}$ between 0 and 34. (a) As $N_I^{(S)}$ increases, the $N_I^{(D)}$ decreases, while the total number $N_I^{(S)} + N_I^{(D)}$ depends only weakly on $N_I^{(S)}$. (b) Although the total number of ions is relatively constant, their distribution is not. To illustrate, we counted the number of diffuse and explicit ions within a cylinder of radius $r = 15$ Å centered on the DNA molecule. The overall effect of adding explicit ions is to increase the total number of ions near the DNA surface, at the expense of the diffuse component further away. The number of diffuse ions near the surface depends strongly on the choice of Stern radius. (c) The size of the ion distribution for the different models was compared by calculating the radius of gyration of the ions about the Z-axis of the DNA, R_{GZ}, defined as $R_{GZ,S}^2 = \left(N_I^{(S)}\right)^{-1} \sum_{i=1}^{N_S} \left(x_i^2 + y_i^2\right)$ for the explicit ions, and $R_{GZ,D}^2 = \left(N_I^{(D)}\right)^{-1} \int_V \rho(x,y,z)\left(x^2+y^2\right) dV$ for the diffuse ions. The value of R_{GZ} for the total ion atmosphere is $R_{GZ}^2 = \left(N_I^{(S)} R_{GZ,S}^2 + N_I^{(D)} R_{GZ,D}^2\right) / \left(N_I^{(S)} + N_I^{(D)}\right)$.

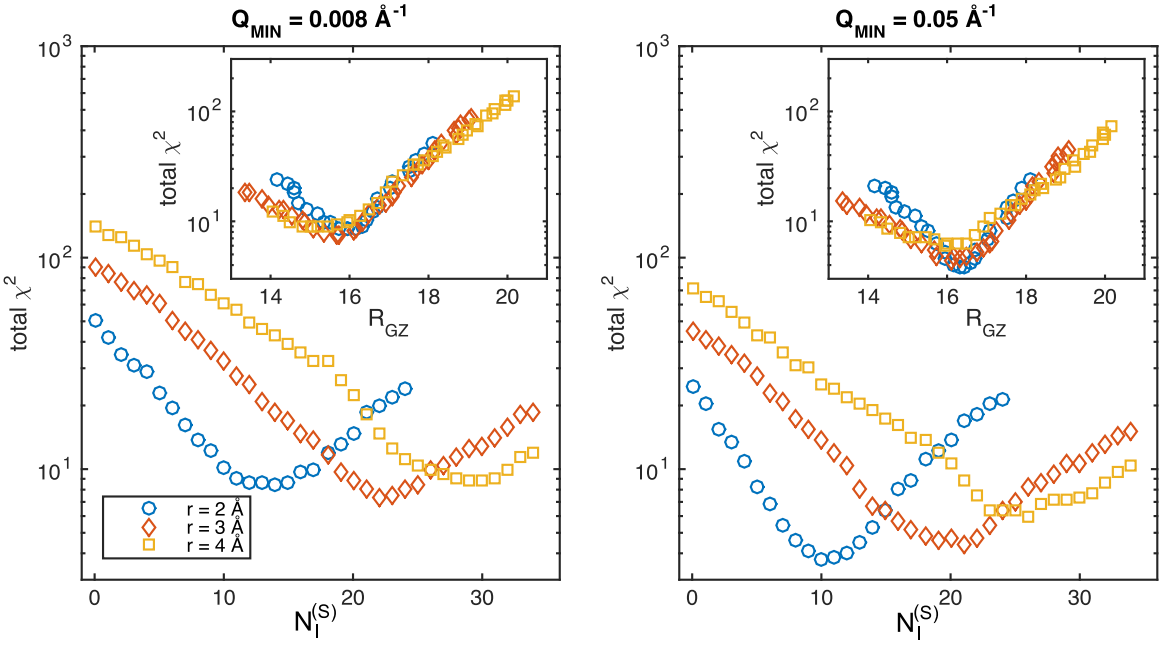

Figure S3: Ad-hoc parameters of the ion atmosphere model and fitting procedure do not affect overall conclusions. (left) Results of fitting the explicit ion / nonlinear Poisson Boltzmann equation (NPBE) hybrid models to the entire heavy atom data set. Results for three Stern layer thickness indicated in the legend are compared (points for $r = 2$ Å are identical to Figure 6b in the Main Text). (right) The first were repeated using a truncated q-range, chosen to minimize the potential effects of residual inter-particle interference (see Figure S1). Truncation improves the overall quality of the fits, but does not alter the interpretation. In both cases, the minimum χ^2 occurs for $N_I^{(S)}$ between 0 and 34. The location of the minimum depends on the choice of Stern layer thickness. As discussed in Figure S2, the Stern layer thickness affects the spatial extent of the ion atmosphere at $N_I^{(S)} = 0$, and as $N_I^{(S)}$ increases the average ion atmosphere size is reduced. Since SAXS data is low resolution, it is reasonable to guess that the primary feature of the ion atmosphere determining χ^2 is the rotational moment about the Z-axis (R_{GZ} in Figure S2c), and not the fine details of the ion placement. Indeed, when χ^2 is plotted as a function of R_{GZ} rather than $N_I^{(S)}$ (insets), models with different Stern radii collapse to a single curve with a minimum around $R_{GZ} \sim 16$ Å (full q-range) or $R_{GZ} \sim 16.4$ Å (truncated q-range). The best-fit value is significantly more compact than the Poisson-Boltzmann (PB) solution ($R_{GZ} \sim 18$ Å with $r = 2$ Å). Thus, our conclusion that PB overestimates the spatial extent of the ion atmosphere does not depend the details of the calculation, i.e. the Stern radius value or the exclusion of low-angle data.