From a Structural Average to the Conformational Ensemble of a DNA Bulge

Xuesong Shi[†], Kyle Beauchamp[‡], Pehr A. B. Harbury[†,*] and Daniel Herschlag[†,*]

[†]Department of Biochemistry, Stanford University, Stanford, California, United States [‡]Biophysics Program, Stanford University, Stanford, California, United States

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Direct experimental measurements of conformational ensembles are critical for understanding macromolecular function, but traditional biophysical methods do not directly report the solution ensemble of a macromolecule. Small angle x-ray scattering interferometry has the potential to overcome this limitation by providing the instantaneous distance distribution between pairs of gold-nanocrystal probes conjugated to a macromolecule in solution. Our x-ray interferometry experiments reveal an increasing bend angle of DNA duplexes with bulges of one, three, and five adenosine residues, consistent with previous FRET measurements, and further reveal an increasingly broad conformational ensemble with increasing bulge length. The distance distributions for the AAA bulge duplex (3A-DNA) with six different Au-Au pairs provide strong evidence against a simple elastic model in which fluctuations occur about a single conformational state. Instead, the measured distance distributions suggest a 3A-DNA ensemble with multiple conformational states predominantly across a region of conformational space with bend angles between 24 and 85 degrees and characteristic bend directions and helical twists and displacements. Additional x-ray interferometry experiments revealed perturbations to the ensemble from changes in ionic conditions and the bulge sequence, effects that can be understood in terms of electrostatic and stacking contributions to the ensemble and that demonstrate the sensitivity of x-ray interferometry. Combining x-ray interferometry ensemble data with molecular dynamics simulations gave atomic-level models of representative conformational states and of the molecular interactions that may shape the ensemble, and fluorescence measurements with 2-aminopurine-substituted 3A-DNA provided initial tests of these atomistic models. More generally, x-ray interferometry will provide powerful benchmarks for testing and developing computational models.

ensemble | helix-junction-helix | bulge | x-ray interferometry | SAXS

INTRODUCTION

A grand challenge in biology is to understand the complex free energy landscape of macromolecules and to decipher the resulting conformational ensembles. To perform their biological functions, macromolecules must adopt a multiplicity of conformations. Balancing and controlling different conformational states is central to biological processes including protein folding, allostery and signaling, and the stepwise assembly and function of macromolecular machines. To understand these complex molecules requires characterization of their free energy landscapes—i.e., their equilibrium conformational ensembles. Precise measurements of conformational ensembles could allow quantitative modeling of the folding and function of biological macromolecules, would provide valuable experimental data to test current computational models and assumptions, and might facilitate the rational design of specifically acting inhibitors (1, 2).

Techniques including NMR and EPR relaxation techniques have been developed to incisively probe motions in the ensemble on different time scales, ranging from picoseconds to milliseconds (3, 4). Nonetheless, such dynamic information represents an average of the dynamics of the molecules across the conformational ensemble. In special cases, where the ensemble contains slow exchanging conformational states, these states can be separately detected (e.g., relaxation dispersion approaches can detect conformational states interconverting at tens of microseconds to hundreds of milliseconds, and single-molecule FRET (smFRET) can characterize conformational transitions at millisecond or slower timescales (5, 6)). But again, each of these states is an average of a more complex local conformational ensemble.

To date, successes in reconstructing equilibrium ensembles have mostly relied on experimental measurement of NMR residual dipolar couplings (RDCs) (7, 8). Compared to other NMR techniques, RDCs provide long-range angular structure information that helps to generate equilibrium ensemble models (10). In combination with molecular dynamic simulations, RDCs have been used to generate ensemble models for small disordered proteins (7, 11), DNA duplexes (12), and a RNA bulge motif (13-15). In addition to RDCs, relaxation dispersion, and paramagnetic relaxation enhancement have been used to detect and characterize conformational states that are in low abundance in an ensemble (16). While powerful, these NMR-based methods, like all approaches, have limitations. For example, RDCs have difficulty distinguishing between conformations with similar angular orientations but different translational displacements (17, 18). Additional methods are needed to construct ensembles that can test and complement these current methods.

To meet this challenge, we continue to develop, test, and apply the capabilities of a recently developed solution x-ray interferometry technique (19, 20). X-ray interferometry can be used

Significance

Obtaining the conformational ensembles of biological macromolecules, beyond average structures, is extremely challenging but necessary for a complete understanding of the folding and functions of biological macromolecules. Such insights may also lead to the rational design of therapeutics that can target less-ordered macromolecules and may advance the design of nanostructures and nanomachines from nucleic acids. We have applied x-ray interferometry to estimate the conformational ensemble of a small model macromolecule, a DNA bulge, representative of helix-junction-helix building blocks of natural RNAs and designed DNA nanostructures. The measured ensemble, in combination with molecular dynamics simulations, allows generation of testable atomic-level models. X-ray interferometry can detect changes in the ensemble arising from different bugle sequences and solution salt conditions.

Reserved for Publication Footnotes

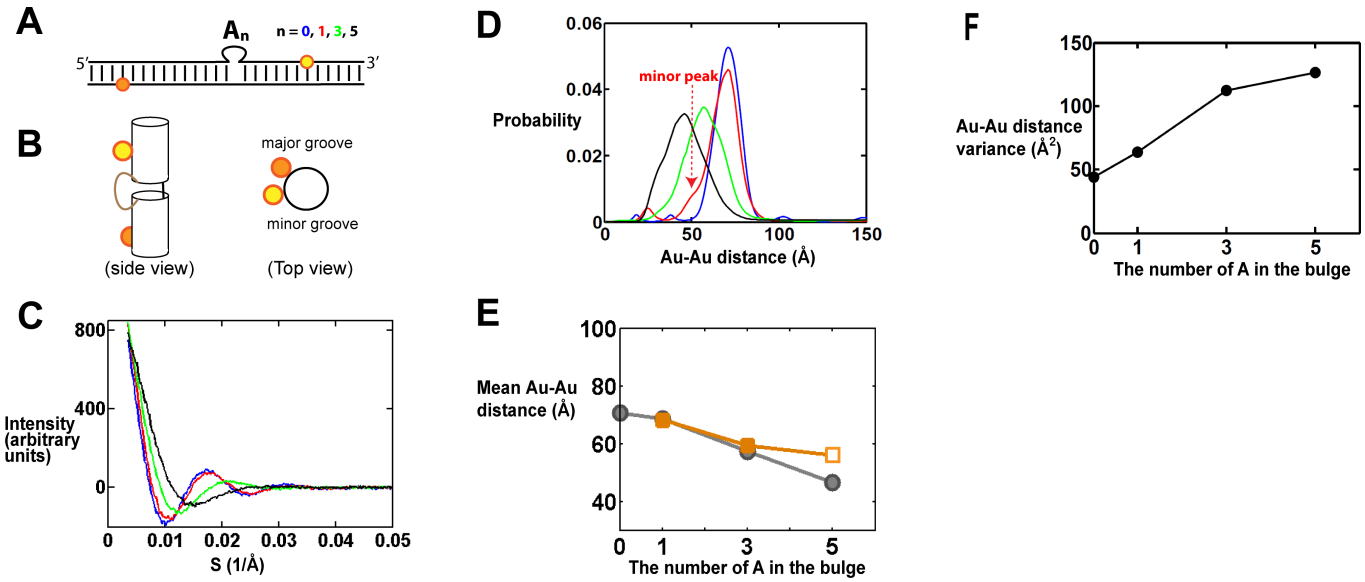

Fig. 1. X-ray interferometry measurement of DNA bulges with different numbers of adenosines. (**A**) Schematic depiction of bulged DNA constructs labeled with a pair of Au probes (yellow and orange sphere). The top (yellow) and bottom (orange) Au probes are labeled on T residues 7 and 11 base steps from the bulge site, respectively. (See *SI Appendix*, Table S1 for the sequences used.) Results for the 0-nt bugle construct (i.e., duplex DNA) are from reference (20) (**B**) A cartoon view of the constructs as a straight duplex. The left panel is the side view from the minor groove of the bulge site. In a continuous DNA helix the bottom gold nanocrystal (orange) would be about 58° clockwise of the top gold nanocrystal (yellow) when viewed from the top (right panel) and is away from the viewer when viewed from the minor groove side (left panel). (**C**) The Au-Au scattering profile for DNA constructs (panel A) with a bulge sequence of 0 (blue), 1 (red), 3 (green) and 5 (black) adenosines. The x-axis is the scattering angle parameter S. (**D**) The Au-Au center to center distance distributions deduced from the Au-Au scattering profiles in panel C, following the same color code as panel C.The minor peaks at the short and long distances, outside of the main distribution, are generally noise that is sample preparation dependent, as described in ref 25. (**E**) The mean Au-Au distance values determined from x-ray interferometry (grey) and predicted from the literature average structure models generated by smFRET (orange) (31). The 5A value from the smFRET model (open symbol) is unreliable as the smFRET data for the 5A construct were poorly fit by the model, as noted by the authors (32). (**F**) The Au-Au distance variance in panel D as a function of bulge length.

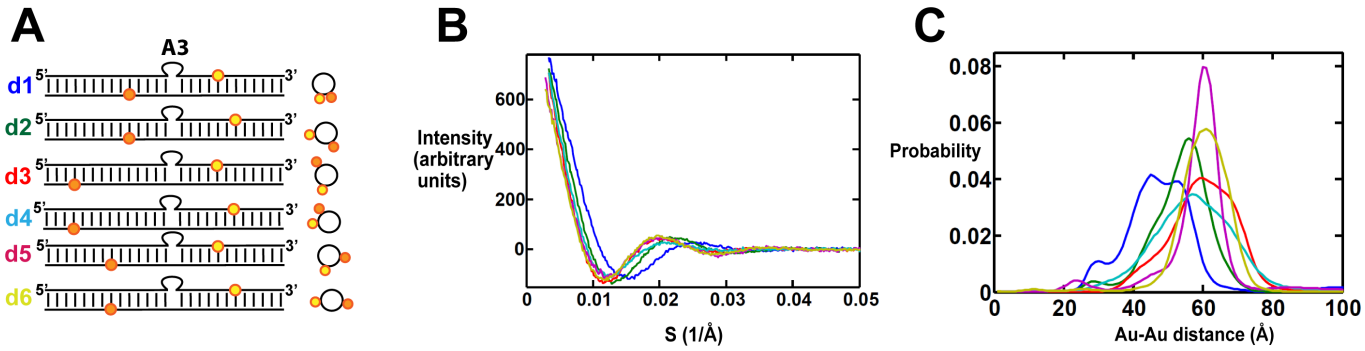

Fig. 2. X-ray interferometry measurement of 3A-DNA. (**A**) Schematic of the 3A-DNA constructs labeled with six pairs of Au nanocrystals (yellow and orange sphere). (See *SI Appendix*, Table S1 for the sequences of constructs d1 through d6). Cartoon views of the constructs show the Au nanocrystal positions (right) analogously to Fig. 1B. (**B**) The Au-Au scattering profiles for constructs d1 to d6, color coded as in panel A. (**C**) The Au-Au center to center distance distributions deduced from the Au-Au scattering profiles in panel B, color coded as in panels A and B.

to determine site-to-site distance distributions instantaneously because it relies on atomic scattering (19, 21-26). Standard small angle x-ray scattering (SAXS) measures the sum of the scattering and scattering interference from all atoms in a macromolecule (27). As it would not be possible to decompose this sum and distinguish contributions from specific atoms or atom pairs, standard SAXS provides no site-specific information and is limited to determining the overall size and shape of macromolecules (27). X-ray interferometry overcomes this limitation through the introduction of a pair of site-specifically labeled gold nanocrystal probes and isolation of the scattering interference from this strongly scattering probe pair. This scattering interference can be directly converted into a distance distribution through a Fourier transformation, without the complications of a nonlinear mapping (9). Multiple pairs of gold nanocrystal probes, in different site-specific locations, provide additional distance information and increase the information content of the technique (e.g., refs. 19, 20).

Unlike standard ensemble-average methods such as FRET that give a single average value for the distance between each probe pair, X-ray interferometry naturally yields a distance distribution between each probe pair. Strategies measuring the time dependence of fluorescence energy transfer (28) or spin echo intensity (double electron-electron resonance; DEER) (29) are powerful but are limited in their ability to determine an ensemble by the complex relationships between the measured values and the desired probe-probe distances. These complications amplify the uncertainty of determining an average value and introduce even greater uncertainty in determining a distance distribution and the underlying conformational ensemble.

Prior results using the DNA double helix as a model experimental system (20) strongly suggest that detailed and quantita-

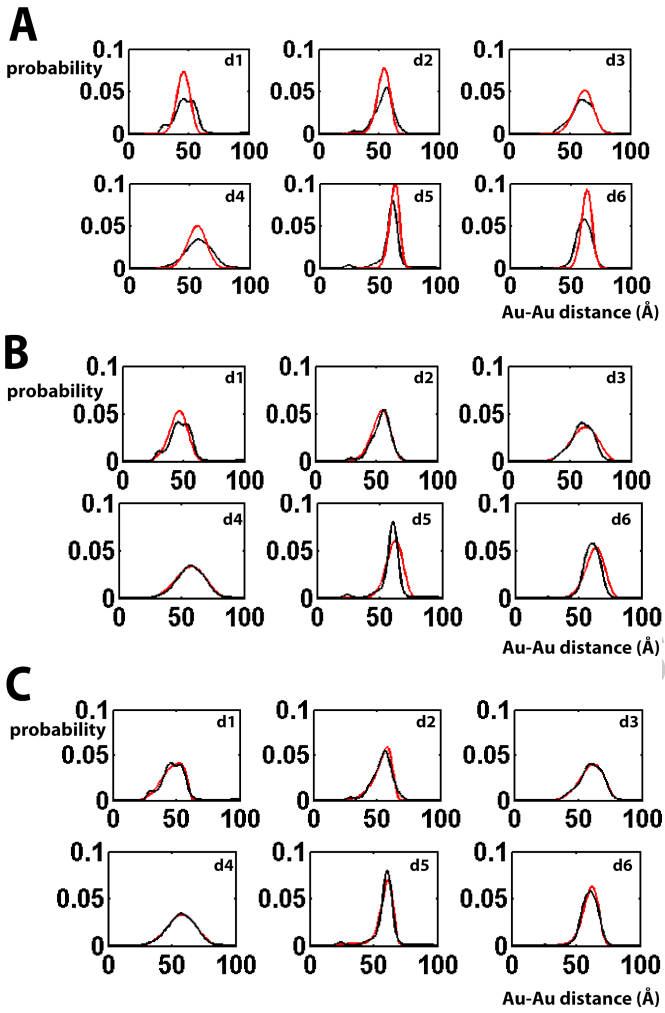

Fig. 3. Fitting of Au-Au distance distributions with a series of physical models. Experimental Au-Au center to center distance distributions (black lines) are plotted together with the best-fit model prediction (red lines) using a one-state approximation (**A** and **B**) or a non-constrained multiple-state ensemble model (**C**). In the one-state approximation models the ensemble is assumed to only contain a single conformation (**A**) or elastic fluctuations around a single conformation (**B**). In the non-constrained multiple-state model (**C**), there is no prior assumption of the ensemble (see "A Procedure for Building the Ensemble" for detail).

tive information about solution ensembles can be obtained. For the DNA helix, x-ray interferometry distance distributions were found to quantitatively agree with consensus elastic parameters of DNA and also revealed new properties of free DNA helices in solution (20). Nevertheless, the ensemble of a DNA double helix is simpler than that for most macromolecules and could be well described by broadening from a single conformation using an elastic potential. The ensembles of most biological macromolecules are likely to contain substantial anharmonicities and multiple local free energy minima.

To further test x-ray interferometry as a general method for probing macromolecule equilibrium ensembles and to determine fundamental properties of basic nucleic acid structures, we have applied x-ray interferometry to a nucleic acid helix-junction-helix (HJH) motif, the DNA bulge. DNA bulges can provide a model for the RNA bulges that are more commonly encoded in biology, and can be used to engineer nanostructures (30, 31). We chose the A-bulge DNA system for this study to allow comparison to a prior smFRET study that provided models for the average structures of these DNAs (32).

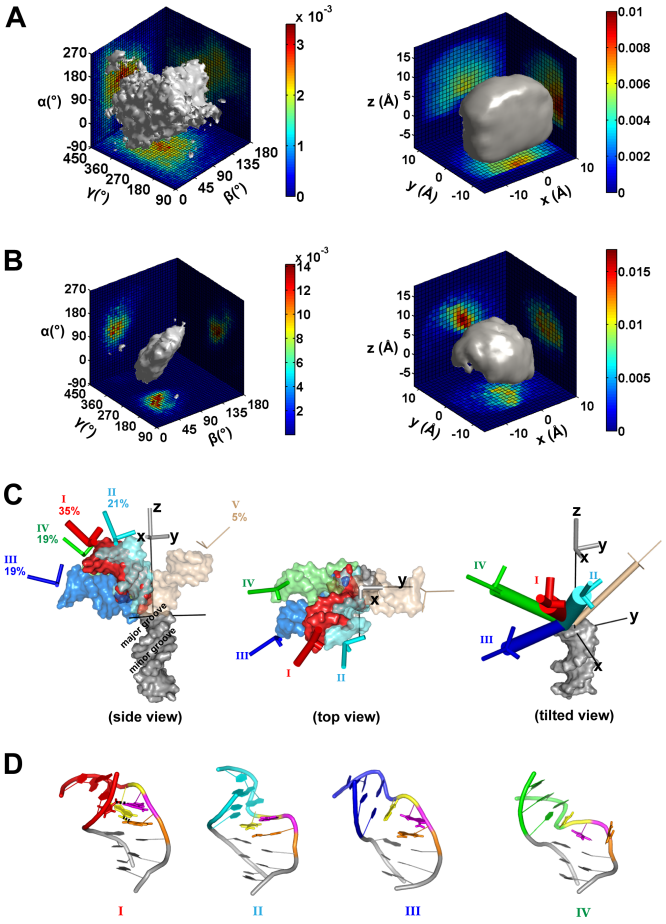

Fig. 4. Model for the 3A-DNA ensemble derived from x-ray interferometry data. (**A**) The geometrically allowed conformational space of the 3A-DNA helices (also see *SI Appendix*, *SI Methods*). The grey surface encloses 80% of the total population. The sharp edge in the right panel is the result of limiting the allowed space within a cube, which was defined by a set of MD conformations to estimate limiting x, y and z values and then extending this by 1 Å to provide a more conservative limit (see *SI Appendix, SI Methods for detail; also see SI Appendix,* Fig. S22)." (**B**) The estimated conformational ensemble of 3A-DNA obtained by reweighting the allowed space ensemble in panel A using the x-ray interferometry data (see Materials and Methods for details). The grey surface encloses 80% of the total population. (**C**) Three-dimensional view of representative conformers of each of the five groups of conformations (*SI Appendix*, Fig. S10). The conformer shown for each group is the one that is closest to the mean of the group in terms of (α, β, γ, x, y, z) (*SI Appendix*, Table S3). (**D**) Atomistic models of the representative bulge conformations for groups I to IV. The three adenosine residues in the bulge are colored in orange, magenta and yellow in the order from the 5′ to 3′ end of that DNA strand.

Results and Discussion

Bulges Bend DNA Helices and Broaden Their Conformational Ensembles

We first studied a series of bulged DNA helix constructs, with the bulge consisting of an increasing number of adenosine residues (0, 1, 3 and 5 adenosine residues; (32)). To use x-ray interferometry to investigate the ensemble of the bulge series, gold-nanocrystal probes were introduced site specifically through amino-modified thymine using an SPDP (N-Succinimidyl 3-[2-pyridyldithio]-propionate)-based linker, following our prior procedures (19, 20). To facilitate comparison of the different bulge constructs, the gold-nanocrystal probe pairs spanned the bulge and were placed at the same positions for each of the bulge constructs (Fig. 1A). The Au-Au scattering interference pattern for each construct was measured (Fig. 1C) and the resulting

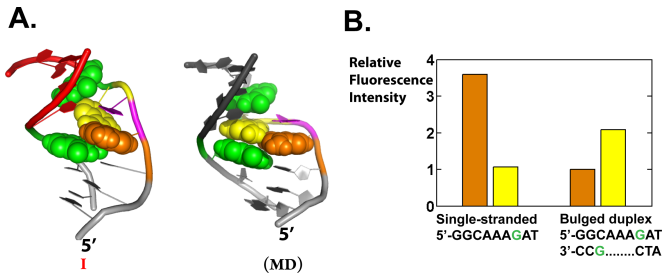

Fig. 5. 2-Aminopurine fluorescence to test the atomic-level models for 3A-DNA. (**A**) The relative position of the two flanking guanines (green) and the 5′ (orange) and 3′ (yellow) bulge adenine in the experimental conformer I (left) and the most stable MD conformer (right). The central A is magenta and not rendered space filling (also see *SI Appendix*, Table S4). (**B**) 2-Aminopurine intensity for DNA constructs with 5′ (orange bar) or 3′ (yellow bar) bulge adenine replaced with 2-aminopurine. See Materials and Methods for experimental conditions. The 3′-A (yellow) is less quenched than the 5′-A (orange) in the bulged duplex and the 3′-A is less quenched in the bulge than in a single strand, consistent with model I (**A**, left) and not expected for the structure corresponding to the MD model in which there is extensive stacking of both A residues (**B**, right). See also *SI Appendix, SI Text* and Table S4.

interference patterns were converted into distance distributions (Fig. 1D), again following previously published procedures (19, 20).

As expected, the average Au-Au distance decreases with increasing bulge length (Figs. 1D and 1E), consistent with increased bending. The observed decrease in distance with increasing bulge size reasonably matches the inferred distances from the literature smFRET average structures (32) (Fig. 1E). This agreement provides further support that x-ray interferometry is able to provide high-resolution average structural information, as it did in determining the average rise and twist per base of the DNA helix in solution (20). Nonetheless, the prior smFRET data (32) do not provide information on the nature of the ensemble. We found that the width of the pairwise distance distributions increases upon introduction of the bulge and further increases with increasing bulge length (Fig. 1D), as can be represented in terms of the variance of the distance distribution (Fig. 1F). This result suggests that bulged DNA has a broader ensemble than that of a continuous duplex and that this distribution broadens as the number of residues that are not restricted in motion by base-paired neighbors increases –i.e., as the number of single-stranded residues increases.

Beyond probing the extent of ensemble broadness (i.e., the variance), these distance distributions provide previously unavailable information about the ensemble. For example, the shape of the distance distribution for DNA with the 1A bulge (1A-DNA; red, Fig. 1D) is asymmetrical compared to that of a regular duplex (blue, Fig. 1D), which immediately suggests that the conformational ensemble of the 1A bulge cannot be represented by a single harmonic potential in its free energy landscape. [An isotropic broadening around a single stable conformation would be expected to broaden the position of gold probes in all directions and result in largely symmetrical broadening of the Au-Au distance distribution, as is the case for the duplex (blue, Fig. 1D, see also ref. (20)).] Thus, the highly asymmetrical small shoulder for 1A-DNA (Fig. 1D, dashed arrow) strongly suggests the presence of at least one minor conformer family. The position of the minor peak also provides information on the nature of this family. Its smaller probe-probe distances compared to that of a straight helix can in principle arise from bending, such that the top helix bends towards the bottom helix, or twisting, such that the top helix twists clockwise to bring the two probes closer (clockwise rotation of the yellow sphere in the right panel of Fig. 1B). We can rule out the twist-only model as twisting could only reduce the mean probe-probe distance from 70.7 Å (for the straight duplex) to about 67 Å (for a twist sufficient to align the nanocrystals directly above and below one another), and not to the observed distance of around 50 Å (Fig. 1D, dashed arrow).

Although this single probe pair provides incisive information about the conformational ensemble, there are also major limitations of the information provided. Using the above discussion as an example, a bend-only model can account for the minor peak but so can a family of models with successively less bending and more twisting. The minor peak with a mean Au-Au distance of 50 Å (Fig. 1D), could arise from bending alone with an angle of 58°, from a smaller bend of 53° together with a twist of 58°, or from a continuous series of intermediate bend and twist angle pairs. Measuring distances between additional probe pairs is needed to remove degeneracies and to obtain the molecule's conformational ensemble. We chose to pursue DNA with the 3A bulge (3A-DNA) because it has a broader and potentially more complex conformational ensemble than the 1A bulge (Figs. 1D and 1F).

Estimating the 3A Bulge DNA Conformational Ensemble

Construct Design and Experimental Results. To estimate the conformational ensemble of 3A bulge DNA (3A-DNA), we labeled the flanking helices with six different pairs of gold nanocrystals (Fig. 2A). Six scattering interference patterns were measured; one for each gold pair (Fig. 2B). As noted above, each scattering profile contains interference intensities across the range of measured scattering angles (s) and not just a single intensity (Fig. 2B). Each profile has a range of Au-Au distances (Fig. 2C), which correspond to the full distribution of pairwise distances from each member of the ensemble. This experimental ensemble information is then used to weight a large set of potential bulge conformations, generated through simple geometric modeling, to obtain an ensemble model for 3A-DNA, as described in the following sections and in *SI Appendix*, *SI Methods*.

Testing the Null Model: Are Multiple States Required to Account for the Conformational Ensemble of 3A-DNA? We first determined if the interference data can be accounted for by a single conformational family, with a single energy minimum and a simple elastic free energy potential akin to that for a simple DNA duplex (20). We generated a pool of ~5 x 10^4 geometrically allowed conformations to extensively sample the allowed conformational space (see "A Procedure for Building the Ensemble" below and *SI Appendix*, *SI Methods* for details). Each conformation corresponds to a unique position of the top helix relative to the bottom helix (Fig. 1B), described by a set of rotational (α, β, γ) and translational (x, y, z) parameters.[1] The data were fit with a single bulge conformation in the allowed space (Fig. 3A and *SI Appendix*, Fig. S1A, $\chi^2 = 12$, $\bar{x}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$ where N = 1898 is the total number of measured data points combined over all six probe pairs) and with an elastic expansion from this single bulge conformation (Fig. 3B and *SI Appendix*, Fig. S1B, $\chi^2 = 4.9$), with the (α, β, γ, x, y, z) of the ensemble conformations normally distributed around the values of (α, β, γ, x, y, z) of the single best-fit conformation. The dispersion of the six parameters, (α, β, γ, x, y, z), were changed independently. These fits, constrained to a single state, gave a nearly twofold larger χ^2 compared to the unconstrained fit described below (Fig. 3C and *SI Appendix*, Fig. S1C, $\chi^2 = 2.6$

[1] A bulge conformation (α, β, γ, x, y, z) is generated by first rotating the top helix by (α, β, γ), followed by a translation of (x, y, z). We use the zyz Euler convention (33), in which the order of rotation is as follows: a clockwise rotation of α along the z axis when viewed from above; a bend of β towards the negative x-axis (i.e., clockwise rotation around the y-axis); and finally a clockwise rotation of γ around the z-axis, with γ = 0°, 90°, 180° and 270° corresponding to the x-, y+, x+ and y- directions, respectively. A slightly different definition of the six-dimensional conformational space of HJH was previously used by Bailor et al. (34); see *SI Appendix*, *SI Methods* for details.

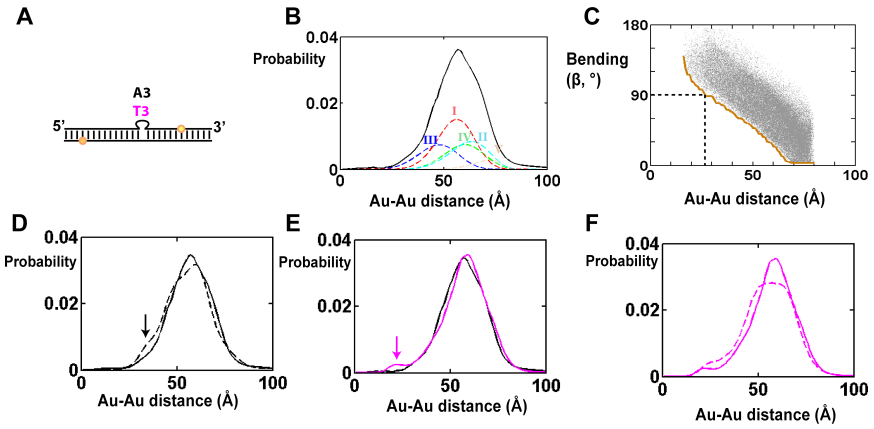

Fig. 6. The effect of sequence composition and ionic conditions on the bulge-DNA ensemble. **(A)** Schematic of the constructs used with bulges of 3A or 3T and Au probe pair d4 as defined in Fig. 2A, the position that is most sensitive to bending. **(B)** The measured Au-Au center to center distance distribution for 3A-DNA with probe pair d4 (black, also in Fig. 2C) and the individual contributions from the five conformational groups (I to V, colored dashed lines (Fig. 4C and *SI Appendix*, Fig. S10)). **(C)** The relationship between d4 Au-Au distances and the possible range of bending angles, β, in the allowed space. Each dot corresponds to a conformation in the allowed space (see also *SI Appendix*, Fig. S13 for a density map); the x-axis is the mean distance of the d4 Au-Au center to center distance distribution of each allowed space conformation. The minimum bending angle at the same d4 Au-Au distance is labeled as a brown line. **(D)** Comparison of the d4 Au-Au center to center distance distributions for 3A-DNA bulge without (solid line) and with (dashed line) 4 mM Mg^{2+} in the background of 160 mM Na^{+}. **(E)** Comparison of the d4 Au-Au center to center distance distributions for 3A-DNA (black) and 3T-DNA (magenta). **(F)** Comparison of the d4 Au-Au center to center distance distributions for 3T-DNA without (solid line) and with (dashed line) 4 mM Mg^{2+} in the background of 160 mM Na^{+}.

for a non-constrained model allowing multiple states), suggesting that the 3A-DNA ensemble is not well described by a single free energy minimum.

To explore the information content of the data beyond a single state, we fit the data with increasing number of discrete conformations. As expected, the fitting improved as the number of discrete conformations allowed in an ensemble was increased. The improvement in fitting becomes negligible beyond about five conformations (*SI Appendix*, Fig. S2). In other words, the distance distributions contain enough information to distinguish about five discrete types of conformations in an ensemble. We emphasize that an ensemble of a few discrete conformations is not a physically realistic model, as it assumes no motion of the bulge around each discrete conformers. Similar to NMR-based ensemble methods, a realistic but underdetermined ensemble model can be generated by refining computationally generated ensemble models with the x-ray interferometry data, as described in the next section (7). We envision that a combination of multiple techniques will improve the confidence in the ensemble models generated in the future.

A Procedure for Building the Ensemble. We briefly outline the procedure we used to build an ensemble for the 3A-DNA here, and provide a more detailed description in *SI Appendix, SI Methods*. We first generated an allowed space, as alluded to above, using simple geometric modeling. MD simulations were used to set boundary conditions but were not directly involved in the modeling (see *SI Appendix*, *SI Methods* for details). We systematically sampled the six-dimensional conformational space (α, β, γ, x, y, z) to generate an initial pool of about 1.8×10^6 conformations, and then eliminated conformations that result in steric clashes of the top and bottom helices and conformations with unreasonably long distances between the connections of the two helices, leaving a total of $\sim 5 \times 10^4$ conformations, representing our basis set of the geometrically allowed conformational space of the bulge (Fig. 4A; (35, 36)).

The allowed space represents a prior model of the ensemble where all conformations in the allowed space have equal probability. An ensemble model is a specific set of probabilities of the allowed space conformations. The experimental scattering profile of an ensemble model, I(S), can be predicted from the weighted sum of the expected scattering profile of individual bulge conformations: $I(S) = \sum_{i=1}^{50k} w_i I_i(S)$. The probability that a given ensemble model is the actual ensemble would then be related to the level of agreement between the predicted and measured scattering profiles, $I_{exp}(S)$. The most probable ensemble model, or the optimum set of w_i weights of the basis set conformations, was estimated using a simplified Bayesian method (37, 38) (see Materials and Methods and *SI Appendix, SI Methods f*or details). The procedure sums over different ensemble models weighted by their likelihood given the experimental data, which was calculated from χ^2 statistics. This weighted summation combines reasonable individual ensemble solutions to provide an estimate of the overall ensemble.

Using the procedure described above, a good fit to the data was obtained (Fig. 3C and *SI Appendix*, Fig. S1C; $\chi^2 = 2.6$), and the corresponding conformational ensemble is illustrated in Fig. 4B. As a test of this ensemble model and the bulge not being disturbed by Au labeling, we used the estimated ensemble to predict (39) the SAXS profile of the unlabeled bulge constructs and determine if it agrees with the measured profile. Good agreement was found between the predicted and measured profiles (*SI Appendix*, Fig. S20, $\chi^2 = 0.8$). We further carried out a cross-validation test of our ensemble model and our fitting procedure, where each of the sets of five distance distributions were used to predict the sixth distribution. The six predicted distributions strongly resembled the corresponding experimental distributions (*SI Appendix*, Fig. S4), and the six conformational ensembles generated using each of the sets of five of the six distance distributions closely resemble that of the conformation ensemble obtained using all six distance distributions (*SI Appendix*, Figs. S5 & S6). This agreement suggests a relatively high confidence in the regions of the conformational space that are populated by the 3A-DNA. The standard deviation of these six ensembles also provides a crude estimate of the error of our 3A-DNA conformational ensemble model (*SI Appendix*, Figs. S6C).

The 3A-DNA Conformational Ensemble

The conformational ensemble of the 3A-DNA, estimated via the approach described in the previous section, is shown in Euler space (Fig. 4B, left) and in translational space (Fig. 4B, right). This ensemble populates a much restricted space compared to the allowed space (Fig. 4A). It will be of great interest to de-

termine the variation in conformational restrictions for different helix-junction-helix (HJH) elements, relative to the maximum covalently and geometrically allowed space, to determine how large this contribution is to folding and how variable it is between junction motifs.

To describe the ensemble, we divided the ensemble into five groups of conformations, labeled I to V, based on the experimentally measureable properties of the conformations –i.e., their Au-Au distance profiles (see *SI Appendix*, Fig. S10 and Table S3). To better visualize the ensemble, representative conformers from the five clusters are shown in Fig. 4C. The four dominant groups of the 3A-DNA ensemble (I to IV, 95%) all bend toward the major groove side of the bottom helix (Fig. 4C, left), mostly within the +x/-y quadrant (Fig. 4C, middle) with the 90% of the bending angles between 24 and 85° (Fig. 4C left, *SI Appendix*, Fig. S10B). More detailed descriptions of the ensemble are presented in *SI Appendix, SI Text*.

To explore the potential molecular interactions that could be responsible for stabilizing different bulge conformers and shaping the ensemble, we used molecular dynamics (MD) to generate atomistic models of the bulge conformations that are consistent with the conformations obtained from the x-ray interferometry-generated ensemble (Fig. 4D). The MD atomistic model for the most abundant region of the ensemble was found to be consistent with results from a 2-AP fluorescence assay (Fig. 5; see *SI Appendix, SI Text* for details).

X-ray Interferometry to Probe the Effects of Bulge Sequence and Ionic Conditions on the Bulge-DNA Ensemble

To probe the sensitivity of a HJH ensemble to ionic conditions and junction sequence and to further probe the ability of x-ray interferometry to distinguish conformational ensembles, we tested the effect of adding Mg^{2+} and of changing the bulge sequence from 3A to 3T. In particular, we assessed their effects on the small population of extremely bent conformers ($\beta > 90°$) that is part of group III.

The 3A-DNA ensemble obtained from the interferometry data exhibits limited bending compared to the allowed space: conformers with bends (β) of greater than 90° made up 29% of our allowed space (Fig. 4A and *SI Appendix*, Fig. S9A), but these conformers populate only 4% of the actual ensemble (Fig. 4B). Two simple factors could contribute to the limited bending in 3A-DNA. Stacking in the bulge and electrostatic repulsion between the helices in the highly bent conformations would tend to favor a roughly continuous arrangement of the helices. If base stacking were important in limiting bending, then the reduced stacking with the 3A bulge replaced with 3T would be expected to increase occupancy in the extremely bent region ($\beta > 90°$). If electrostatic repulsion were important, screening by added Mg^{2+} would reduce this repulsion and likewise be expected to increase occupancy of this region.

We tested these models using the d4 gold labeling pair (Fig. 2A & Fig. 6A), as this labeling pair is highly sensitive to bending (Fig. 6C & *SI Appendix*, Fig. S13). Figure 6C shows the bending angles for all conformers that share the same d4 distance. As d4 distance gets smaller, the values of β becomes larger, and distances of <27 Å require that β be greater than 90° (Fig. 6C, dotted line). Thus, populations of strongly bent conformations can be probed by the population at small d4 distances (Fig. 6C and *SI Appendix*, Fig. S13).

Addition of 4 mM Mg^{2+} had no significant effect on Au-Au distance distributions of the DNA helix lacking bulged residues (*SI Appendix*, Fig. S14) but did alter the distribution for A3-DNA (Fig. 6D solid vs dashed lines). Upon addition of Mg^{2+} there is a small but measureable increase in conformers with d4 distances of about 33 Å (Fig. 6D, arrow). This distance corresponds to strongly bent conformations with β values of 80° or greater (Fig. 6C, brown line). The modest magnitude of this change is consistent with Poisson–Boltzmann (PB) model predictions. We estimated that an addition of 4 mM Mg^{2+} would preferentially stabilize the strongly bent ($\beta > 90°$) over the less bend (β = 30-90°) bulge conformations by about 0.2-0.5 kcal/mol (*SI Appendix*, Fig. S15 and *SI Methods*), which would correspond to stabilization and enrichment of the strongly bend conformations by about 30-110%. PB tends to underestimate the effects of divalent cation screening (40-42), and the observed ~2 fold effect (Fig. 6D) is at the upper end of the predicted range.

Changing the base sequence from 3A to 3T led to a new d4 distance peak at 23 Å (Fig. 6E, magenta line & arrow), consistent with increased bending, which could arise via a reduction of stacking, different hydrogen bonding patterns, and/or a reduction of steric constraints from the purine bases. Conformers that give rise to a peak at 23 Å require β values of >100° (Fig. 6C, brown line). We also added 4 mM Mg^{2+} with the 3T bulge to additionally test electrostatic factors and observed an increase in populations with small d4 distances (<50 Å, Fig. 6F), as expected from enhanced electrostatic screening, though not an increase of the small fraction of the most bent conformers.

The effects from changing the bulge sequence and increasing electrostatic screening are small, but they are readily detected by appropriately placed gold nanocrystals, providing support for the above-noted models and the sensitivity of x-ray interferometry. Future experiments, in conjunction with atomic level models, will be required to dissect the origins of these effects in greater detail.

It is of interest to compare our results on DNA bulge and literature results on the average bending of RNA bugles. Transient electric birefringence (TEB) and gel mobility results (43) suggest that a 3U-RNA bulge is slightly less bent than a 3A-RNA bulge. Our data, although revealing an increase in a strongly bent subpopulation, are consist with a slightly less bent average conformation, as the major peak (larger peak on the right, Fig. 6E) of 3T-bulge DNA (Fig. 6E, magenta) is slightly shifted to longer distance compared to the 3A-bulge DNA (Fig. 6E, black). Conversely, Zacharias and Hagerman (43) observed a decrease in bending for 3U-RNA with an increase in Mg^{2+} concentration, whereas our data suggest that 3T-DNA is slightly more bent in the presence of Mg^{2+}. This difference could result from specific metal binding to the RNA but not DNA bulge or from stacking or other differences between this DNA and RNA (44, 45). The ability of x-ray interferometry to probe beyond structure averages should render this technique particularly valuable in determining the origin of such differences.

Conclusions and Implications

Determining a molecule's conformational ensemble, beyond that of an average structure, is a major and necessary step towards a predictive and quantitative understanding of macromolecule structure, folding, and function. This task is extremely challenging even for simple helix-junction-helix (HJH) motifs, which represent the building blocks of complex nucleic acids structure and an ideal starting point for developing such approaches (14, 46). The average structural information from single-molecule FRET experiments with bulged DNAs (32) was reproduced by x-ray interferometry (Fig. 1 and *SI Appendix*, Fig. S16) and extended to provide incisive information about the ensemble of conformers present that could not be obtained from FRET.

We obtained information about the conformational ensembles of HJH motifs with increasing numbers of A residues and we estimated the 3A-DNA conformational ensemble. The 3A-DNA ensemble populates a limited region of its geometrically allowed conformational space. Predominant in the ensemble are conformers with bend angles ranging between 24 and 85° and with characteristic bending directions and helical twists as well as helical displacements. X-ray interferometry also revealed changes to the conformational ensemble from perturbations in

ionic conditions and the bulge sequence, and our results suggest that stacking and electrostatics limit bending in 3A-DNA.

X-ray interferometry instantaneously assays distances and allows a direct transformation from interference pattern to distance. Thus, x-ray interferometry can be readily predicted from MD simulations and should therefore be valuable in testing and refining MD-based models. A community-wide competition for predicting flexible nucleic acids structures using x-ray interferometry as experimental benchmark would be a powerful addition to the current RNA CASP (47), which currently predicts only folded RNA structures; obtaining the correct weighting of an ensemble of structures is a more stringent test than the correct prediction of a single most-stable structure and is necessary for understand and effectively predict thermodynamics and kinetics.

X-ray interferometry provides elusive ensemble information of macromolecules and complements existing NMR-based approaches. X-ray interferometry readily provides information about translational displacements, which are difficult to assess with RDC measurements (48), as well as angular movements, and it is more straightforward to extend to larger structures (24, 26) and other classes of macromolecules. However, x-ray interferometry is limited in detecting rare conformers (19, 20, 25) so that techniques that can trap or assess rare excursions, such as NMR relaxation dispersion (49), paramagnetic relaxation enhancement (50), H/D exchange (51) and cyclization (52) are powerful complements to x-ray interferometry. Full atomic-level resolution of conformational ensembles and free energy landscapes of macromolecules will require continued synergy between the development of x-ray interferometry, RDC measurements, and other experimental techniques as well as MD and other computations approaches.

Materials and Methods

Materials. Au-labeled DNA oligonucleotides were prepared following procedures described previously (19, 20). Briefly, the DNA oligonucleotides were synthesized using ABI 394 DNA synthesizer, and purified by Poly-Pak (Glen research) followed by anionic exchange HPLC. Internal thiol groups were introduced through derivatization at amino-allyl dT (Glen research) using succinimidyl 3-(2-pyridyldithio)propionate (SPDP, Pierce), followed by DTT reduction and desalting. The thiolated DNA oligonucleotides were coupled to thioglucose-passivated gold nanocrystals (19) for 2 h at pH 9.0, purified by anion exchange HPLC, and desalted by centrifugal buffer exchange with water. Complementary strands were annealed at room temperature for 30 minutes, then purified and desalted as above. 2-Aminopurine modified DNA oligonucleotides were synthesized as above and purified by anionic exchange HPLC.

SAXS Measurements and Data Processing. Small-angle x-ray scattering measurements were carried out at the Stanford Synchrotron Radiation Lab (SSRL, Beamline 4-2) using a sample to detector distance of 1.7 meters. The buffer conditions for all experiments are 150 mM NaCl, 70 mM Tris-HCl, pH 8.0, with 10 mM sodium ascorbate. Data were acquired and analyzed following procedures described previously (20, 25).

Predicting the Au-Au Distance Distribution and Scattering Profile for a Bulge Conformation. Each bulge conformation was identified by a set of (α, β, γ, x, y, z) and the equivalent matrix M. The rotation and translation indicated by (α, β, γ, x, y, z) is equivalent to applying the rotational and translational matrix M to the top helix at (0, 0, 0, 0, 0, 0), a standard duplex. At (0, 0, 0, 0, 0, 0), the position of the gold probe on either the bottom or the top helix is not a point but a point cloud, due to internal conformational fluctuation of the DNA duplex (20). For a pair of gold probes j, one on the bottom helix and one on the top helix, we generated the two probe position clouds at (0, 0, 0, 0, 0, 0) based on our prior DNA double helix model (20). For a bulge conformation i, the bottom helix gold position cloud is unchanged and the top helix cloud can be calculated by applying the rotational and translational matrix Mi to the top helix cloud at (0, 0, 0, 0, 0, 0). The pairwise distance distribution between the bottom and top helix gold clouds gives rise to the Au-Au distance distribution for gold pair j and conformer i. A Au-Au distance distribution ij can be transformed back to its equivalent Au-Au scattering interference profile, Ii,j(S), as previously described (20, 25).

Estimating the Conformational Ensemble. The conformational ensemble is determined using a procedure simplified from a published Bayes approach (38). A detailed description of the procedure is included in *SI Appendix*, *SI Methods*. Briefly, the optimum weight for i = 1 to N conformers in the allowed space was calculated as

$$w_i = \int w_{i,E}\, f(m|E)\, dE \qquad ,1$$

where the weight for conformer i in each ensemble solution E, $w_{i,E}$, were averaged over all E and weighted by f(m|E), the probability of obtaining data m with ensemble solution E, which can be determined using χ^2 statistics. Here data m is I(S), the scattering angle dependence in scattering intensity.

Due to the complexity of the ensemble solution space, Eq. 1 cannot be solved by directly sampling the entire ensemble solution space as in ref. (38). Instead, we simplified by sampling the ensemble solution space in hierarchical stages and used a method similar to empirical Bayes approximation (37) in which a smaller sub-ensemble solution space is represented only by its maximum likelihood solution. Specifically, we first divided the ensemble space into smaller sub-ensemble spaces. We randomly select 100 conformers out of the 50k allowed conformational space to be allowed to have non-zero weights. These 100 conformers can have different weight vectors and in itself is a sub-ensemble solution space. We then approximated this sub-ensemble solution space with its maximum likelihood solution, the set of optimum weight that maximize f(m|E), which was determined using the lsqnonneg function of Matlab. The above sampling procedure was repeated by randomly select different 100 conformers from the 50k allowed conformational space. Convergence was found among five separate 200k step samplings. The five 200k step samplings were combined to calculate the final weight vector using Eq. 1, where each randomly selected sub-ensemble space E is weighted by f(m|E).

Molecular Dynamics Modeling of the Bulge. Molecular dynamics (MD) simulations were used to generate a library of plausible bulge conformations and provide a MD estimate of the bulge ensemble. Simulations were performed using Gromacs 4.5.5 (53) and the AMBER99 force field (54). The simulated construct included three bulged A's flanked by three base pairs at each end of the helix. Distance restraints were used to fix the secondary structure of each closing base pair and to enforce the helicity of the adjacent bases in the non-bulge strand. The TIP3P explicit solvent model was used to model water. PME was used to calculate electrostatic forces. Simulations were performed in a cubic box of length 55.9 Å. The box contained DNA with 13 backbone phosphates, 5600 water molecules, 29 sodium ions, and 16 chloride ions. Multiple simulations were performed at 288 K with the temperature controlled by a Langevin integrator. A total of ~2 μs of simulation was performed; conformations were saved every 100 ps, leading to a total of ~200,000 conformations in the resulting bulge conformation library.

For each MD bulge conformer, the corresponding (α, β, γ, x, y, z) was determined as described in *SI Appendix*, *SI Methods*. The d1 to d6 (Fig. 2A) Au-Au distances were then predicted as described above ("Predicting the Au-Au Distance Distribution and Scattering Profile for a Bulge Conformation"). For each representative x-ray interferometry conformer (Fig. 4C), the closest matched MD conformer was found as the MD conformer with the smallest d1 to d6 RMSD from the x-ray interferometry conformer. This MD conformer then provides an atomistic structure model for an x-ray interferometry conformer (Fig. 4D).

Fluorescence Measurement Steady-state fluorescence intensity of 2-aminopurine-modified DNA was measured using a Fluorolog-3 spectrometer (Horiba) with excitation and emission wavelengths set to 320 and 380 nm, respectively. Fluorescence lifetimes were measured using an Easylife fluorometer (OBB). All measurements were carried out with 150 mM NaCl and 20 mM Tris-HCl, pH 8.0, at 20 °C (steady state) or room temperature (time resolved).

ACKNOWLEDGMENTS.

We thank H. Tsuruta (deceased), T. Matsui and T. Weiss at beamline 4-2 of the Stanford Synchrotron Radiation Lab (SSRL) for technical support in synchrotron small angle X-ray scattering experiments, members of the Herschlag and the Harbury labs, R. Das, and R. S. Mathew for helpful discussions, and an anonymous reviewer for suggest the control in Fig. S20. This work was supported by NIH grants DP-OD000429-01 (P.B.H.) and PO1 GM066275 (D.H.).

1. Stelzer AC, *et al.* (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7(8):553-559.
2. Lee GM & Craik CS (2009) Trapping moving targets with small molecules. *Science* 324(5924):213-215.
3. Al-Hashimi HM & Walter NG (2008) RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* 18(3):321-329.
4. Nguyen P & Qin PZ (2012) RNA dynamics: perspectives from spin labels. *Wires Rna* 3(1):62-72.
5. Clore GM (2011) Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation. *Protein Sci.* 20(2):229-246.
6. Roy R, Hohng S, & Ha T (2008) A practical guide to single-molecule FRET. *Nat. Methods* 5(6):507-516.
7. Fisher CK & Stultz CM (2011) Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 21(3):426-431.
8. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, & Vendruscolo M (2005) Simultaneous determination of protein structure and dynamics. *Nature* 433(7022):128-132.

9. Cantor CR & Schimmel PR (1980) *Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function* (W. H. Freeman and Company, New York).
10. Dosset P, Hus JC, Marion D, & Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* 20(3):223-231.
11. Salmon L, *et al.* (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* 132(24):8407-8418.
12. Schwieters CD & Clore GM (2007) A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry-Us* 46(5):1152-1166.
13. Frank AT, Stelzer AC, Al-Hashimi HM, & Andricioaei I (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res.* 37(11):3670-3679.
14. Zhang Q, Stelzer AC, Fisher CK, & Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450(7173):1263-1267.
15. Salmon L, Bascom G, Andricioaei I, & Al-Hashimi HM (2013) A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. *J. Am. Chem. Soc.* 135(14):5457-5466.
16. Clore GM (2011) Exploring sparsely populated states of macromolecules by diamagnetic and paramagnetic NMR relaxation. *Protein Sci.* 20(2):229-246.
17. Bailor MH, *et al.* (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat Protoc* 2(6):1536-1546.
18. Kruschel D & Zagrovic B (2009) Conformational averaging in structural biology: issues, challenges and computational solutions. *Mol Biosyst* 5(12):1606-1616.
19. Mathew-Fenn RS, Das R, & Harbury PA (2008) Remeasuring the double helix. *Science* 322(5900):446-449.
20. Shi X, Herschlag D, & Harbury PA (2013) Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. *Proc Natl Acad Sci U S A* 110(16):E1444-1451.
21. Vainshtein BK, *et al.* (1980) Determination of the distance between heavy-atom markers in hemoglobin and histidine-decarboxylase in solution by small-angle X-ray-scattering. *FEBS Lett.* 116(1):107-110.
22. Capel MS, Kjeldgaard M, Engelman DM, & Moore PB (1988) Positions of S2, S13, S16, S17, S19 and S21 in the 30-S-ribosomal subunit of Escherichia-coli. *J. Mol. Biol.* 200(1):65-87.
23. Miakelye RC, Doniach S, & Hodgson KO (1983) Anomalous X-ray-scattering from terbium-labeled parvalbumin in solution. *Biophys. J.* 41(3):287-292.
24. Hura GL, *et al.* (2013) DNA conformations in mismatch repair probed in solution by X-ray scattering from gold nanocrystals. *P Natl Acad Sci USA* 110(43):17308-17313.
25. Mathew-Fenn RS, Das R, Silverman JA, Walker PA, & Harbury PAB (2008) A molecular ruler for measuring quantitative distance distributions. *Plos One* 3(10) e3229(10).
26. Mastroianni AJ, Sivak DA, Geissler PL, & Alivisatos AP (2009) Probing the conformational distributions of subpersistence length DNA. *Biophys. J.* 97(5):1408-1417.
27. Lipfert J & Doniach S (2007) Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* 36:307-327.
28. Hochstrasser RA, Chen SM, & Millar DP (1992) Distance distribution in a dye-linked oligonucleotide determined by time-resolved fluorescence energy-transfer. *Biophys. Chem.* 45(2):133-141.
29. Jeschke G (2012) DEER distance measurements on proteins. *Annu Rev Phys Chem* 63:419-446.
30. Lilley DMJ (1995) Kinking of DNA and RNA by base bulges. *P Natl Acad Sci USA* 92(16):7140-7142.
31. Gohlke C, Murchie AIH, Lilley DMJ, & Clegg RM (1994) Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy-transfer. *P Natl Acad Sci USA* 91(24):11660-11664.
32. Wozniak AK, Schroder GF, Grubmuller H, Seidel CA, & Oesterhelt F (2008) Single-molecule FRET measures bends and kinks in DNA. *Proc Natl Acad Sci U S A* 105(47):18337-18342.
33. Goldstein H (2002) *Classical mechanics* (Addison Wesley, San Francisco :) 3rd ed. Ed.
34. Bailor MH, Mustoe AM, Brooks CL, & Al-Hashimi HM (2011) 3D maps of RNA interhelical junctions. *Nat Protoc* 6(10):1536-1545.
35. Bailor MH, Sun X, & Al-Hashimi HM (2010) Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* 327(5962):202-206.
36. Mustoe AM, Bailor MH, Teixeira RM, Brooks CL, 3rd, & Al-Hashimi HM (2012) New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation. *Nucleic Acids Res.* 40(2):892-904.
37. Carlin BP & Louis TA (2000) The empirical Bayes approach. *Bayes and Empirical Bayes Methods for Data Analysis, Second Edition*, (Chapman and Hall).
38. Fisher CK, Huang A, & Stultz CM (2010) Modeling intrinsically disordered proteins with bayesian statistics. *J. Am. Chem. Soc.* 132(42):14919-14927.
39. Schneidman-Duhovny D, Hammel M, Tainer JA, & Sali A (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* 105(4):962-974.
40. Anthony PC, *et al.* (2012) Electrostatics of nucleic acid folding under conformational constraint. *J Am Chem Soc* 134(10):4607-4614.
41. Bai Y, *et al.* (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J Am Chem Soc* 129(48):14981-14988.
42. Shklovskii BI (1999) Screening of a macroion by multivalent ions: Correlation-induced inversion of charge. *Physical Review E* 60(5):5802-5811.
43. Zacharias M & Hagerman PJ (1995) Bulge-induced bends in RNA - qantification by transient electric birefringence. *J Mol Biol* 247(3):486-500.
44. Olejniczak M, *et al.* (2002) The bulge region of HIV-1 TAR RNA binds metal ions in solution. *Nucleic Acids Res.* 30(19):4241-4249.
45. Ippolito JA & Steitz TA (1998) A 1.3-angstrom resolution crystal structure of the HIV-1 trans-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *P Natl Acad Sci USA* 95(17):9819-9824.
46. Zhang Q, Sun X, Watt ED, & Al-Hashimi HM (2006) Resolving the motional modes that code for RNA adaptation. *Science* 311(5761):653-656.
47. Cruz JA, *et al.* (2012) RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18(4):610-625.
48. Bax A & Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr Opin Struc Biol* 15(5):563-570.
49. Bouvignies G, *et al.* (2011) Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477(7362):111-114.
50. Clore GM, Tang C, & Iwahara J (2007) Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr Opin Struc Biol* 17(5):603-616.
51. Chamberlain AK, Handel TM, & Marqusee S (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Bio.* 3(9):782-787.
52. Vafabakhsh R & Ha T (2012) Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* 337(6098):1097-1101.
53. Pronk S, *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845-854.
54. Wang JM, Cieplak P, & Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21(12):1049-1074.