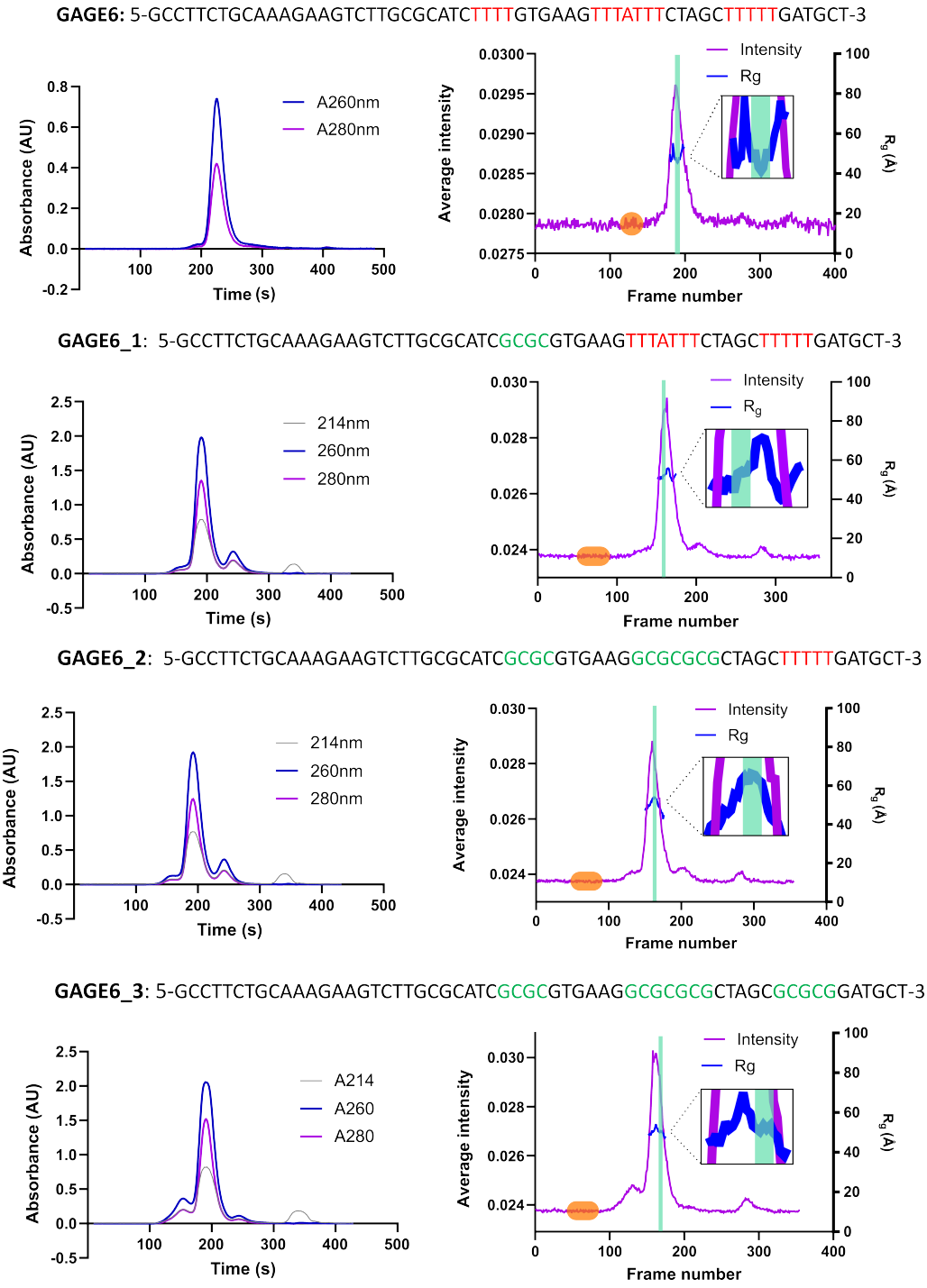
Supplementary Materials:

Probing the statistics of sequence-dependent DNA conformations in solution using SAXS

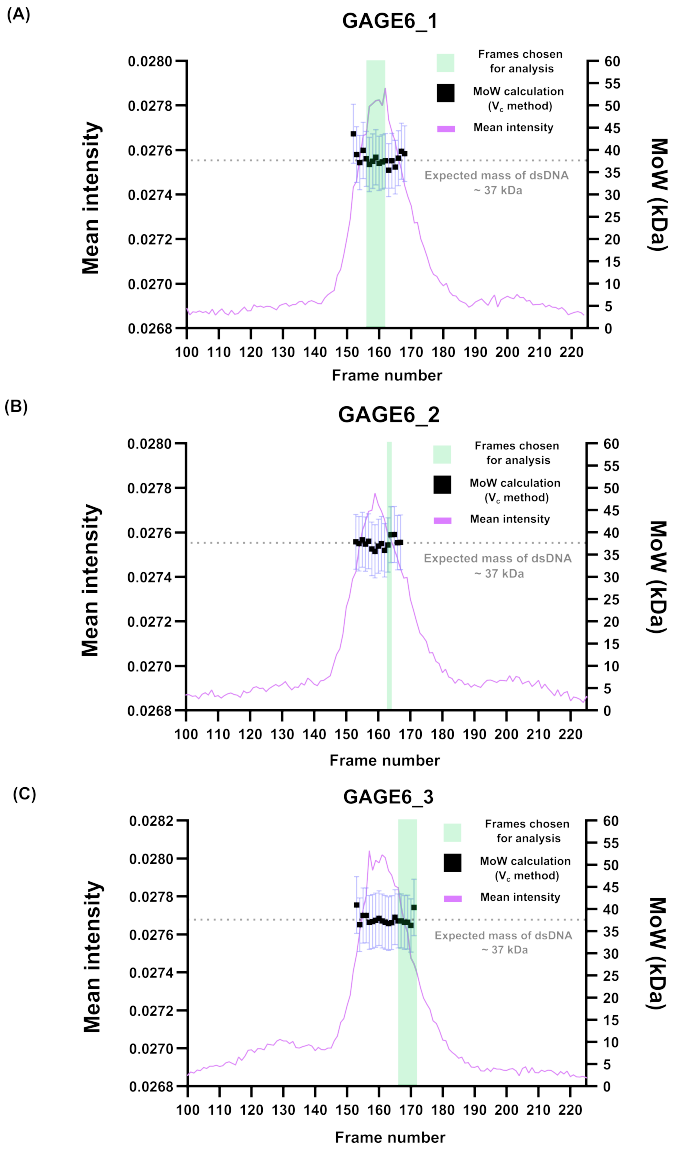
Heidar J. Koning, Anuradha Pullakhandam, Andrew E. Whitten, Charles S. Bond and Michel Peyrard



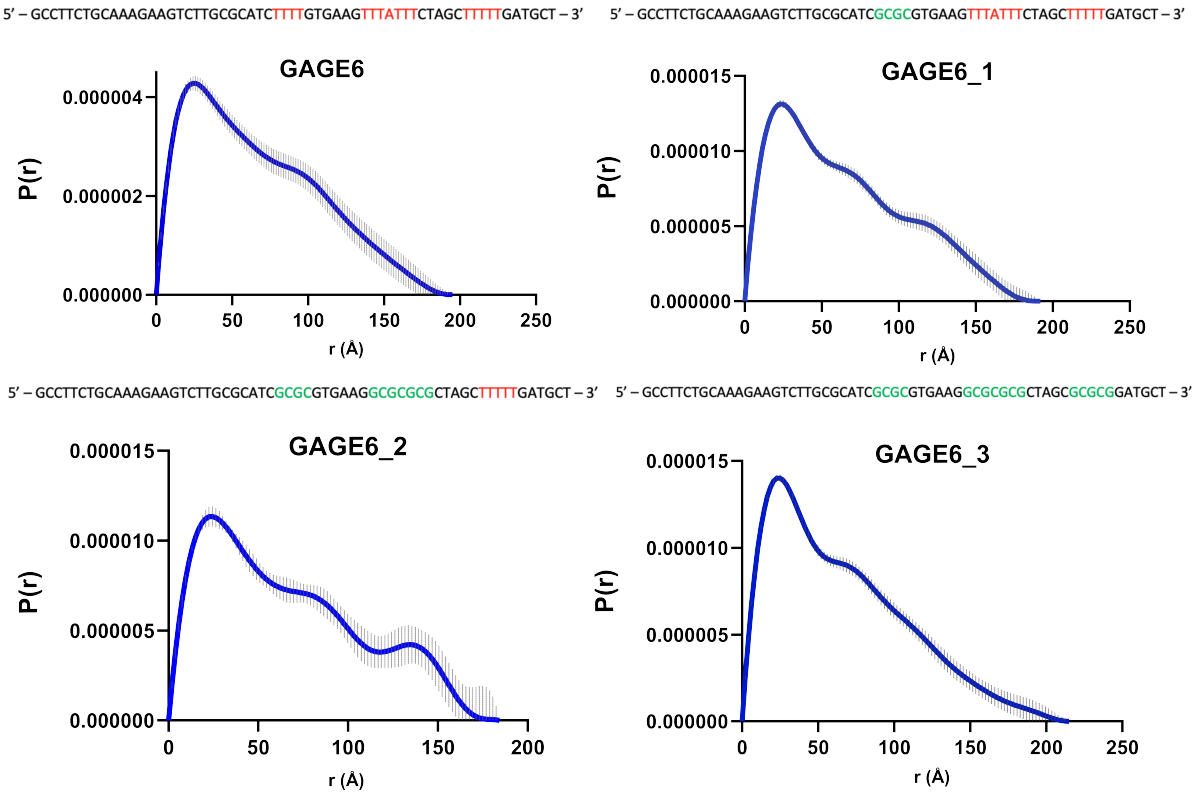
Supplementary Figure S1: Figure S1 shows the absorbance traces of all the duplexes at 214nm, 260nm and 280nm vs time (seconds) followed by scattering traces of all the duplexes plotted from CRHOMIXS showing average X-ray intensity against frame number. Predicted Rg values are shown as blue curves and were predicted using CHROMIXS. The regions used for buffer subtraction are shown in orange and the regions used for sample processing in green.

$\Delta q = 0.001$ Å^{-1}	GAGE6	GAGE6_1	GAGE6_2	GAGE6_3
GAGE6	-	-	-	-
GAGE6_1	0.493	-	-	-
GAGE6_2	0.280	0.493	-	-
GAGE6_3	0.078	0.493	0.753	-

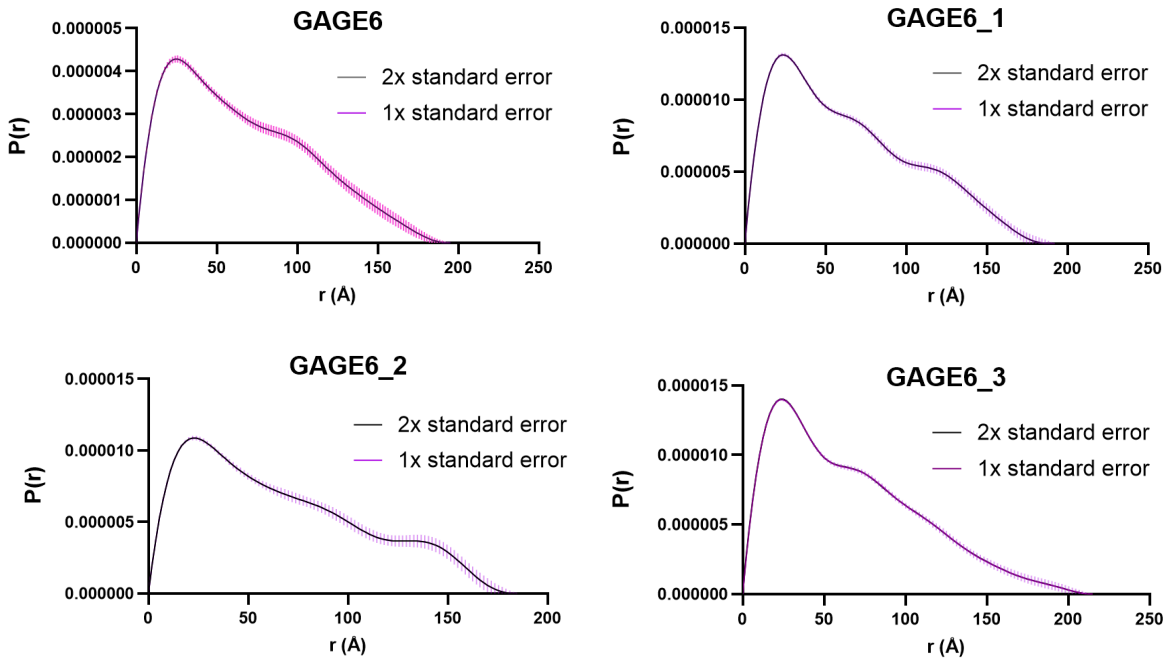
Supplementary Table 1: Correlation Map Test P-values after regridding the data truncated to $q \leq 0.18$ Å^{-1} which is the domain that we use in the analysis.



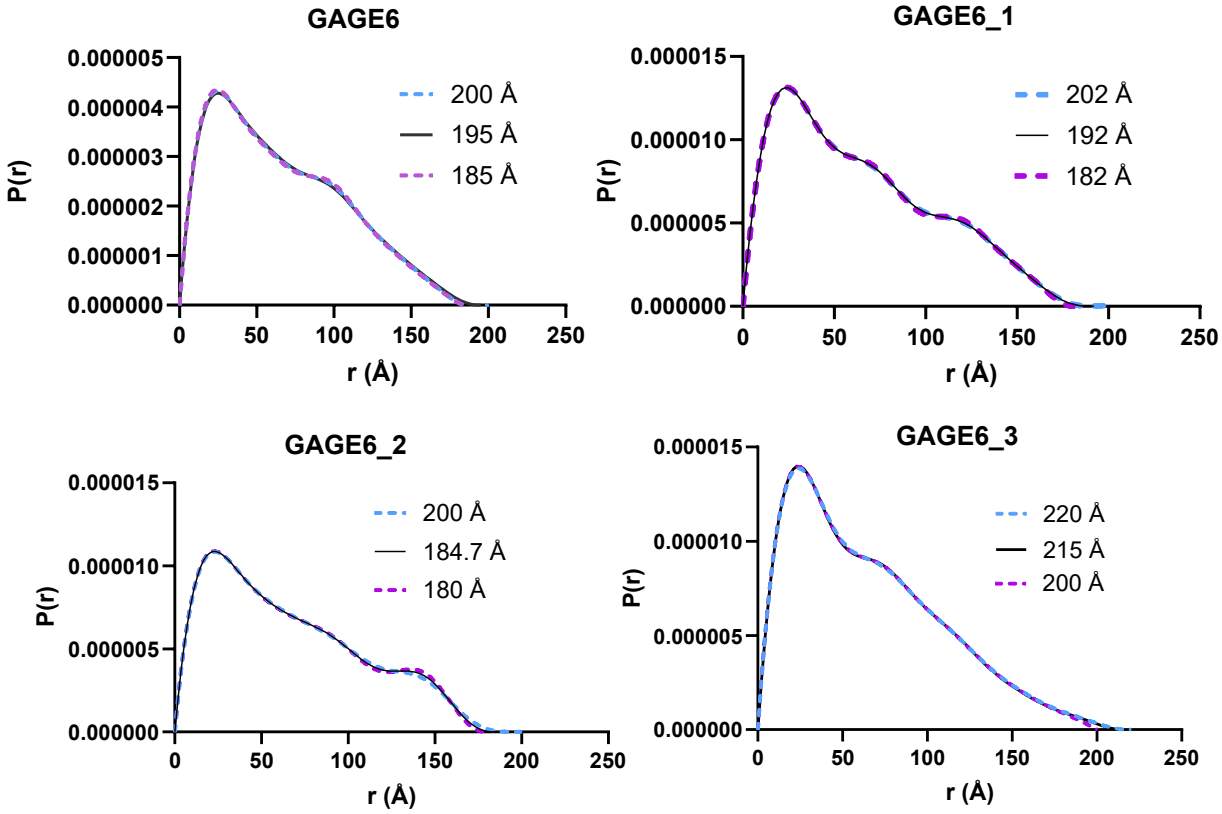
Supplementary Figure S2: (A) The GAGE6_1 mean intensity vs. frame number chromatogram (pink) shown alongside the results of the MoW calculation (black squares with light blue error bars) as performed by BioXTAS RAW. The expected mass of a 60bp dsDNA duplex is shown as a grey dotted horizontal line. The frames chosen for downstream analysis of P(r) are highlighted by a light green box. (B) The same as above but for GAGE6_2. (C) The same as above but for GAGE6_3.



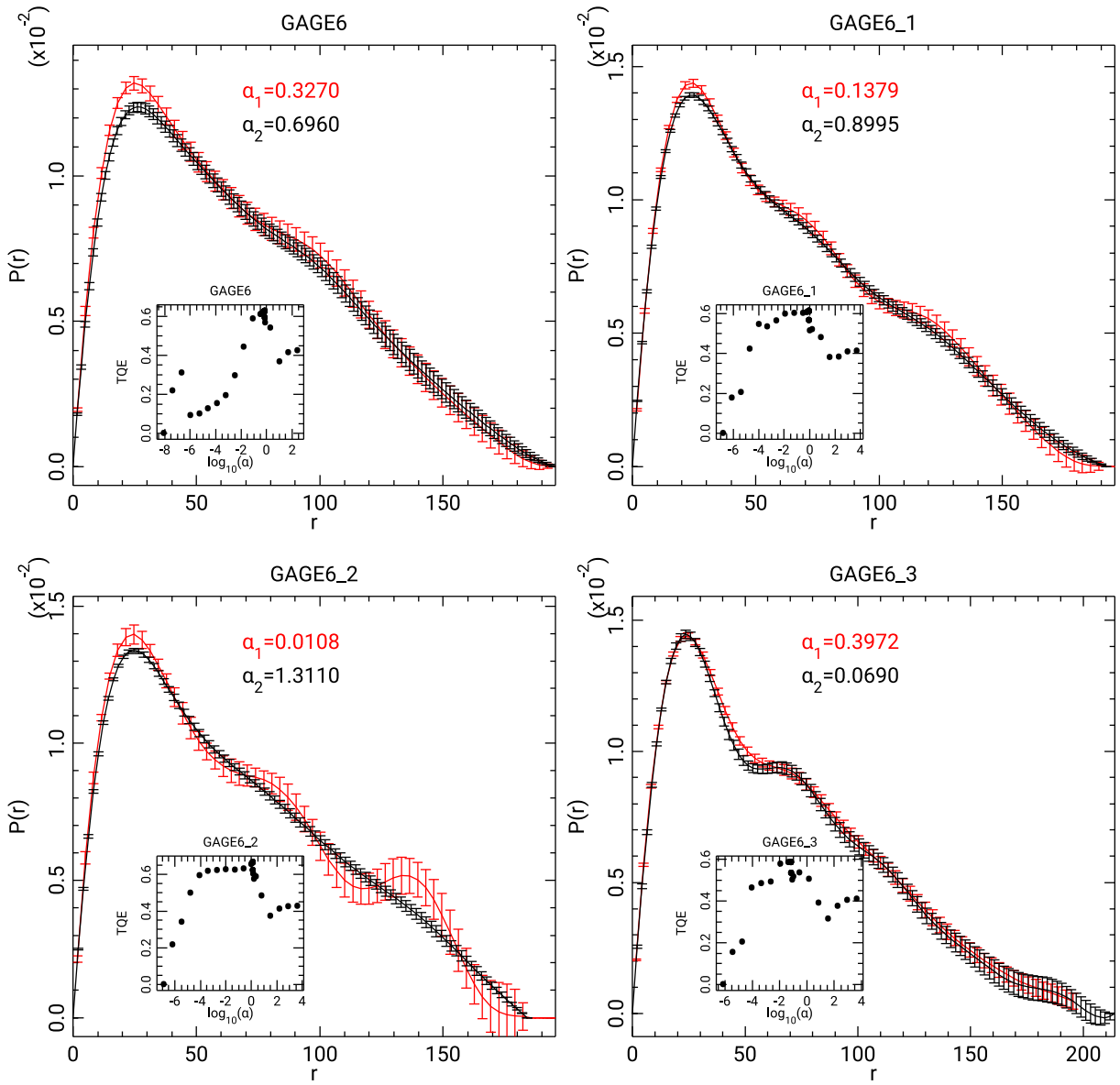
Supplementary Figure S3: Comparative distance distribution functions AutoGNOM-5-$P_{\mathrm{exp}}(r)$ of the different oligonucleotides used in this study. Sequence of each oligonucleotide indicated above $P_{\mathrm{exp}}(r)$ function with AT tracts highlighted in red and GC tracts in green. Sequence-dependent changes in the $P_{\mathrm{exp}}(r)$ function indicate the position of multiple bumps in the AT tract containing duplexes and a smooth descent for GAGE6_3.



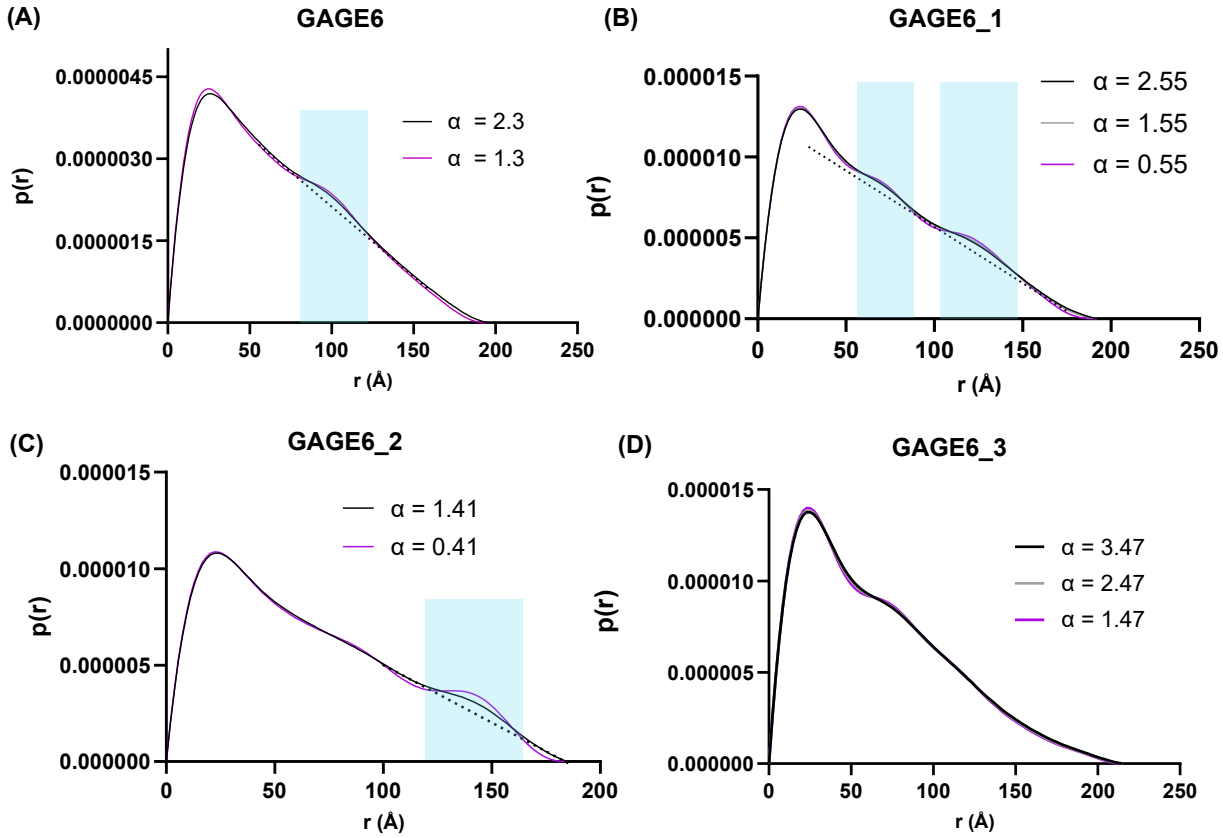
Supplementary Figure S4: Comparative AutoGNOM-5-$P_{\mathrm{exp}}(r)$ functions for datasets accompanied by 1x (pink line) and 2x (black line) standard error show consistency in the shape of P(r). Errors bars only plotted for data from 1x standard error (pink).



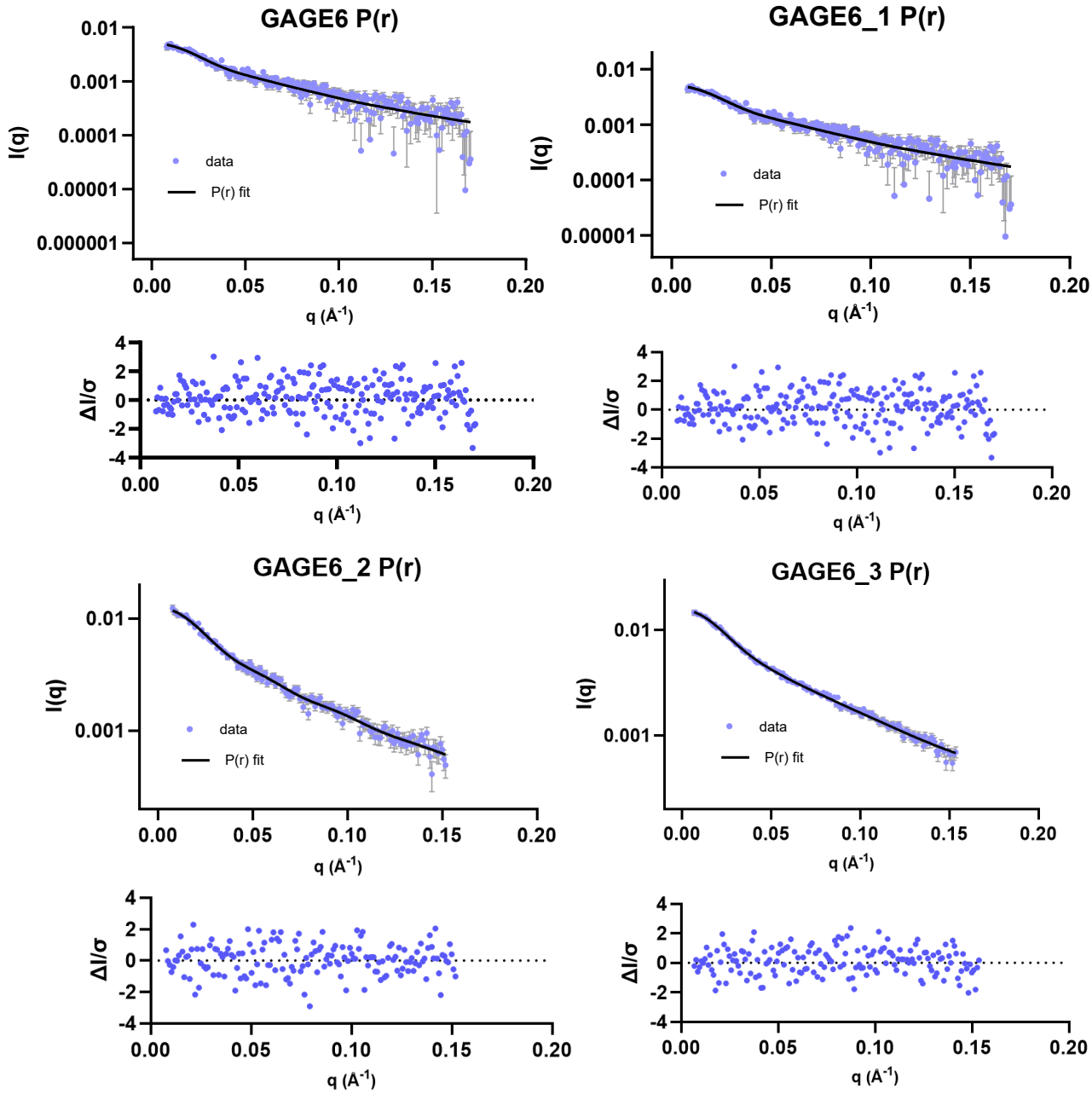
Supplementary Figure S5: Variation of D_{max} parameter around the reported values shows that AutoGNOM-5-$P_{\mathrm{exp}}(r)$ features remain consistently present.



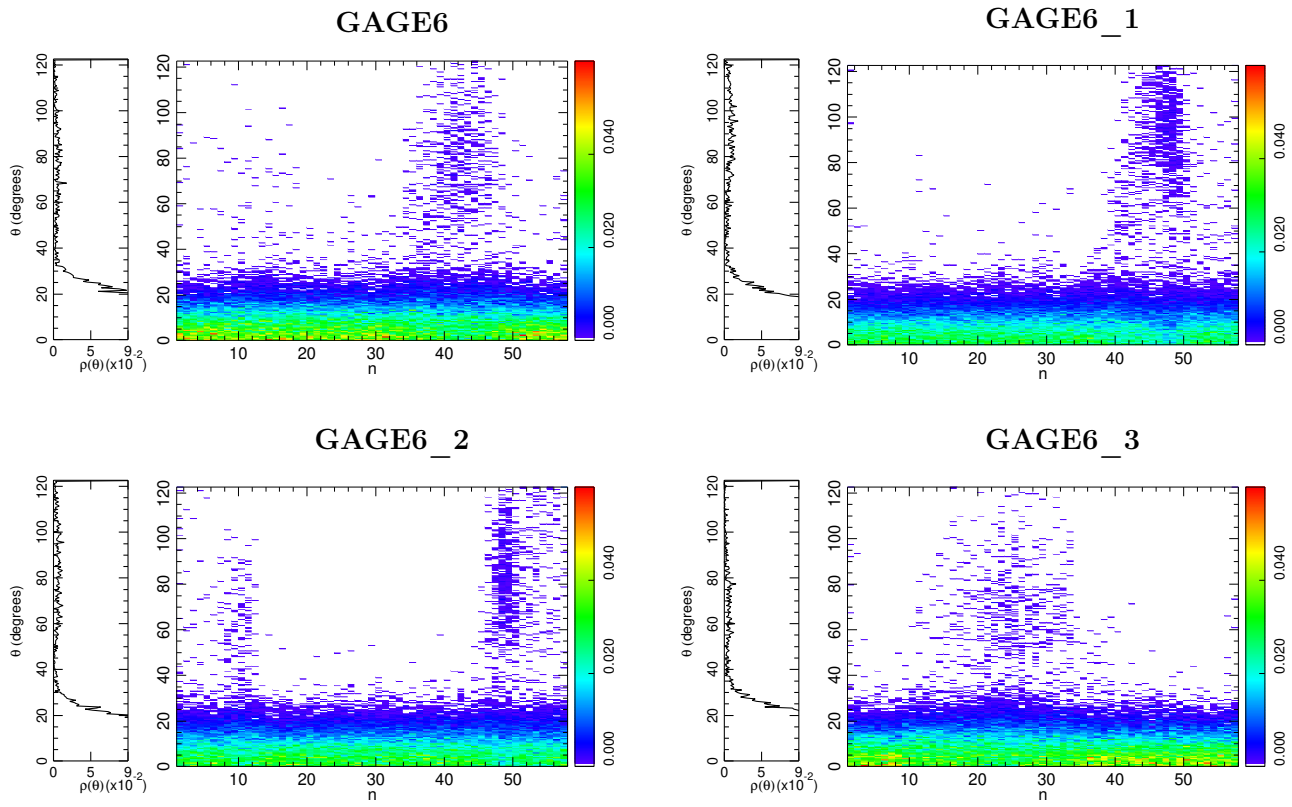
Supplementary Figure S6: Comparison between the functions $P_{\exp}(r)$ calculated with AutoGNOM-5 (red curve with error bars) and with AutoGNOM-4 (black curve with error bars) for the four DNA samples. The values of the regularisation parameter α used for each curve are indicated inside each panel. The inset panels show the variation of the Total Quality Estimate (TQE) versus α (in logarithmic scale) calculated by AutoGNOM-4.



Supplementary Figure S7: Variation of the GNOM alpha parameter around the fitted value indicates that while the magnitude of features changes, their position relative to r remains consistent.



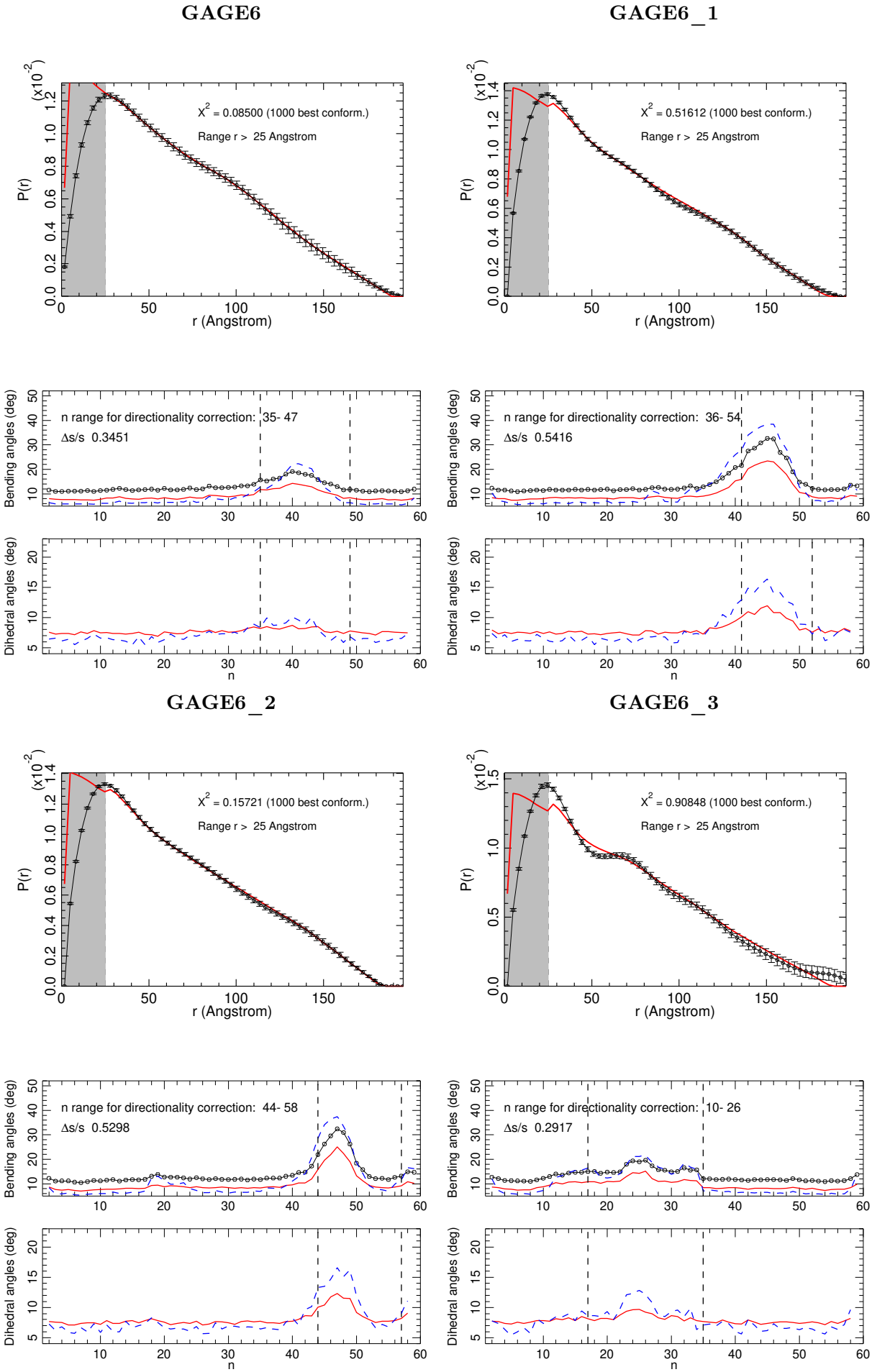
Supplementary Figure S8: Plots of the GNOM fits to the scattering data and their corresponding residual plots.



Supplementary Figure S9: Another view of the three-dimensional probabilities $P(n, |\theta|)$ which highlights some features that are difficult to see on the three-dimensional plots. The left part shows the sum over n of the projections of $P(n, |\theta|)$ on a single plane which displays the probability of a given value of $|\theta|$, whatever the position in the sequence:

$$\rho(\theta) = \sum_{n=2}^{N-1} P(n, |\theta|) \tag{1}$$

The right part uses a color code to show the three-dimensional probability in a two-dimensional image. This is easier to get an idea of the map of the bending angles along the sequence.



Supplementary figure S10: Analysis of the $P_{\exp}(r)$ obtained with Auto-GNOM-4 (see Fig. S6). For each sample the top panel shows the function $P_{\exp}(r)$ with error bars (black) and its fit by the polymer model (red full line). The bottom panel shows the statistics of the local bending and dihedral angles, as on Figs. 6, 7, 8, 10. The vertical dashed lines, limiting the domain showing the largest bending are at the same positions as on the corresponding figures of the paper. The particular case of GAGE6_2 is also shown in the paper as Fig. 9.