

**Supporting information for**

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

Xuesong Shi, Kyle A. Beauchamp, Pehr B. Harbury and Daniel Herschlag

Supporting Methods (page 2-5)

Supporting Text (page 6-9)

4 Supporting Information Tables (page 10-12)

22 Supporting Information Figures (page 13-30)

## Supporting Methods

**Defining the Allowed Space Through Statistical Sampling of the Six-dimensional Conformational Space.** The allowed space for 3A-DNA (Fig. 4A) was generated using direct statistical sampling of the six-dimensional,  $(\alpha, \beta, \gamma, x, y, z)$ , conformational space of the helices connected by the bulge. The three Euler angles  $(\alpha, \beta, \gamma)$  are defined in detail in “Procedure Used to Building the Ensemble” below. Briefly, any orientation of a helix can be generated from three sequential rotations starting from a helix along the  $z^+$  axis: a rotation of  $\alpha$  around the  $z$  axis, followed by a rotation of  $\beta$  around the  $y$  axis, and then followed by a rotation of  $\gamma$  around the  $z$  axis. The angles  $\alpha$  and  $\gamma$  were randomly sampled over the range of  $-180$  to  $180^\circ$  and  $\beta$  was sampled over the range of  $0$  to  $180^\circ$  with frequency  $\sin \beta$ . The translational space of  $(x, y, z)$  was also randomly sampled within a set of upper and lower boundary values of  $x, y,$  and  $z$  that limit sampling in the  $(x, y, z)$  space to a cube. The upper and lower boundary used,  $[9.2\ 6.5\ 14.1]$  and  $[-13.3\ -9.7\ -3.1]$  (in Å), respectively, were set at  $1$  Å beyond the boundaries determined by the MD-generated conformational library. The MD simulations were carried out not only at a realistic temperature (288 K, see Materials and Methods) but also at elevated temperatures of 375 K and 450 K (see Materials and Methods) to more exhaustively sample the geometrically allowed bulge conformational space. We generated our allowed space based on the limiting values in the multi-temperature MD simulations. Nonetheless, as the multi-temperature MD sampling of the allowed space might still be incomplete, we relaxed the MD-determined limiting values to cover a broader range of conformations and to achieve a more complete coverage of the allowed space.

After the initial statistical sampling of  $(\alpha, \beta, \gamma, x, y, z)$  described above, two additional criteria were used to select for geometrically allowed conformations. The first criteria were distances of less than  $23.2$  Å and  $8.5$  Å for the 3 nt and 0 nt bulge strands, respectively, between the two intra-strand C1'-C1' atoms of the residues of the base pairs immediately 5' and 3' to the bulge. These limits of  $23.2$  Å and  $8.5$  Å are  $0.5$  Å relaxed from  $22.7$  Å and  $8.0$  Å, the respective maximum C1'-C1' distances in the MD library generated from the multi-temperature MD simulations mentioned above. The second criterion is that there be no steric clashes between the top and the bottom helices. Surveying of the MD library indicated that the minimum pairwise distance between two non-H atoms, one on the top helix and one on the bottom helix, is  $2.1$  Å, and the maximum number of pairwise distances of less than  $3$  Å appears to be 7. We again relaxed the criteria relative to the MD limits and defined steric clashing as having any pairwise distance between the top and bottom helices of less than  $2$  Å or having more than 10 pairwise distances of less than  $3$  Å for two non-H atoms. We sampled a total of 1.6 million combinations of  $(\alpha, \beta, \gamma, x, y, z)$ ; about 3% of the 1.6 million conformations satisfied both of the steric criteria above. The final allowed space for 3A-DNA contained 52,347 conformations (Fig. 4A).

An allowed space was also generated using the MD limiting values without any relaxation or using criteria that are independent of MD. In the latter case, there is no pre-set boundary in  $(x, y, z)$ ; the upper limit of the 3 nt and 0 nt intra-strand C1'-C1' were increased to the maximum distances for fully extended chains of  $27.5$  and  $10.0$  Å, respectively, and steric clashes were identified as any pairwise non-H atom distance to be more than  $0.4$  Å smaller than the sum of their Van der Waal radii (S1). Using these alternative allowed spaces did not lead to large changes in the fitted ensemble (Fig. S17).

**Procedure Used to Build the Ensemble.** We start by defining a bulge conformation, which is the position of the top helix, the helix 3' to the bulge, relative to the reference bottom helix, the helix 5' to the bulge. To define this relative position requires a six-dimensional conformational space, including three rotational Euler angles,  $(\alpha, \beta, \gamma)$ , and three translational displacement,  $(x, y, z)$ . The vector  $(\alpha, \beta, \gamma, x, y, z) = (0, 0, 0, 0, 0, 0)$  corresponds to a starting conformation that is equivalent to a standard DNA helix without the bulge. The standard coordinate (S2) of the first base pair 5' to the bulge in the bottom helix is used as the reference coordinate, where the y-axis is along the long axis of the base pair and the positive x-axis points towards the major groove (Fig. S3). A bulge conformation  $(\alpha, \beta, \gamma, x, y, z)$  is generated by first rotating the top helix by  $(\alpha, \beta, \gamma)$ , followed by a translation of  $(x, y, z)$ . We use the zyz Euler convention (S3), in which the order of rotation is as follows: a clockwise rotation of  $\alpha$  along the z axis when viewed from above; a bend of  $\beta$  towards the negative x-axis (i.e., clockwise rotation around the y-axis); and finally a clockwise rotation of  $\gamma$  around the z-axis, with  $\gamma = 0^\circ, 90^\circ, 180^\circ$  and  $270^\circ$  corresponding to the x-, y+, x+ and y- direction, respectively. Conceptually,  $\alpha$  is related to the twisting of the top helix around its own helical axis;  $\beta$  is the bending angle ranges from straight at  $0^\circ$  to a maximum bend of  $180^\circ$ ; and  $\gamma$  represents the direction of the bend. Both  $\alpha$  and  $\gamma$  have a periodicity of  $360^\circ$ . When  $\beta$  approaches  $0^\circ$ ,  $\alpha$  and  $\gamma$  represents similar motions and when  $\beta = 0^\circ$ ,  $\alpha$  and  $\gamma$  are redundant. A slightly different definition of the six-dimensional conformational space of HJH was previously used by Bailor et al. (S4). Bailor et al. (S4) break the two-fold degeneracy of  $(\alpha, \beta, \gamma)$  within  $\pm 180^\circ$  by minimize  $\alpha^2 + \beta^2 + \gamma^2$ , and starts the rotation with the top helix aligned to the bottom helix. Herein work, we break the two-fold Euler degeneracy by limit  $\beta$  to be non-negative, and we start the rotation from a straight conformation with the top helix co-helix stacked with the bottom helix.

We next generate an allowed conformation space as described in detail in ‘Defining the Allowed Space Through Statistical Sampling of the Six-dimensional Conformational Space’ above. Briefly, we systematically sampled the six-dimensional conformational space  $(\alpha, \beta, \gamma, x, y, z)$  to generate a pool of about  $1.8 \times 10^6$  conformations. The angles  $\alpha$  and  $\gamma$  were uniformly sampled over  $360^\circ$  and the bending angle  $\beta$  was sampled from 0 to  $180^\circ$  with frequency of  $\sin \beta$  in order to uniformly sample in probability space (S5). We eliminated conformations would result in steric clashes of the top and bottom helices and conformations with distances between the connection points of the two helices being unreasonably long, leaving a total of  $\sim 5 \times 10^4$  conformations. These conformations represent the geometrically allowed conformational space (S6, 7) of the bulge and are used as the basis set conformations (Fig. 4A).

To predict the scattering profile for the ensemble models, we first calculated the expected Au-Au distance distribution and scattering profile for each conformer within the basis set defined above (also see Materials and Methods). As noted in the Introduction, a major strength of an interferometry approach is that distance distributions, in this case for the pair of site-specifically attached gold nanocrystals, can be translated into a scattering profile and vice versa (S8, 9, 10). Each bulge conformer gives a distribution of states because we know that the flanking DNA helices themselves are not fully rigid. Fortunately, our prior results indicate that the ensemble of DNA helix can be well described by a set of elastic potentials (S8), and we make the simplifying assumption that these distributions are unaffected by the particular conformation of the 3 nucleotide (nt) bulge. Thus, for each bulge conformation,  $i$ , the scattering profile,  $I_i(S)$ , can be

predicted based on our previously generated ensemble model of DNA helix (S8). The expected scattering profile of an ensemble model is the weighted sum of the scattering profile of the  $5 \times 10^4$  conformations within the allowed space:  $I(S) = \sum_{i=1}^{50k} w_i I_i(S)$ . The expected scattering profile of an ensemble,  $I(S)$ , can then be compared with experimentally measured scattering profile,  $I_{\text{exp}}(S)$ .

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 (S11, 12) (see Materials and Methods and “Estimating the Conformational Ensemble of 3A-DNA“ below for details). The procedure sums over different ensemble models weighted by their likelihood of being true given the experimental data, which was calculated from  $\chi^2$  statistics. This weighted summation combines reasonable ensemble solutions to provide an estimate of the ensemble.

**Determining ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $x$ ,  $y$ ,  $z$ ) for Each MD Conformer.** The bottom helix of a MD-simulated conformation was first aligned to a standard DNA duplex created by 3DNA (S13). The transformation matrix (M) for moving the top helix of the standard duplex to align with the top helix of the MD conformation was determined using the Kabsch algorithm (S14), and the transformation matrix M was used to determine the corresponding ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $x$ ,  $y$ ,  $z$ ) by Matlab.

**Estimating the Conformational Ensemble of 3A-DNA.** The 3A-DNA conformational ensemble was determined using a simplified version of a published procedure (S11). Specifically, the posterior probability for an ensemble E of N conformers with a weight vector w given experimental data m,  $f(E|m)$ , can be determined from Bayes theorem using Equation 2 from ref (S11):

$$f(E|m) = \frac{f(m|E)f(E)}{\int dE f(m|E)f(E)} = \frac{f(m|E)}{\int dE f(m|E)} \quad [\text{S1}]$$

where  $f(m|E)$  is the likelihood function, which is the probability of obtaining the experimental data m with an ensemble E with weight vector w.  $f(E)$  is the prior distribution. To avoid introducing bias from incorrect energy functions, we assumed no prior knowledge of the distribution and used a uniformly distributed  $f(E)$  over the allowed space. The  $f(E)$  term is then a constant and consequentially canceled out, and  $f(E|m)$  then simply represents the normalized  $f(m|E)$  (right hand side of Eq. S1).

The estimated weight for the  $i$ th conformation in the conformational space of N conformers,  $w_i$ , is then given by:

$$w_i = \int w_{i,E} f(E|m) dE = \int w_{i,E} \frac{f(m|E)}{\int dE f(m|E)} dE, \quad [\text{S2}]$$

where  $i = 1$  to N. Here N is the size of the geometrically allowed space of about 50,000;  $w_{i,E}$  is the weight of conformer  $i$  in the ensemble solution E;  $f(m|E)$  is as defined above. Since  $\int dE f(m|E)$  is the same for all  $w_i$ , it will be eliminated when  $w_i$  is normalized. Thus, Eq. S2 can be simplified to

$$w_i = \langle w_{i,E} \rangle = \int w_{i,E} f(m|E) dE. \quad [\text{S3}]$$

For each ensemble solution, E,  $f(m|E)$  can be calculated using  $\chi^2$  statistics as in Eq. **S4-S6**, analogous to Equations 7 and 8 in reference (S11).

$$f(m|E) = \prod_{j=1}^6 f(m_j|E) \quad [\text{S4}]$$

$$f(m_j|E) = [\prod_{s=1}^n f(m_{j,s}|E)]^{\frac{1}{n}} \quad [\text{S5}]$$

$$f(m_{j,s}|E) = [2\pi\varepsilon^2]^{-0.5} \exp\left[-\frac{(m_{j,s}-m_{j,s,E})^2}{2\varepsilon^2}\right] \quad [\text{S6}]$$

Here  $j=1$  to 6 corresponds to the six scattering profiles for the six Au-Au pairs;  $n$  is the number of data points per gold pair  $j$ , all measured scattering intensities at  $n \sim 300$  different scattering angles ranges between  $s = 0$  and  $0.05 \text{ \AA}^{-1}$  (Fig. S1);  $m_{j,s}$  is the measured scattering value for gold pair  $j$  at scattering angle  $S$ ; and  $m_{j,s,E}$  is the corresponding value predicted value from an ensemble solution E. The error terms,  $\varepsilon^2$ , is the sum of two error terms  $\varepsilon_1^2$  and  $\varepsilon_2^2$ , where  $\varepsilon_1$  is the standard error of 10 repeated measurements of the same sample and  $\varepsilon_2$  is the error resulting from the uncertainty in modeling the flanking B-DNA helices and the uncertainty in modeling the osition of the Au nanocrystals with respect to its attachment base (S8). More specifically, to account for this uncertainty, we generated 100 sets of different B-DNA helices and Au nanocrystal parameters based on the standard error of these parameters in reference (S8). These 100 synthetic parameter sets were then used to generate 100 sets of predicted scattering intensity curves based on a common ensemble model to get the scale of the error resulted from the uncertainty in these parameters.  $\varepsilon_2$  was then determined as the standard error of the 100 sets of predicted ensemble bulge scattering intensity curves. The common ensemble model was generated by setting  $\varepsilon_2$  to zero.

To calculate Eq. **S3** and to estimate the weight for each conformer in the allowed space, we sampled the ensemble solution space of E as described in Materials and Methods.

**Estimating the Free Energy Change of a Bulge Conformation Upon Varying Salt Conditions.** The electrostatic free energy of a bulge conformation at a specific salt condition,  $\Delta G_{\text{elec}}$ , is estimated using the PDB2PQR online Poisson-Boltzmann calculation server (S15). The input pdbs includes the top and bottom flanking helices, each 13 base pairs long. The bulge sequence was not included. The effect of salt condition on a bulge conformer is then calculated as

$$\Delta\Delta G = \Delta\Delta G_{\text{elec}} = \Delta G_{\text{elec, salt condition2}} - \Delta G_{\text{elec, salt condition1}}. \quad [\text{S7}]$$

## Supporting Text

### 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 determine 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.

Correlations in Euler space have been suggested to be a fundamental property of RNA helix-junction-helices (HJH) and have been reported for RNA bulges, based on NMR-RDC measurements and evaluation of x-ray structural databases (S6, 7, 16), and would be expected for DNA HJHs as well. The preferred conformations of 3A-DNA do exhibit correlations among the Euler angles, albeit weakly (Fig. 4B). For example, as bending ( $\beta$ ) increases, the two twists,  $\alpha$  and  $\gamma$ , both increase (Fig. 4B left). In other words, as the helix bends it is also twisting clockwise. The two twist angles ( $\alpha$  and  $\gamma$ ) are also negatively correlated, reflecting constraints in the overall twist (sum of  $\alpha$  and  $\gamma$ , also see Figs. S7 & S8). Euler correlations for 3A-DNA are largely absent prior to reweighting, but are observed when translational space is restricted (Fig. 4A, Fig. S9 and Table S2; ref. (S7)). Thus, our results suggests that restrictions in translational space, presumably as the result of stabilized local interactions, such as stacking, hydrogen-bonding, sterics within the bulge, or preferred bond rotations of bulge residues, significantly influences the preferred conformations.

The distance-based x-ray interferometry method complements the angle-based NMR-RDC method and can more directly provide information about translation of the helices with respect to one another. Indeed, the interferometry-reweighted ensemble revealed correlations in translation (Fig. 4B right and Table S2). For example, a negative correlation between x and z emerges with the interferometry constraints. This correlation can be accounted for with a crude physical model in which as the top helix moves towards the major groove (increase in x), it moves into the major groove (decrease in z) in order to be in closer contact with the bottom helix.

To better describe the ensemble and go beyond general statements of correlations, we divided the ensemble into five groups of conformations, labeled I to V, based on the experimentally measurable properties of the conformations –i.e., their Au-Au distance profiles (see Fig. S10 and Table S3). To better visualize the ensemble, representative conformers from the five clusters are shown in Fig. 4C. Groups I to IV dominate the ensemble (95%) and can be considered as four regions of an overall free energy basin. Group I (35%) is the most populated. Group V corresponds to only 5% of the modeled conformational space but is well separated from the main free energy basin. A more detailed shape of each region of the free energy landscape cannot be confidently determined from the data, and, of the five groups, we have the least confidence in the minor group (V, 5%) as it could result from over-interpretation of stochastic baseline noise ( Fig. S11). Additional measurements specifically designed to test group V would be needed to confirm and best characterize group V.

Each of the four dominant groups of the 3A-DNA ensemble (groups I to IV) contains thousands of conformers with characteristic distributions in  $(\alpha, \beta, \gamma, x, y, z)$  (Fig. S10). The representative conformations shown in Fig. 4C have  $(\alpha, \beta, \gamma, x, y, z)$  values closest to the mean value of their respective group. As each of the four groups have roughly normally distributed  $(\alpha, \beta, \gamma, x, y, z)$ , the representative conformations are roughly at the center of respective sub-ensembles and provide an easier way to visualize the groups and the difference between them. The bending angle  $\beta$  can be readily understood in Fig. 4C (left). The twist angle  $\gamma$  indicates the bending direction and is best seen in the middle panel of Fig. 4C;  $\gamma$  is the clockwise twist angle from the negative x-axis viewed from the top. The sum of the Euler angles  $\alpha$  and  $\gamma$ ,  $\alpha + \gamma$ , reports the net inter-helical twist, which is harder to visualize and is easier to see via the colored coordinates on top of each upper helix (Fig. 4C, right, also see Fig. S8).

The four dominant groups of the 3A-DNA ensemble (I to IV) 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, Fig. S10B). We can describe the four major groups in terms of hypothetical bending transitions within this quadrant. As the degree of bend,  $\beta$ , increases from II to IV to III, the bending direction,  $\gamma$ , oscillates: II to IV ( $\beta = 37^\circ$  to  $53^\circ$ , with bending direction  $\gamma = 176^\circ$  to  $262^\circ$ ); IV to III ( $\beta = 53^\circ$  to  $71^\circ$ ,  $\gamma = 262^\circ$  to  $241^\circ$ ). Group I ( $\beta = 47^\circ$  and  $\gamma = 217^\circ$ ) is between groups II and IV, closer to the II in  $\beta$  but nearly at the midpoint of these groups in  $\gamma$ , again suggesting the absence of a single smooth transition between the groups (Fig. 4C right). This point is underscored when helical twist is considered in addition to bending: group I is more twisted, with a larger  $\alpha + \gamma$ , than the groups II to IV ( $-60^\circ$  for I and  $-21$  to  $-40$  for II to IV). The net twist of group I,  $-60 \pm 28^\circ$ , which is of the same sign as the right-handed DNA helical twist from 5' to 3' or from bottom to top ( $-34^\circ$  per base step), is equivalent to over-twisting of about two base steps. We speculate that group I, as the most populated state (35%), may represent a specifically stabilized conformational state that is distinct from conformers along the hypothetical motional pathway described above.

In addition to rotation, there are also translational movements, most notably from group IV to III where there is a shift in  $x$  by  $+8 \text{ \AA}$  (Table S3); in other words, as the top helix bends toward the direction perpendicular to the groove (group IV), it also slides away from the major groove.

### Atomistic Models of the Bulge Conformations

To explore the potential molecular interactions that could be responsible for stabilizing different bulge conformers and shaping the ensemble, atomistic modeling was carried out. We used molecular dynamics (MD) in combination with our x-ray interferometry results to generate atomistic models of the bulge conformations. Specifically, we first used MD to generate a large library of bulge conformations and for each conformation we predicted the expected six Au-Au distance distributions (each bulge conformation associates with a distribution rather than a single distance due to fluctuations within the flanking helices) for labeling pair d1 to d6 (Fig. 2A). We then compared Au-Au distances predicted from the MD simulations with those observed experimentally. Finally, we determined the best-fit MD conformer for each of the representative x-ray interferometry conformers (Figs. 4C & 4D), and we subjected these atomistic models to an initial test using 2-aminopurine (2AP) fluorescence.

We carried out approximately 2 microseconds of MD simulations using the AMBER99 force field (S17). The ensemble predicted by these MD simulations differs substantially from the experimental ensemble (Fig. S12). For example, the most stable experimental conformer, I, is not highly populated in MD (Fig. S12). This difference could in principle arise from limitations in either sampling time or the force fields themselves. Due to these limitations, we primarily used our MD simulations to generate atomic-level hypotheses for the key interactions stabilizing the bulge geometries, and did not try to evaluate MD. Nevertheless, our results underscore that empirical and unbiased tests of MD and other methods not based on direct observation are critical. Whereas the importance of such tests may seem obvious, the need for blind tests, well documented in some areas (S18, 19, 20), has not been adopted by all computational sub-fields. One reason for this absence appears to be the difficulty of relating experimental readouts to the structural information from atomic simulations. The ease of translating MD ensembles into probe-probe distance distributions renders x-ray interferometry exceptionally well suited for future tests and improvements of MD methods. Our data and data from analogous future studies can be used to test true state-of-art computation, and we are willing to test predictions from computational models of bulges or other nucleic acids structures in the future.

Aided by MD, we generated atomistic structural models for the four major representative bulge conformations, as described in the Materials and Methods (Fig. 4D). We found that the conformers for groups I to III (Fig. 4C) share a set of common stacking interaction between the first (orange) and second (magenta) bulge adenine and between the first (orange) bulge adenine and the 5'-flanking base pair (Fig. 4D). Conformer IV, which is part of a group that accounts for 21% of the total ensemble, differs from conformations I to III as it does not share the common base stacking in group I to III and is stabilized by alternative stacking interactions between the second (magenta) bulge adenine and the 5' flanking base pair (Fig. 4D).

As a test of general features of these atomistic models, we compared the most stable conformation I with the most favored MD conformation. The two conformations differ substantially. The most favored MD conformer has the 3'-A stacking between two flanking guanines (green spheres in Fig. 5A), whereas this residue is largely unstacked in I. We distinguished between these two models with 2-aminopurine (2AP) fluorescence assays, as 2AP is highly sensitive to stacking (Fig. 5B and Table S4; (S21, 22, 23)). In particular, 2AP is most strongly quenched by stacking with guanine (S21, 22, 23). If the stacking-heavy MD model (Fig. 5A right) were dominant, we would expect the 3'-A to be more strongly quenched than the 5'-A, and the opposite would be expected if the stacking arrangement of conformer I were prevalent (Fig. 5A left).

The greater fluorescence quenching of 5'-2AP than the 3'-2AP (Fig. 5B) is consistent with the simplest expectations from the predominant experimental model (I) and not from the MD model. Furthermore, 3'-2AP in A3-DNA is also less quenched than 3'-2AP in single-stranded DNA (Fig. 5B), consistent with the experimental model I in which hydrogen-bonding interactions (Fig. 4D and Fig. S12) steer the 3'-A away from its neighboring guanosine residues which likely give its strong quenching in the single-stranded form (S21, 22, 23).

We also carried out fluorescence lifetime measurements (Table S4), as the fluorescence lifetime of 2AP in an oligonucleotide often has multiple exponential components that can report on different types of environments of the 2AP probe (S21, 22, 24). The long lifetime component of 2AP is often associated with unstacked conformations and short lifetime components are often associated with stacked conformation, especially with guanine, or unstacked conformations that are under exchange in the low nanosecond or faster timescale with stacked conformations (S21, 22, 25, 26). 3'-2AP in conformer I (Fig. 5A left) is mostly free of base stacking and would therefore be expected to have a longer lifetime than the 3'-2AP if it were in the stacked MD model conformation (Fig. 5A right). Our 2AP fluorescence lifetime measurements show that the stronger fluorescence for the 3' bulge adenine over the 5' bulge adenine arises with a longer average lifetime (2.0 vs 0.7 ns, see Table S4), mostly due to a long lifetime component (7.9 ns, ~7% of the total population, see Table S4) of the 3'-2AP. In comparison, such long lifetime component is absent for single-stranded DNA with or without neighboring guanine (Table S4, and ref. (S27)), where interactions needed to prevent rapid base conformational exchange are absent.

This exercise, using x-ray interferometry, MD, and 2AP fluorescence, suggests that combining lower resolution global information from experiments and the atomic-level high-resolution capability of MD can provide atomic-level models that can be further tested. Clearly additional and more incisive atomic-level tests will be needed to further test and to refine the atomistic features of 3A-DNA. While providing reasonable discrimination between the conformer I and the MD structure, there are many conformers that differ subtly or substantially from conformer I that would also yield low quenching of a 3'-2AP. We anticipate that as MD results become more and more reliable, through direct comparisons with predictions from x-ray interferometry and other experimental results, a point will be reached when less extensive experimental testing will be required and higher-resolution models and ensembles will be obtained.

**Table S1. Sequences used in this study**

| Construct series | Sequence   |
|------------------|--|
| 0-5A DNA         | 5'- CGAACCGTGAAGGCA <b>AAA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'         |
|                  | 5'- CGAACCGTGAAGGCA <b>AA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'          |
|                  | 5'- CGAACCGTGAAGGCA <b>A</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'           |
|                  | 5'- CGAACCGTGAAGGCGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'                        |
| 3A DNA           | 5'- CGAACCGTGAAGGCA <b>AA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'          |
|                  | 5'- CGAACCGTGAAGGCA <b>AA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'          |
|                  | 5'- CGAACCGTGAAGGCA <b>AA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'          |
|                  | 5'- CGAACCGTGAAGGCA <b>AA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'          |
|                  | 5'- CGAACCGA <b>GA</b> AAGGCA <b>AA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>TCT</u> TCCGCTAGAGACGCCG-5' |
|                  | 5'- CGAACCGA <b>GA</b> AAGGCA <b>AA</b> AAGATCTCTGCGGC -3'<br>3'- GCTTGGC <u>TCT</u> TCCGCTAGAGACGCCG-5' |
| 3T DNA           | 5'- CGAACCGTGAAGGCT <b>TTT</b> GATCTCTGCGGC -3'<br>3'- GCTTGGC <u>CA</u> CTTCCGCTAGAGACGCCG-5'           |

The bulge sequence is in bold. The gold labeling T are in italic. The non-bulge sequences for these constructs are the same expect for the underlined residues. The 0A-DNA (row four) is from reference (S8).

**Table S2. Correlation coefficients among the six degrees of conformational freedom for the allowed conformational space (black) and the x-ray interferometry estimated conformational space (brown)**

| R        | $\alpha$ | $\beta$                    | $\gamma$                     | x                           | y                            | z                            | $\alpha + \gamma$           |
|----------|----------|----------------------------|------------------------------|-----------------------------|------------------------------|------------------------------|-----------------------------|
| $\alpha$ |          | <b>0.00</b><br><b>0.26</b> | <b>-0.09</b><br><b>-0.36</b> | <b>0.23</b><br><b>0.26</b>  | <b>0.00</b><br><b>0.25</b>   | <b>-0.01</b><br><b>-0.08</b> |                             |
| $\beta$  |          |                            | <b>-0.04</b><br><b>0.09</b>  | <b>0.04</b><br><b>0.27</b>  | <b>-0.20</b><br><b>-0.06</b> | <b>0.34</b><br><b>0.24</b>   | <b>-0.02</b><br><b>0.42</b> |
| $\gamma$ |          |                            |                              | <b>-0.01</b><br><b>0.06</b> | <b>0.00</b><br><b>0.06</b>   | <b>0.01</b><br><b>0.08</b>   |                             |
| x        |          |                            |                              |                             | <b>-0.02</b><br><b>0.01</b>  | <b>-0.02</b><br><b>-0.28</b> | <b>0.52</b><br><b>0.60</b>  |
| y        |          |                            |                              |                             |                              | <b>0.06</b><br><b>0.12</b>   | <b>0.06</b><br><b>0.11</b>  |
| z        |          |                            |                              |                             |                              |                              | <b>0.01</b><br><b>-0.16</b> |

See also Fig. 4 and Figs. S7, S9 and S18 for plots of correlations.

**Table S3. The five clusters of the estimated ensemble of the 3A bulge**

| Conformation cluster | Population   | $\alpha$ ( $^{\circ}$ ) | $\beta$ ( $^{\circ}$ ) | $\gamma$ ( $^{\circ}$ ) | x ( $\text{\AA}$ ) | y ( $\text{\AA}$ ) | z ( $\text{\AA}$ ) | $\alpha + \gamma$ ( $^{\circ}$ ) |
|----------------------|--------------|-------------------------|------------------------|-------------------------|--------------------|--------------------|--------------------|----------------------------------|
| I                    | 35 $\pm$ 13% | 84 $\pm$ 33             | 47 $\pm$ 13            | 217 $\pm$ 28            | -3 $\pm$ 3         | -4 $\pm$ 4         | 5 $\pm$ 3          | -60 $\pm$ 28                     |
| II                   | 19 $\pm$ 11% | 161 $\pm$ 42            | 37 $\pm$ 27            | 176 $\pm$ 34            | -3 $\pm$ 2         | -3 $\pm$ 4         | 4 $\pm$ 3          | -40 $\pm$ 38                     |
| III                  | 19 $\pm$ 9%  | 102 $\pm$ 40            | 71 $\pm$ 12            | 241 $\pm$ 20            | 0 $\pm$ 6          | -3 $\pm$ 5         | 7 $\pm$ 5          | -21 $\pm$ 43                     |
| IV                   | 21 $\pm$ 10% | 69 $\pm$ 49             | 53 $\pm$ 15            | 262 $\pm$ 15            | -8 $\pm$ 9         | -1 $\pm$ 6         | 5 $\pm$ 4          | -30 $\pm$ 43                     |
| V                    | 5 $\pm$ 2%   | 28 $\pm$ 52             | 52 $\pm$ 21            | 77 $\pm$ 56             | 4 $\pm$ 5          | -9 $\pm$ 6         | 4 $\pm$ 3          | 103 $\pm$ 26                     |

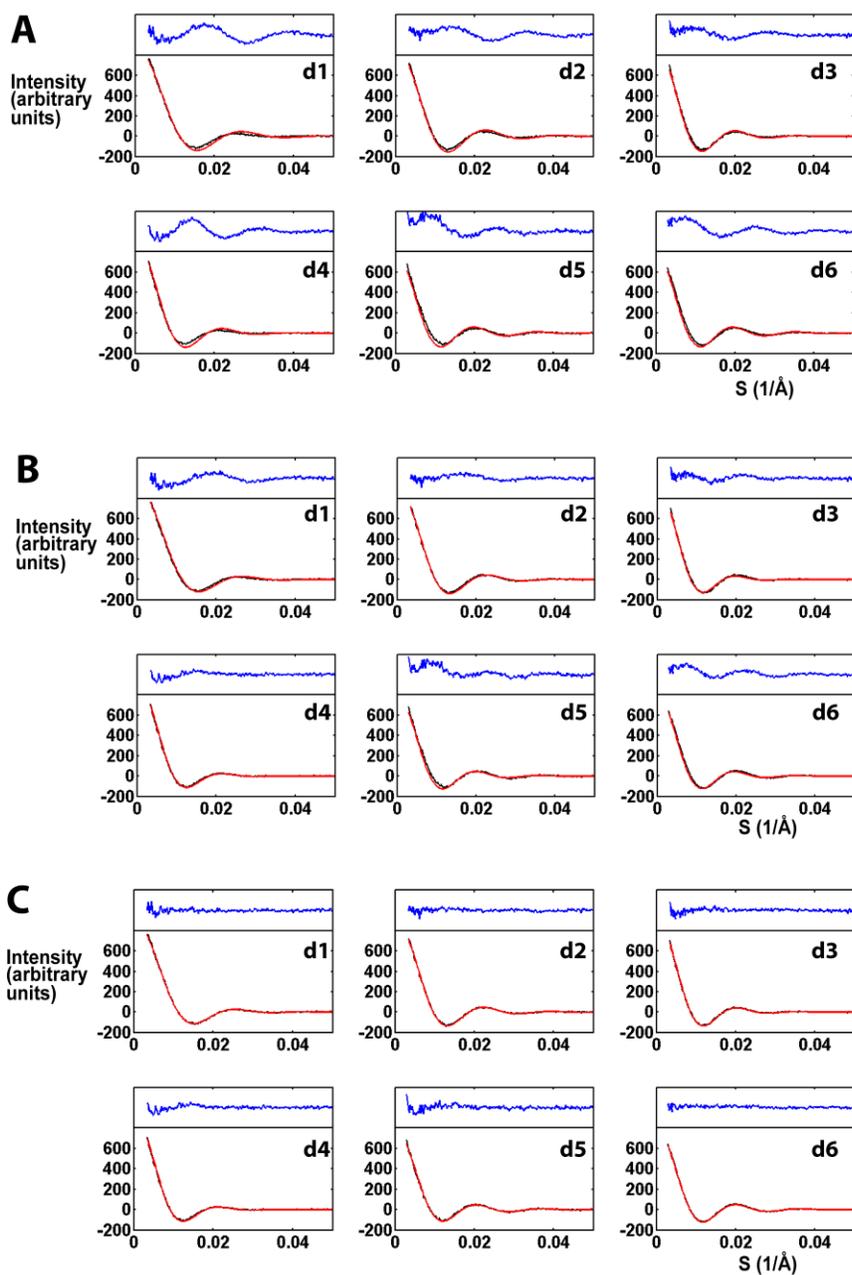
The values for the population are mean values and errors reported as standard deviations estimated from ensemble models generated using experimental data with addition of Gaussian-distributed error (see SI Methods for the process of estimating experimental error). The values for the other columns (the rotational and translational parameters) are means and not errors but rather the standard deviation representing the spread of the population in each cluster (Fig. S10B).

**Table S4. Fluorescence studies of 2-aminopurine (2-AP) containing DNA constructs**

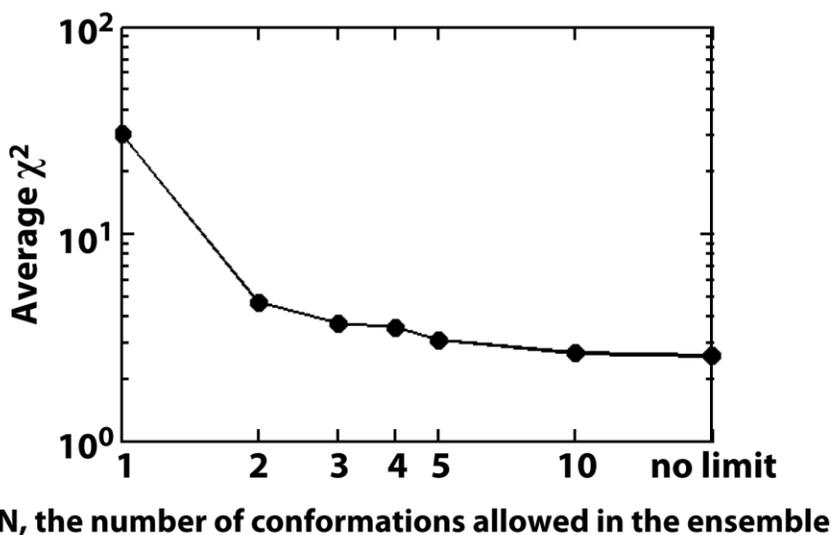
| 2-AP(*) position          | Lifetime (ns)       |                     |                   |                     | Relative steady state fluorescence, $F_{ss}$ | Unresolvable short lifetime component <sup>#</sup> |
|---------------------------|---------------------|---------------------|-------------------|---------------------|--|--|
|                           | $t_1$               | $t_2$               | $t_3$             | $\langle t \rangle$ |  |  |
| 5'-A*AA (single-stranded) | 0.98 (75%)<br>(75%) | 2.19 (25%)<br>(25%) |                   | 1.28                | 3.6±0.1                                      | 0%   |
| 5'-AAA* (single-stranded) | 0.35 (64%)<br>(22%) | 1.07 (20%)<br>(7%)  | 3.8 (15%)<br>(5%) | 1.03                | (1)  | 65%  |
| 5'-A*AA (bulge duplex)    | 0.20 (60%)<br>(32%) | 0.81 (33%)<br>(17%) | 4.7 (7%)<br>(4%)  | 0.74                | 1.1±0.1                                      | 47%  |
| 5'-AAA* (bulge duplex)    | 0.37 (62%)<br>(25%) | 2.01 (20%)<br>(8%)  | 7.9 (18%)<br>(7%) | 2.03                | 2.3±0.4                                      | 60%  |
| 5'-AA*A (single-stranded) |                     |                     |                   |                     | 5.7  |  |
| 5'-AA*A (bulge duplex)    |                     |                     |                   |                     | 4.3  |  |

See Fig. 5 for the flanking base sequences for the four constructs, two single-stranded and two double-stranded. The position of the A that is replaced by 2-AP is labeled A\*. Multiple exponential fitting of time resolved fluorescence decays were used to determine lifetime  $t_i$  and the weight factors  $a_i$  (black percentage in paraphrases). The average lifetime is calculated as  $\langle t \rangle = \sum t_i a_i$ .

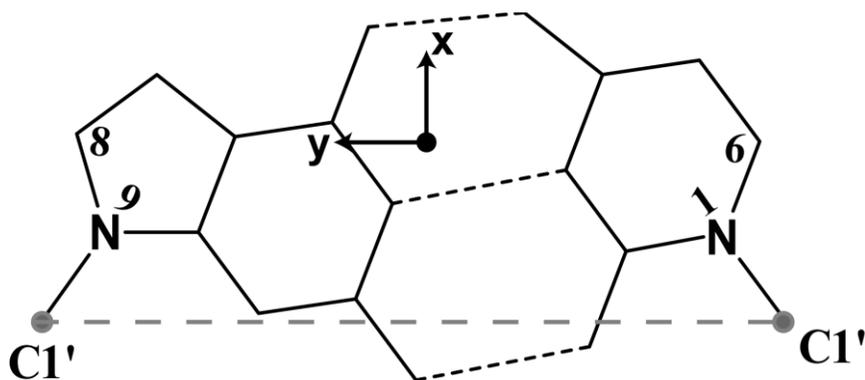
<sup>#</sup>2-AP can have a very short lifetime component, such as at 10 ps, due to efficient hole transfer quenching through stacking with guanine (S23). This short lifetime component includes both the G-stacked conformation and conformations that are under fast exchange (low ps) with the G-stacked conformation. This short ps lifetime component is not detectable in our lifetime measurements (limited to ~0.15 ns), but its quenching is reflected in the steady state fluorescence intensity. Thus, the fraction of 2-AP with the unresolvable short lifetime can be deduced from average lifetime  $\langle t \rangle$  and steady state fluorescence  $F_{ss}$  as described below. The steady state fluorescence is expressed as  $F_{ss} = F \cdot f_{short} t_{short} + F \cdot (1 - f_{short}) \langle t \rangle$ , where  $f_{short}$  is the fraction of the molecules that has a short lifetime of low ps and F is the fluorescence intensity of free 2-AP. As  $t_{short}$  is small and its exact value is unknown, the first term is set to zero. Thus,  $F_{ss} = F \cdot (1 - f_{short}) \langle t \rangle$ . The value F can be calculated using the value of the single-stranded 5'-A\*AA as 2-AP in this construct does not directly neighbor an intra- or inter-strand guanine and its  $f_{short}$  can be assumed to be 0. With F and  $\langle t \rangle$  known,  $f_{short}$  (last column) was estimated for all constructs. The fraction (f) for the longer lifetime components can then be calculated (percentage in parentheses in blue) with the sum of the blue percentages for each construct (across each row) equaling 1.



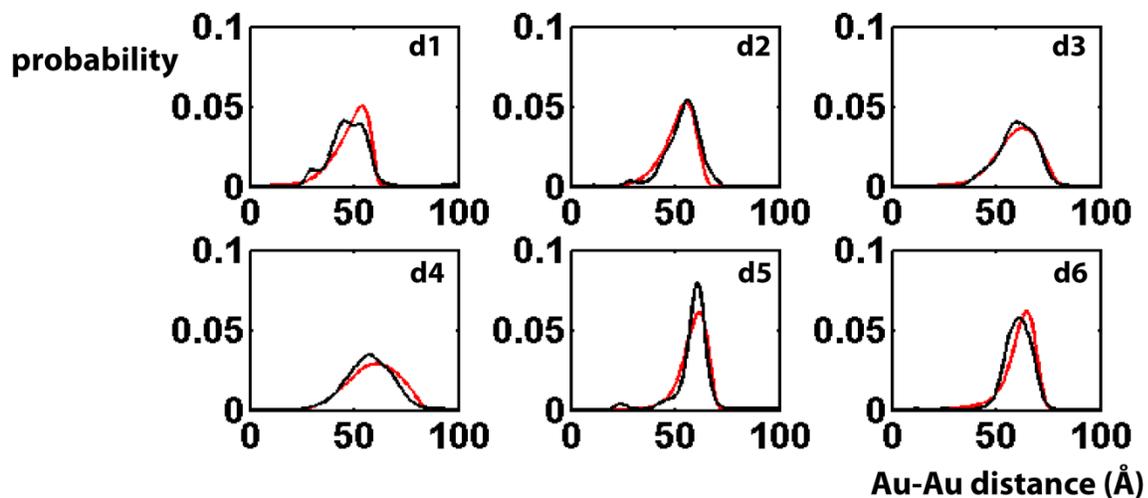
**Figure S1. Fitting of the gold-gold scattering profiles.** The six (d1 to d6, see Fig. 2A) measured gold-gold scattering profiles (black line) for 3A-DNA are plotted together with the best-fit model predictions (red lines) using a one-state approximation (**A** and **B**) or a non-constrained multiple-state ensemble model (**C**) and the corresponding residuals, multiplied by a factor of three to aid visualization (blue lines). In the one-state approximation models the ensemble is assumed to only contain a single conformation (**A**,  $\chi^2 = 12$ ) or elastic fluctuations around a single conformation (**B**,  $\chi^2 = 4.9$ ). In the non-constrained multiple-state model (**C**,  $\chi^2 = 2.6$ ), there is no prior assumption of the ensemble (see “A Procedure for Building the Ensemble” for detail).



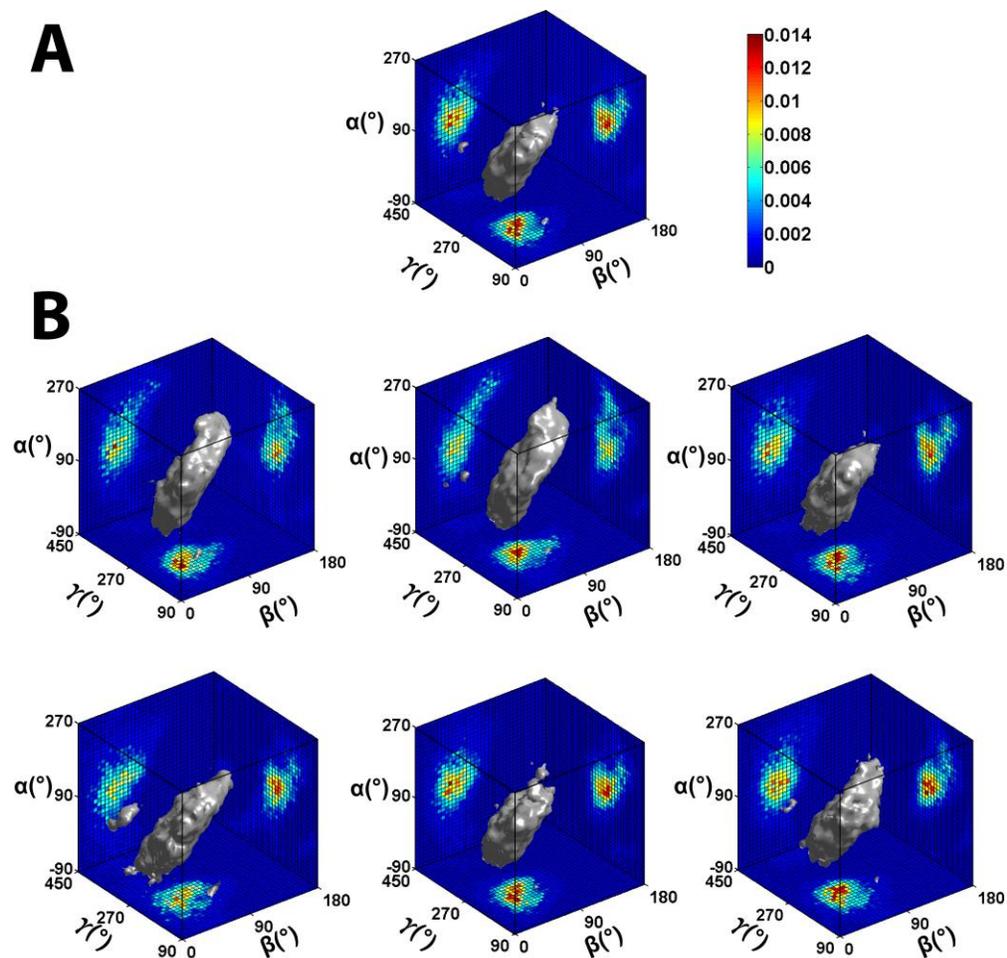
**Figure S2. The best fit achievable using ensemble models with a set number of discrete conformations.** The best-fit  $\chi^2$  is plotted against the number of discrete conformations used in the fitting (black dots).



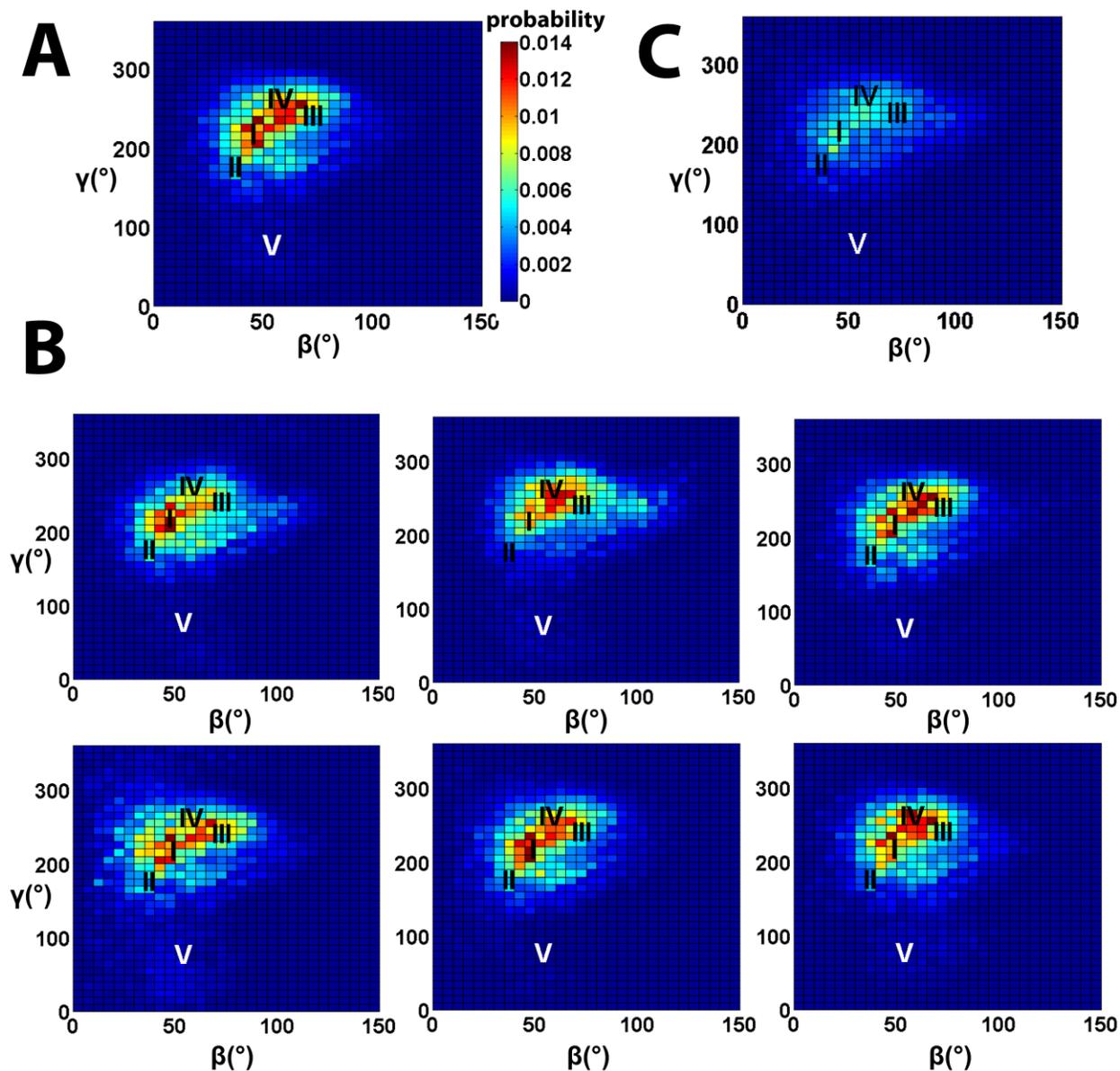
**Figure S3. The standard coordinate of a base pair, as defined in reference (S2).** The x-axis is along the perpendicular bisector of  $C1' \cdots C1'$ ; the y-axis is parallel to  $C1' \cdots C1'$ ; the origin is at the intersection of x-axis and the line connecting purine C8 and pyrimidine C6 (S2).



**Figure S4. Cross-validation tests of the ensemble models.** Each of the six (d1-d6, see Fig. 2A) experimental Au-Au distance distributions (black lines) are plotted together with the distance distribution (red lines) predicted from the ensemble model generated with the other 5 distance distributions.



**Figure S5. The estimated 3A-DNA ensemble.** The conformational ensemble of the 3A-DNA was estimated using all six (**A**) or five of the 6 Au-Au distance distributions (**B**, the excluded distribution in panel top left, top middle, top right, bottom left, bottom middle and bottom right, are distribution d1 to d6 (Fig. 2A), respectively). The grey surface encloses 80% of the total population.



**Figure S6. The estimated 3A-DNA ensemble.** The conformational ensemble of the 3A-DNA was estimated using all six (A) or five of the six Au-Au distance distributions (B, the excluded distribution in panel top left, top middle, top right, bottom left, bottom middle and bottom right, are distribution d1 to d6 (Fig. 2A), respectively). The standard deviations of the six ensembles in panel B are plotted in panel C. The five representative conformations (black and white letters and numbers, see also Fig. 4C) in panel A are labeled at the same coordinates in panel B and C for ease of comparison. See Fig. S5 for the ensembles plotted in the three-dimensional Euler space.

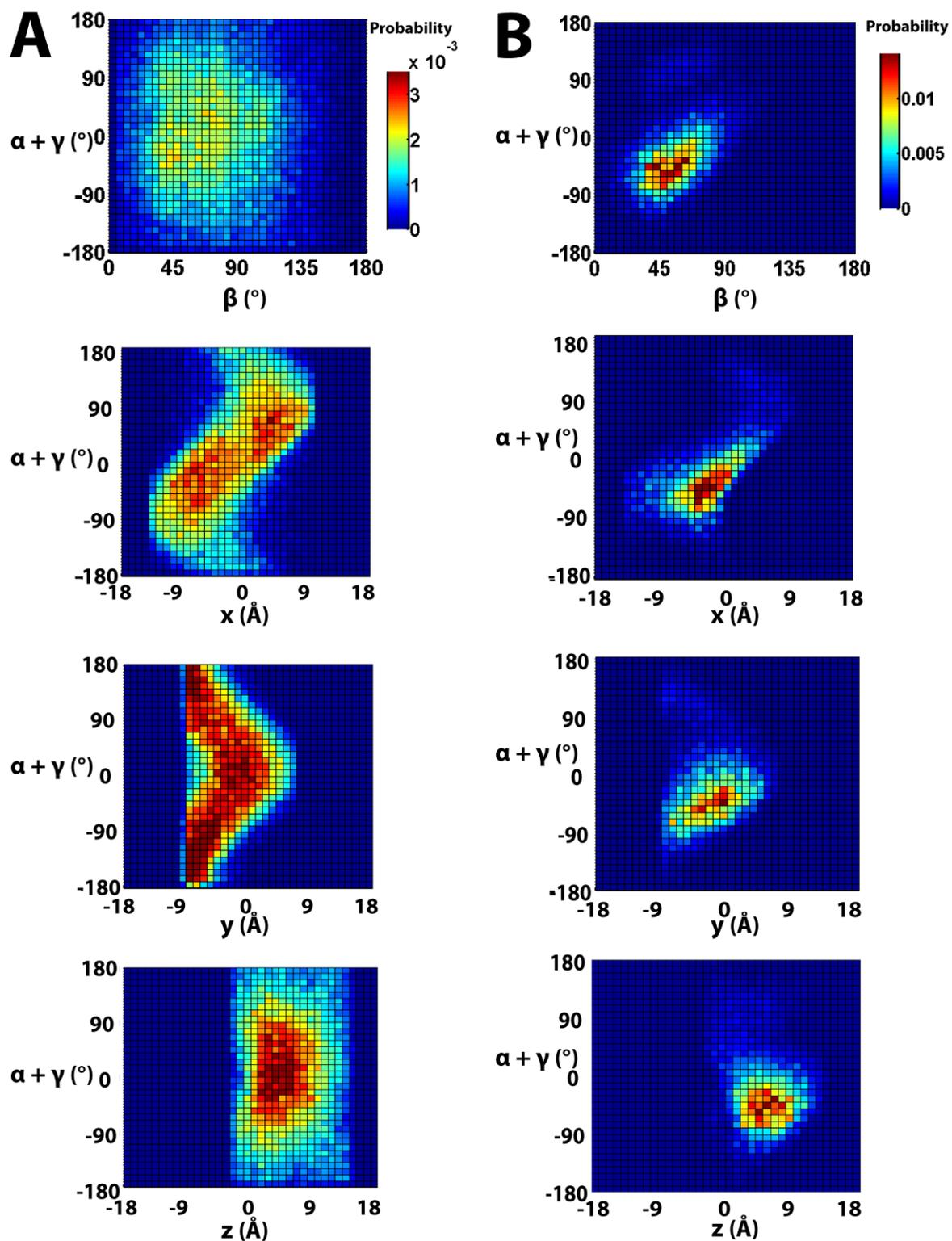
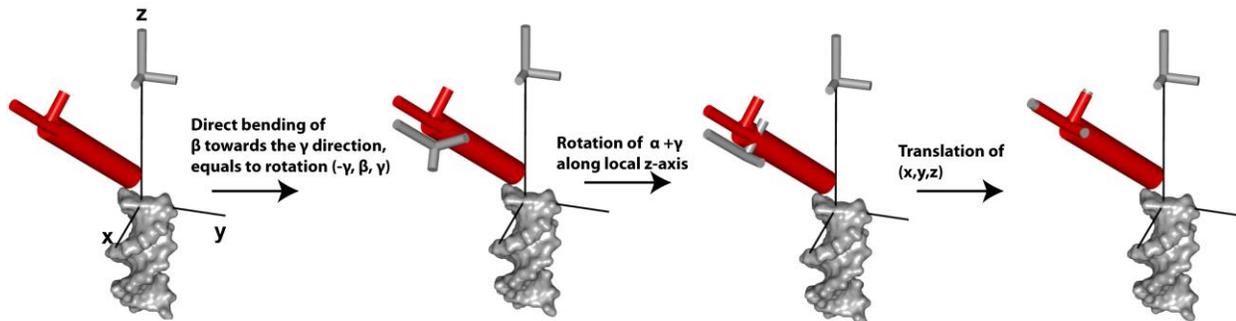
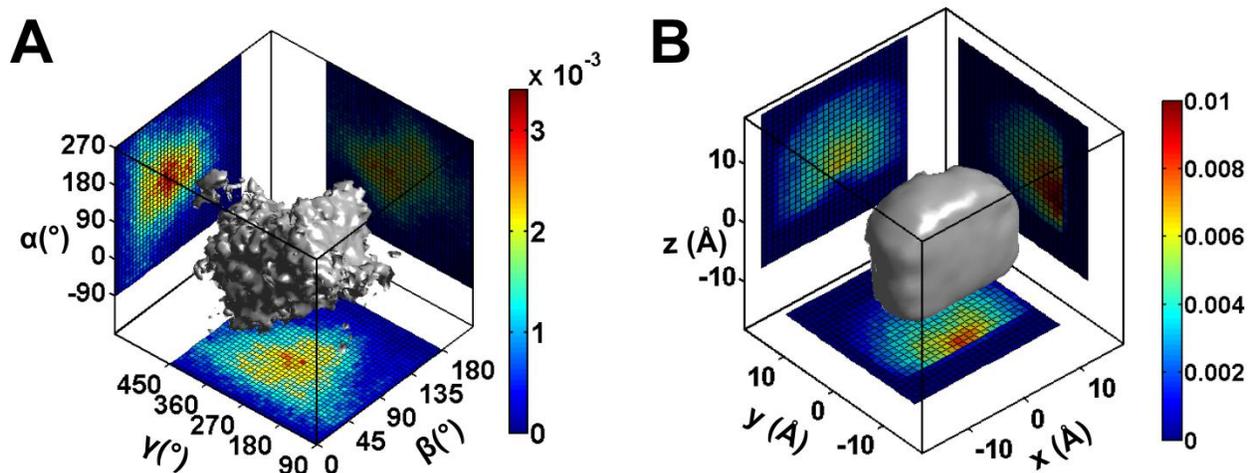


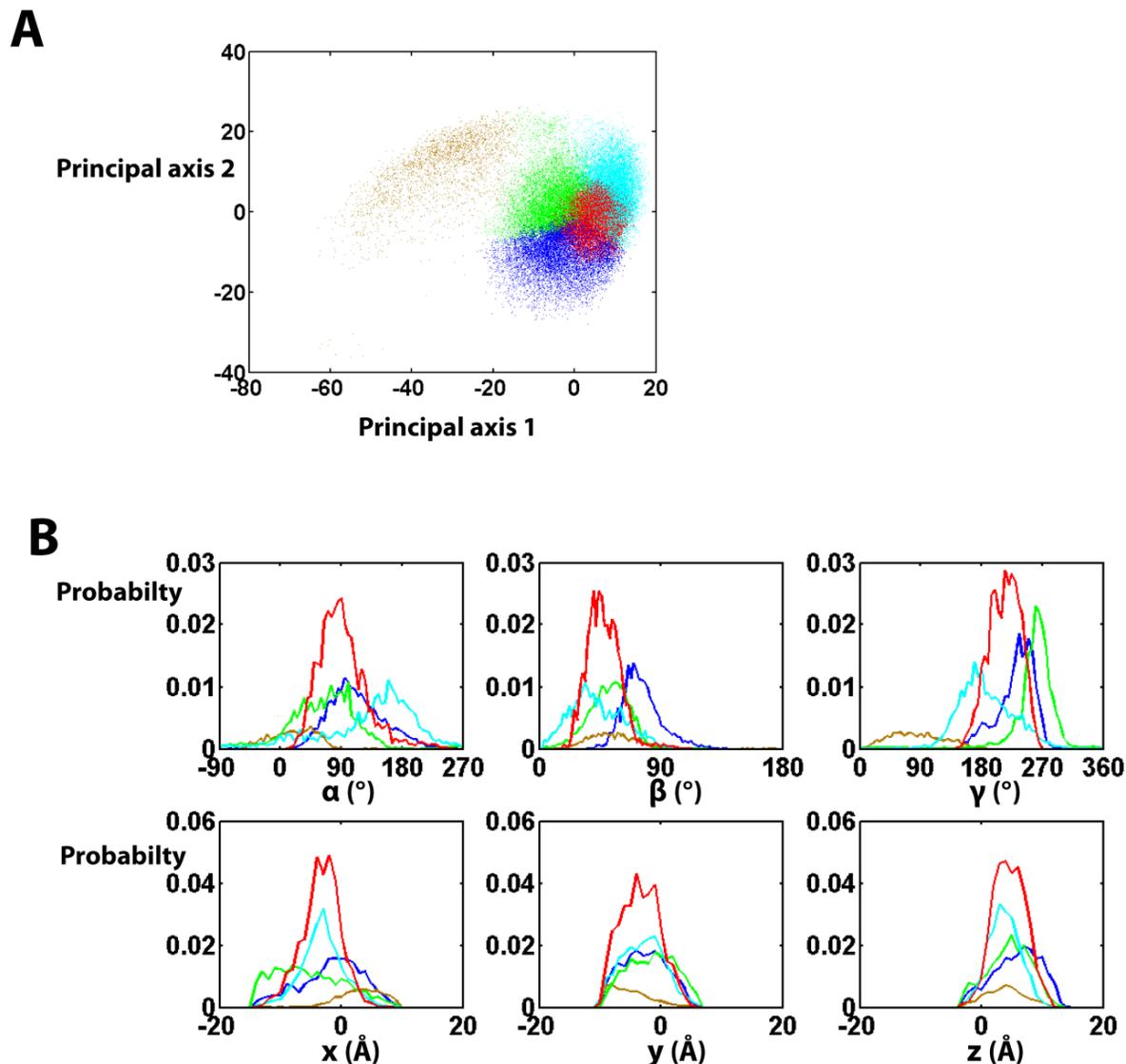
Figure S7. The geometrically allowed conformational space (A) and the estimated 3A-DNA ensemble (B).



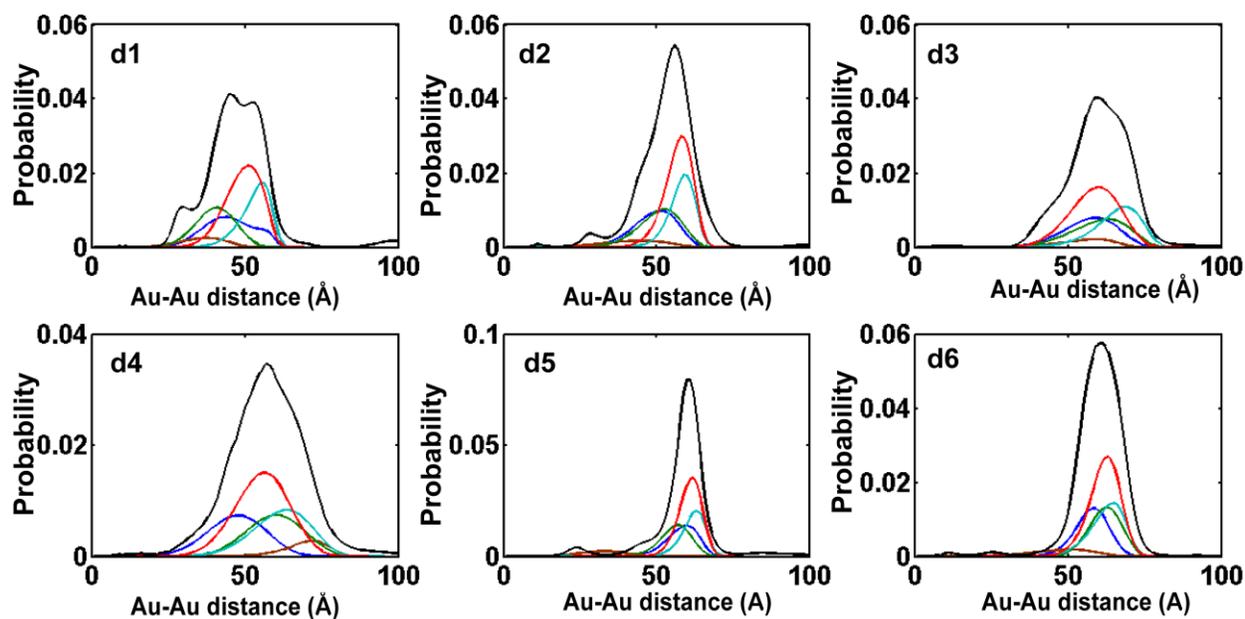
**Figure S8. Understand the meaning of  $\alpha + \gamma$  through a rotational sequence.** The red conformation (conformation I in Figure 4C) can be obtained from the starting conformation (grey coordinate, left panel) by first rotating the starting conformation (grey coordinate, left panel) by  $\beta$  towards the direction indicated by  $\gamma$  ( $217^\circ$  clockwise from the negative x-axis) along an axis that is normal to the z-axis and the  $\gamma$  direction on the xy plane (Panel 1 to 2, from left to right). In the second rotational step (Panel 2 to 3, from left to right), the rotated grey coordinate twists along its z-axis by  $\alpha + \gamma$  (i.e.,  $\alpha + \gamma$  reports the twisting of top helix along its own long axis). The rotated grey coordinate can then be aligned to the red coordinate through a simple translation step (Panel 3 to 4, from left to right). This rotational sequence generates the same conformation as the regular Euler rotation series of  $(\alpha, \beta, \gamma)$  used in the text (see Supporting Methods for description).



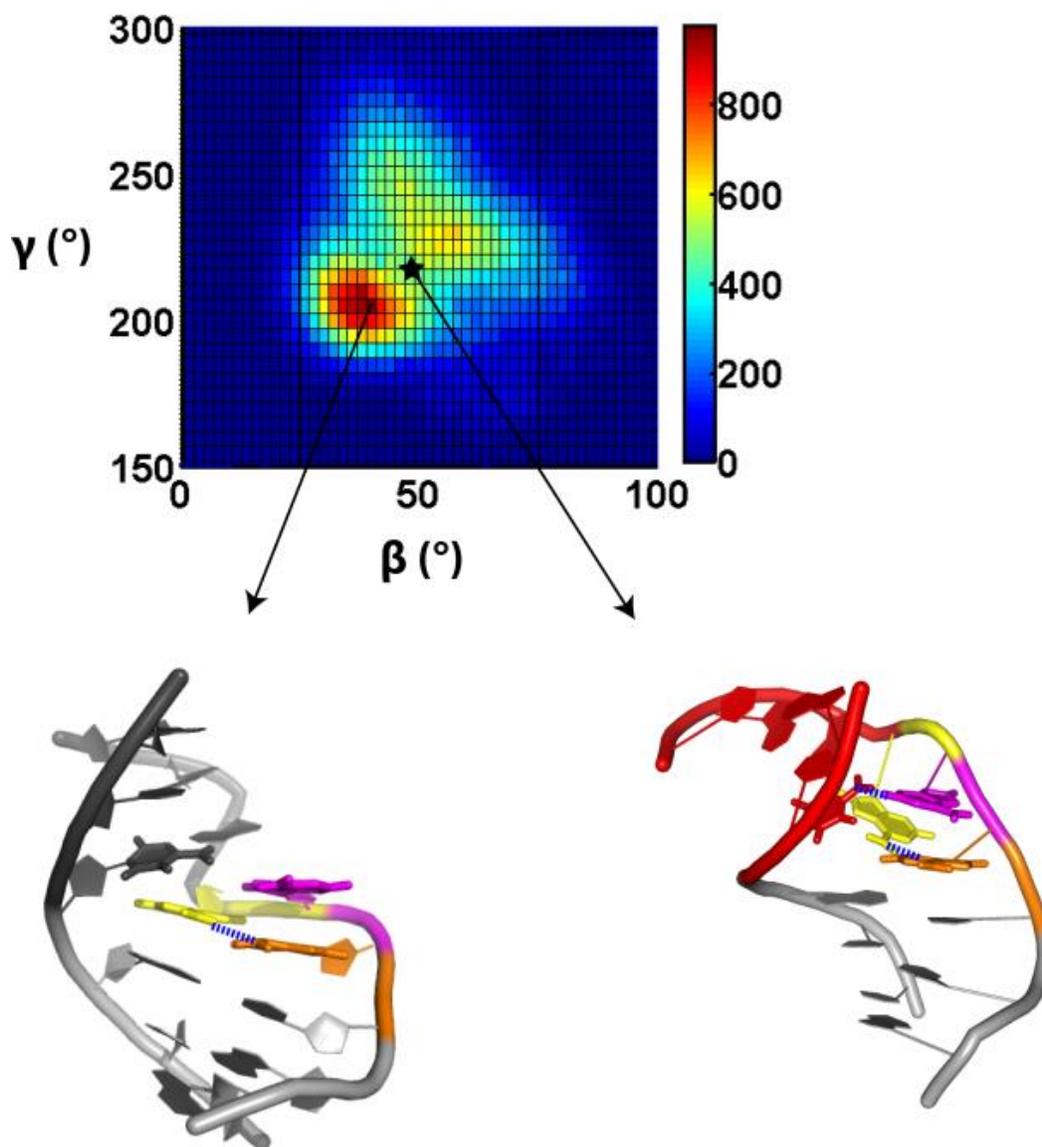
**Figure S9. The geometrically allowed conformational space of 3A-DNA bulge.** The grey surface encloses 80% of the total population in (A) the Euler angle space and (B) the xyz translational space.



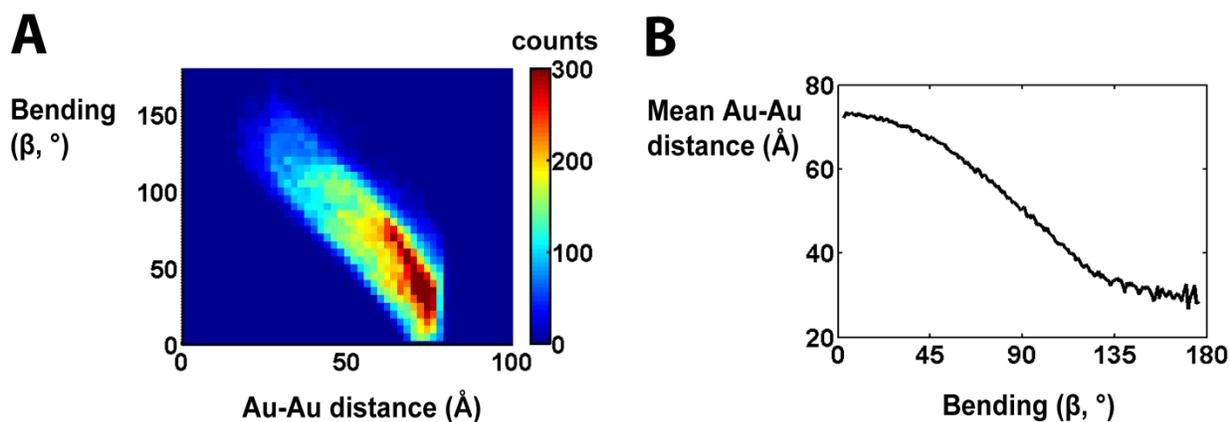
**Figure S10. The estimated 3A-DNA ensemble contains five clusters of conformations. (A)** The five clusters plotted in the first two principal axes of the six dimensional gold-gold distance space,  $d_1$  to  $d_6$  (Fig. 2A). The principal axes were generated using the principal component analysis (PCA) function, `princomp`, in Matlab. PCA transforms the six-dimensional distance data to a new coordinate system of principal axes, such that the projections of the data onto the first and second principal axis give rise to the greatest and second greatest variance, respectively. **(B)** The five clusters plotted against  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $x$ ,  $y$ , and  $z$ . The color code used is the same as in Fig. 4C: I (red), II (cyan), III (blue), IV (green) and V (brown).



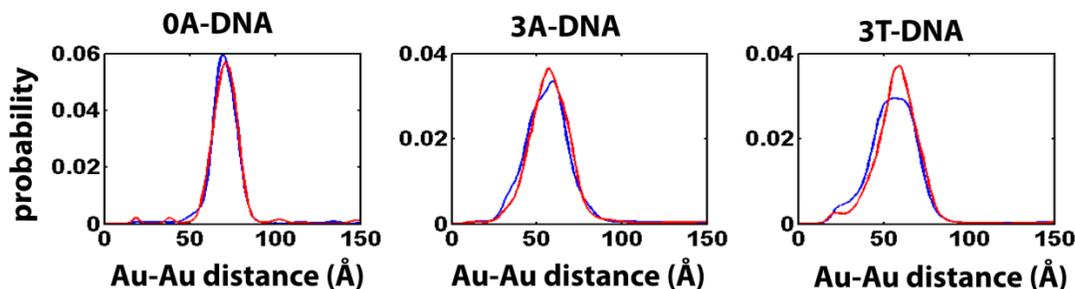
**Figure S11. Au-Au distance distributions from individual conformational groups.** The measured Au-Au distance distribution (black) and the individual contributions from the five groups (I red, II cyan, III, blue, IV green and V brown) for probe pairs d1 to d6 (also see Fig. 2A).



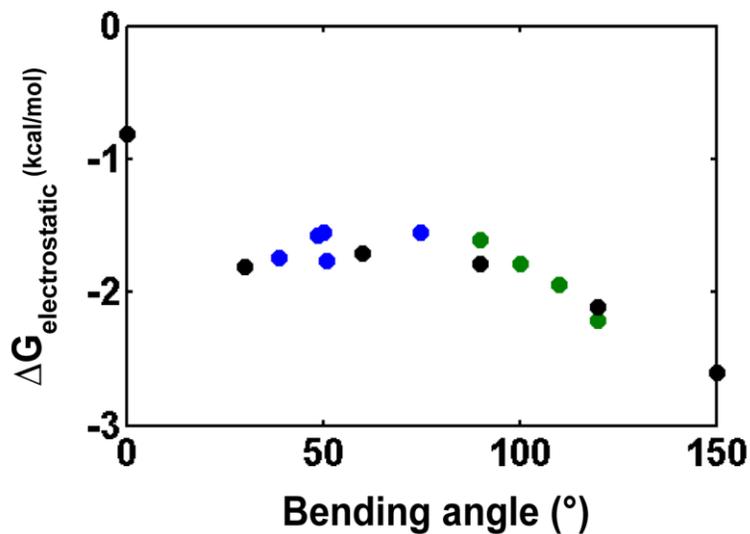
**Figure S12. The conformational ensemble predicted by MD simulation (see Materials and Methods).** The most stable MD conformer (bottom left) is stabilized by stacking and one hydrogen bond (blue dashed line). The most stable experimental conformer I (bottom right and black star, from Fig. 4D –i.e., the MD structure most closely matching the most populated species of the experimentally derived ensemble) is stabilized by stacking and two additional hydrogen bonds between N3 of the first bulge adenine and N6 of the third bulge adenine (lower blue dotted line in bottom right, N3 to N7 distance is 3.0 Å) and between N3 of the second bulge adenine and N4 of a cross-strand cytosine (higher blue dotted line in bottom right, N3 to N4 distance is 2.9 Å).



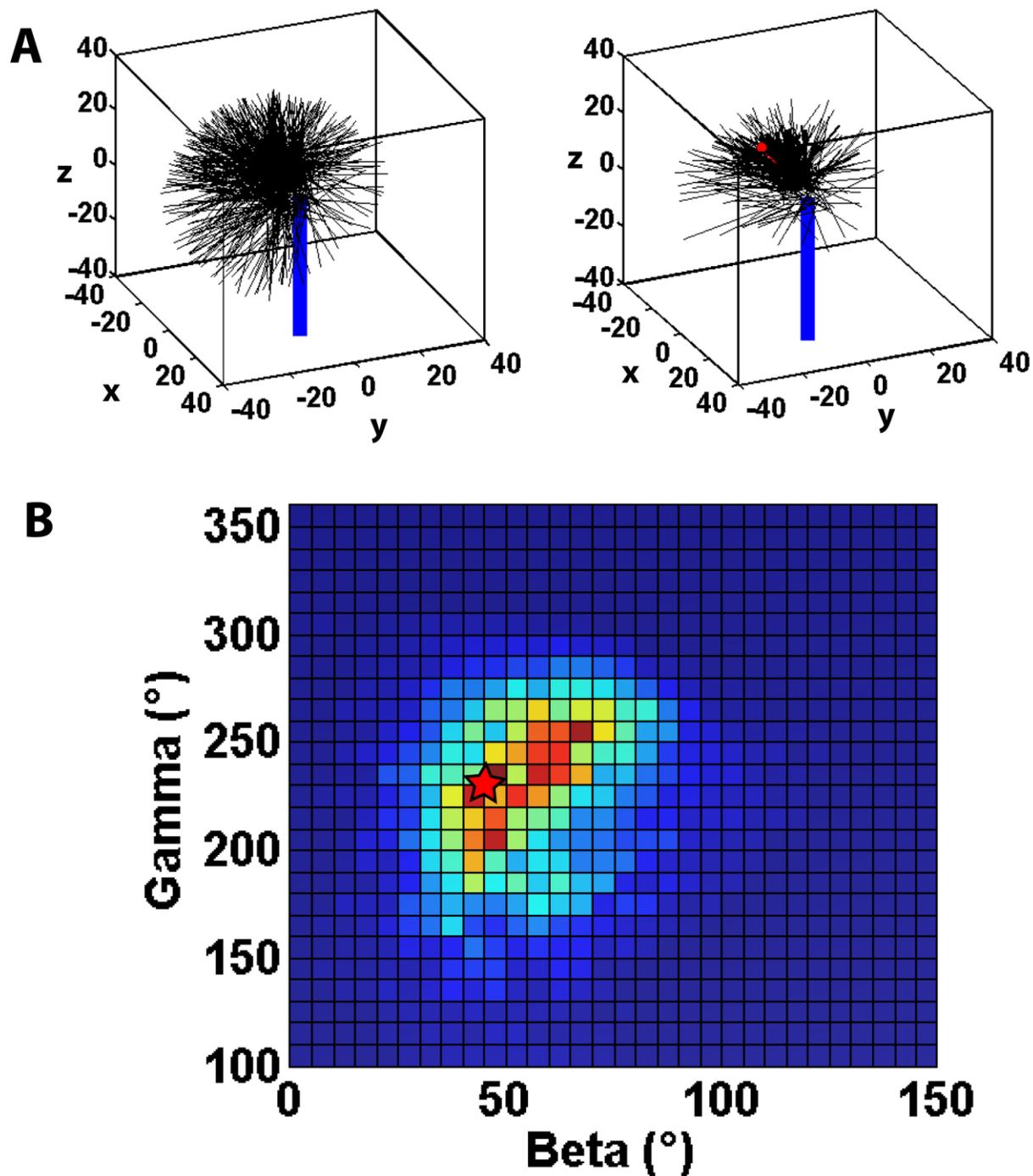
**Figure S13. The relationship between bending angle  $\beta$  and the gold-gold distance for the allowed space conformations.** The gold-gold distances are the mean distances of the gold-gold distance distributions (probe pair d4, Fig. 2A) for the allowed space conformations. The relationships shown here are (A) the distribution of allowed space bending angles for conformations with the same gold-gold distances, and (B) the average Au-Au distance for conformations with different bending angles.



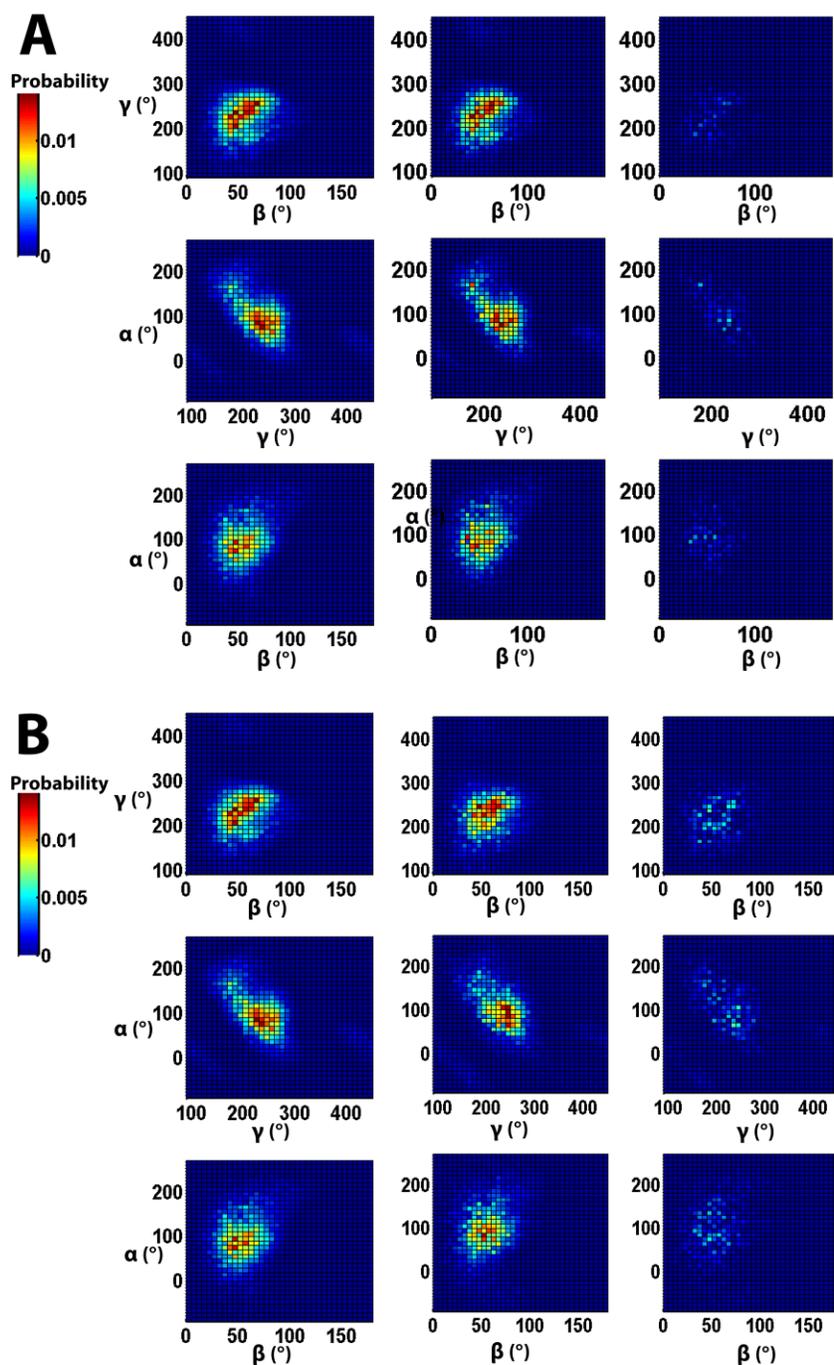
**Figure S14. The effect of solution salt conditions on DNA helix alone is minor compared to the effect on DNA bulges.** Comparison of the gold-gold distance distributions for DNA helix without (left) and with a 3A (middle) or a 3T (right) bulge, with 0 (red) or 4 (blue) mM  $Mg^{2+}$ . The gold pairs are of the labeling scheme d4 (Fig. 2A).



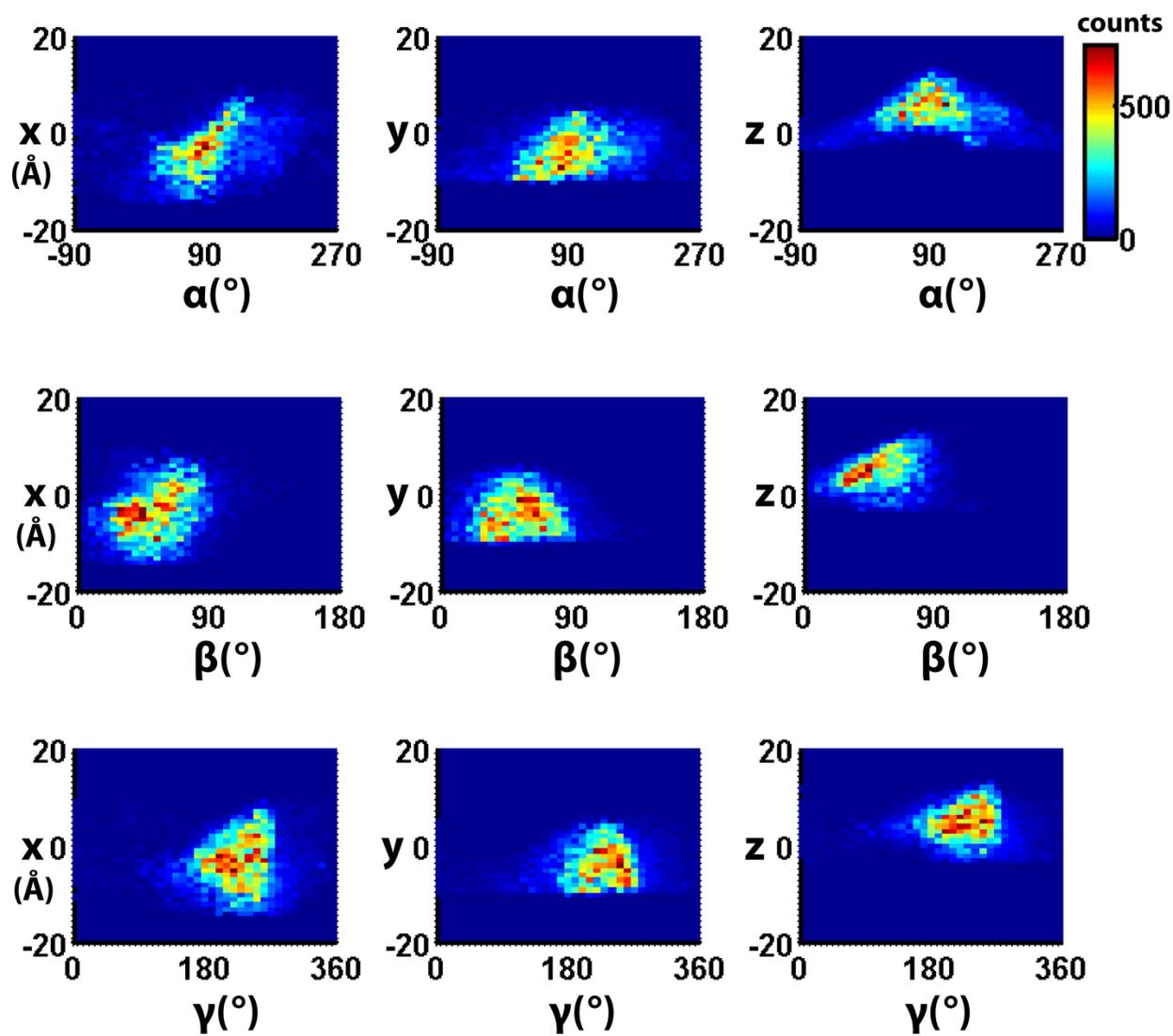
**Figure S15. The changes in electrostatic free energy as a function of bending angle.** The change in electrostatic free energies from increasing the  $\text{Mg}^{2+}$  concentration from 0 to 4 mM in the background of 160 mM  $\text{Na}^+$  were calculated using an online Poisson-Boltzmann calculation server (S15) for different bulge conformations (blue circles: conformation I to V in Table S3; green circles: randomly chosen conformations from the allowed space; black circles: hypothetical conformations that have the same  $[\alpha, \gamma, x, y, z]$  as conformation I in Table S3 but with different bend angles  $\beta$ ). The Poisson-Boltzmann calculation only considered the charges on the flanking helices, not on the 3nt bulge.



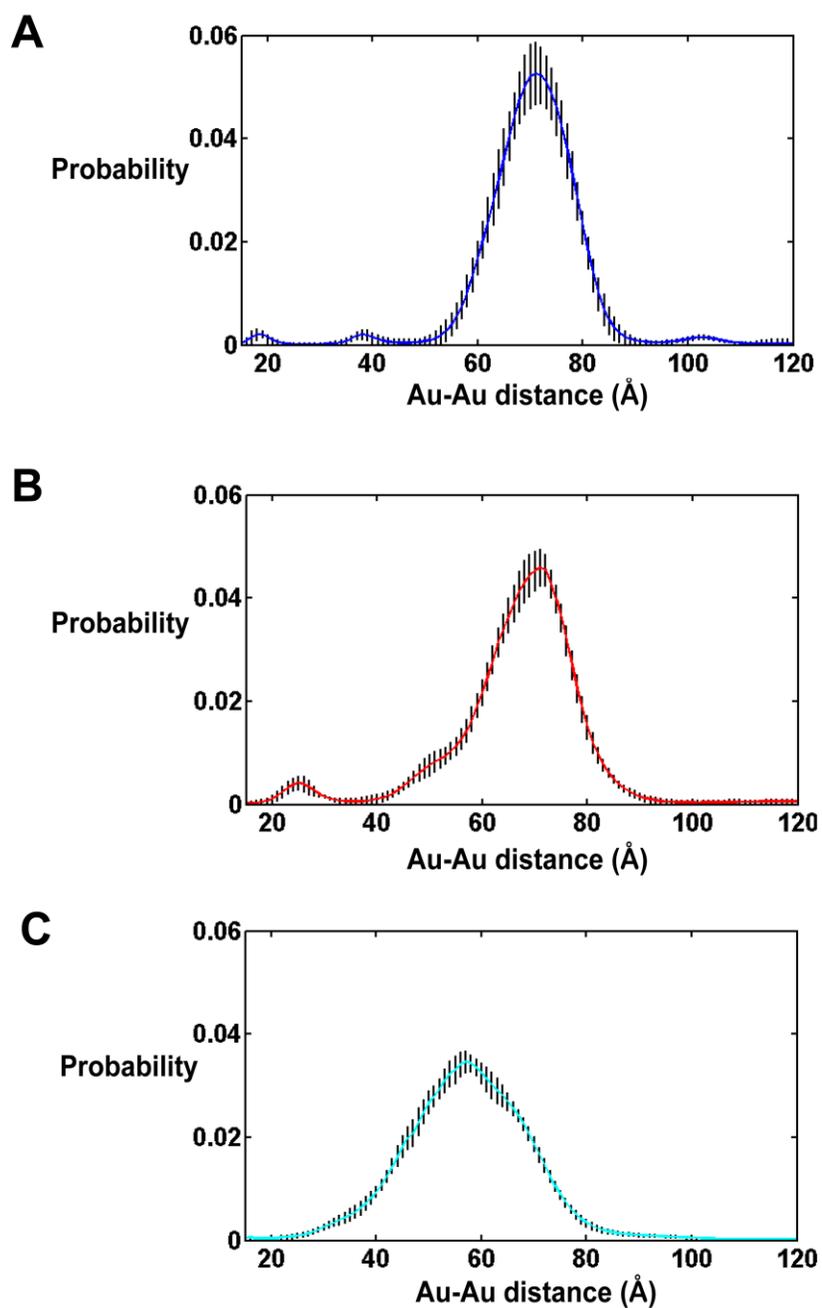
**Figure S16. Visualizing the literature smFRET average conformation in the x-ray interferometry ensemble (A)** The helical axis of the top (black lines) and bottom (blue lines) helix for the allowed space (left) and the x-ray interferometry ensemble (right) for 3A-DNA. The literature smFRET average conformation (S28) (red lines) is shown in the right panel. **(B)** The heat map for the x-ray interferometry reweighted ensemble. The position of the literature smFRET average conformation (S28) is labeled with a red star.



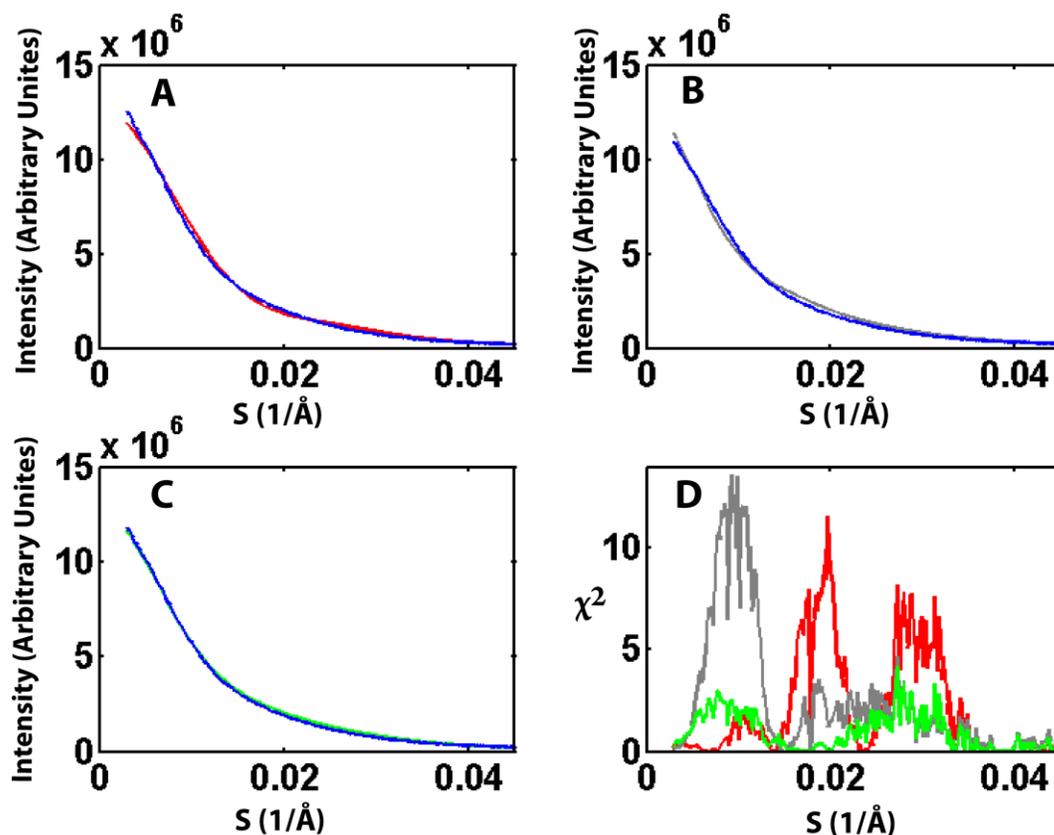
**Figure S17. The 3A-DNA ensemble obtained with different initial allowed spaces.** The 3A-DNA ensemble obtained with the allowed space used in the main text (Panel **A** and **B**, left column) were compared with the ensemble (Panel **A** and **B**, middle column) obtained with two alternative allowed spaces (Panel **A**, allowed space with MD-limiting values; Panel **B**, allowed space generated with MD-independent physical limiting values. See SI Methods for detail). The right column is the absolute difference between the two ensembles from the left and middle columns.



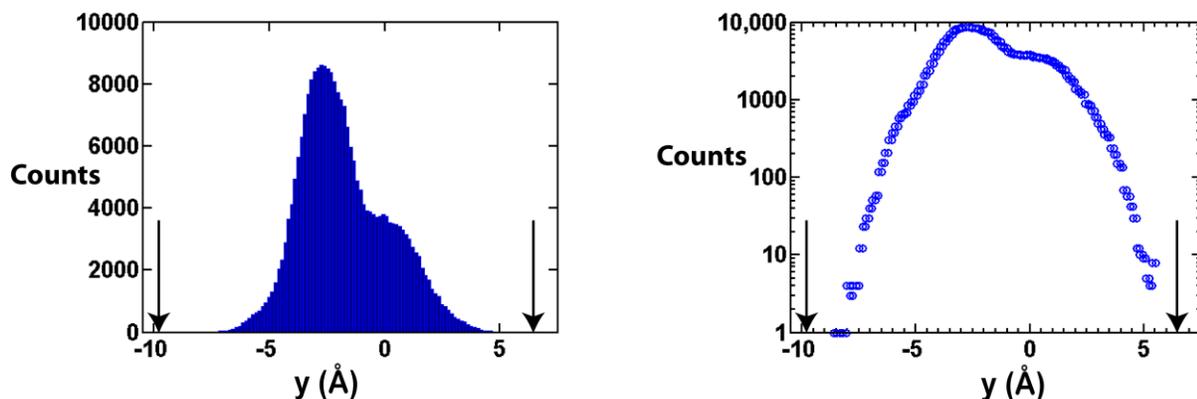
**Figure S18.** Correlations between rotation ( $\alpha$ ,  $\beta$  and  $\gamma$ ) and translation ( $x$ ,  $y$  and  $z$ ) for the estimated 3A-DNA ensemble. See Table S2 for the correlation coefficients.



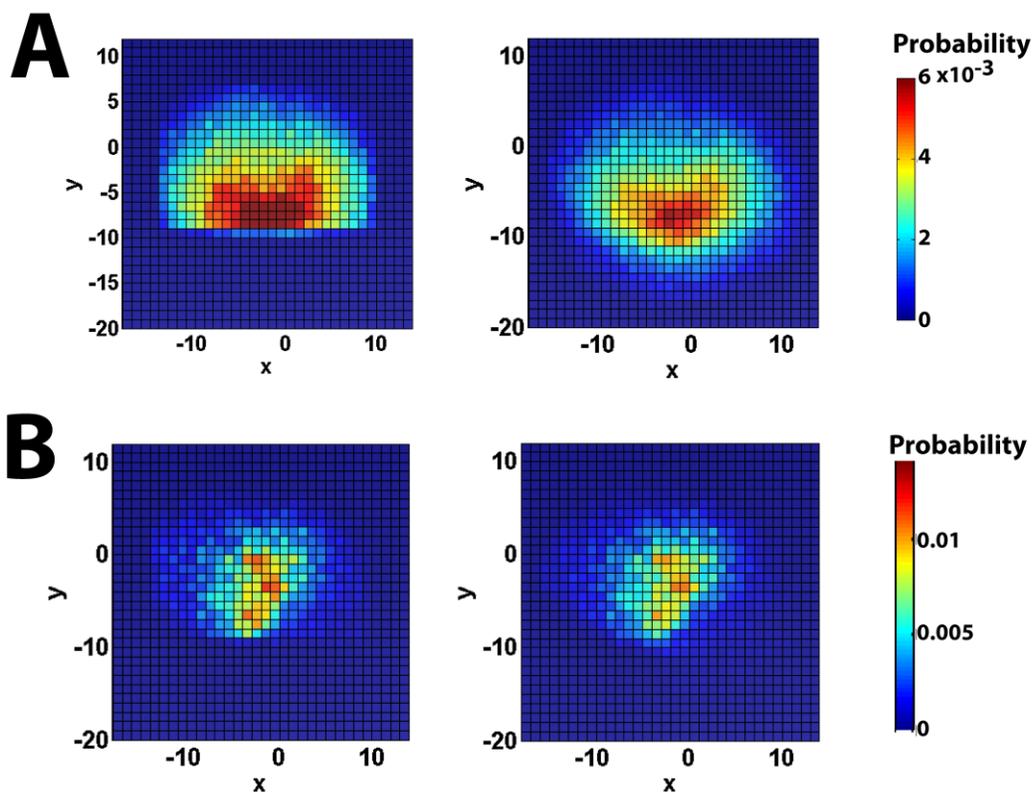
**Figure S19. Typical experimental errors in the Au-Au distance distribution.** To illustrate the typical experimental errors in the Au-Au distance distribution, we re-plotted the 0A (A, blue) and 1A (B, red) bulge data in Figure 1D and the 3A (C, cyan) bulge data in Figure 1D and 2C with error bars (black vertical lines). The experimental errors are calculated as the standard deviation from 10 repeated measurements of the same sample.



**Figure S20.** Comparison of the experimental SAXS profile of non-labeled DNA-3A bulge (blue in **A**, **B**, and **C**) and the predicted SAXS profile of (**A**) the literature (S28) smFRET model (red), (**B**) a straight helix of the same molecular weight (grey) and (**C**) the ensemble model determined in this paper (green). The SAXS profiles were predicted with the FoXS software (S29). The  $\chi^2$  values are as a function of scattering angle ( $S$ ) (**D**). The average  $\chi^2$  is 0.8, 1.4 and 1.7 for the ensemble model (green), the smFRET average structure model (red) and the straight helix (grey), respectively.



**Figure S21. Obtaining the y-limit of the allowed space from MD conformation library.** The histogram of the MD conformation library generated from MD simulation at 288, 375 and 450 K are plotted against the y translation values in linear (left) and log (right) scale. The maximum and minimum allowed y values (arrows) used for the allowed space are set as 1 Å beyond the MD limit.



**Figure S22. The 3A-DNA ensemble obtained with different initial allowed conformational space.** The 3A-DNA ensemble (**B**, left panel) obtained with the allowed space used in the main text (**A**, left panel) were compared with the ensemble (**B**, right panel) obtained with an alternative allowed spaces (**A**, right panel) that does not have a MD-derived preset limit in x, y and z values, but rather allows all sterically and connectively accessible conformations.

## References

1. Davis IW, *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:W375-W383.
2. Olson WK, *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* 313(1):229-237.
3. Goldstein H (2002) *Classical mechanics* (Addison Wesley, San Francisco :) 3rd ed. Ed.
4. Bailor MH, Mustoe AM, Brooks CL, & Al-Hashimi HM (2011) 3D maps of RNA interhelical junctions. *Nat Protoc* 6(10):1536-1545.
5. Margenau H & GM M (1956) *The Mathematics of Physics and Chemistry* (D. van Nostrand, New York).
6. Bailor MH, Sun X, & Al-Hashimi HM (2010) Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* 327(5962):202-206.
7. 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.
8. 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.
9. 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).
10. Mathew-Fenn RS, Das R, & Harbury PA (2008) Remeasuring the double helix. *Science* 322(5900):446-449.
11. Fisher CK, Huang A, & Stultz CM (2010) Modeling intrinsically disordered proteins with bayesian statistics. *J. Am. Chem. Soc.* 132(42):14919-14927.
12. Carlin BP & Louis TA (2000) The empirical Bayes approach. *Bayes and Empirical Bayes Methods for Data Analysis, Second Edition*, (Chapman and Hall).
13. Lu XJ & Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31(17):5108-5121.
14. Kabsch W (1976) Solution for best rotation to relate 2 sets of vectors. *Acta Crystallogr A* 32(Sep1):922-923.
15. Dolinsky TJ, Nielsen JE, McCammon JA, & Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32:W665-W667.
16. Bailor MH, Mustoe AM, Brooks CL, & Al-Hashimi HM (2011) Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation. *Curr Opin Struc Biol* 21(3):296-305.
17. 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.
18. Cruz JA, *et al.* (2012) RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18(4):610-625.
19. Moutl J (2005) A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struc Biol* 15(3):285-289.

20. Lensink MF, Mendez R, & Wodak SJ (2007) Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins* 69(4):704-718.
21. Hall KB (2009) 2-Aminopurine as a probe of RNA conformational transitions. *Methods Enzymol.* 469:269-285.
22. Jean JM & Hall KB (2002) 2-Aminopurine electronic structure and fluorescence properties in DNA. *Biochemistry* 41(44):13152-13161.
23. Larsen OFA, van Stokkum IHM, Gobets B, van Grondelle R, & van Amerongen H (2001) Probing the structure and dynamics of a DNA hairpin by ultrafast quenching and fluorescence depolarization. *Biophys. J.* 81(2):1115-1126.
24. Bharill S, *et al.* (2008) Fluorescence intensity decays of 2-aminopurine solutions: lifetime distribution approach. *Anal. Biochem.* 377(2):141-149.
25. Nordlund TM, *et al.* (1989) Structure and dynamics of a fluorescent DNA oligomer containing the EcoRI recognition sequence - fluorescence, molecular-dynamics, and NMR-studies. *Biochemistry* 28(23):9095-9103.
26. Wan CZ, Fiebig T, Schiemann O, Barton JK, & Zewail AH (2000) Femtosecond direct observation of charge transfer between bases in DNA. *P Natl Acad Sci USA* 97(26):14052-14055.
27. Jean JM & Hall KB (2004) Stacking-unstacking dynamics of oligodeoxynucleotide trimers. *Biochemistry* 43(31):10277-10284.
28. 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.
29. 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.