Supplementary Information for "Absolute Intra-Molecular Distance Measurements with Ångström-Resolution using Anomalous Small-Angle X-ray Scattering"

Thomas Zettl[†], *Rebecca S. Mathew*[‡], *Sönke Seifert*[§], *Pehr A.B. Harbury*[&], *Sebastian Doniach*[^], *and Jan Lipfert*[†,*]

[†] Department of Physics, Nanosystem Initiative Munich, and Center for Nanoscience, LMU Munich, Amalienstrasse 54, 80799 Munich, Germany

[‡] Department of Cell Biology, Harvard Medical School, Harvard University, Boston, MA, USA

[§]Experimental Facility Division, Argonne National Laboratory, Argonne, IL, USA

[&]Department of Biochemistry, Stanford University, Stanford, CA, USA

[^]Departments of Applied Physics and Physics, Stanford University, Stanford, CA, USA

[*]Corresponding author. E-mail address: Jan.Lipfert@lmu.de Phone: + 49-89-2180-2005

SUPPLEMENTARY METHODS

Sample preparation

Double-stranded DNA molecules 10, 20, and 30 bp in length and labeled at both ends with 7 Å radius thioglucose passivated gold nanoparticles were prepared as described previously[1, 2]. Briefly, water-soluble gold nanoparticles were synthesized as previously described[3]. Single-stranded oligonucleotides (ssDNA) were prepared on an automated ABI 394 DNA synthesizer and thiols were incorporated using a C3-thiol-modifier (Glen Research, part # 20-2933-41). ssDNA was purified by high-pressure liquid chromatography (HPLC) and coupled to gold nanoparticles in a 5:1 ratio (gold nanoparticle:ssDNA) at room temperature in 100 mM Tris-HCl, pH 9.0 for two hours. Gold nanoparticle-oligonucleotide conjugates were purified by ion-exchange HPLC and stored at -20°C. Concentrations were determined by absorbance using a NanoDrop ND-1000 (NanoDrop Technologies). The DNA sequences used in this study are reported in Table S1. All SAXS measurements used 70 mM Tris-HCl buffer, pH 8.0, with 100 mM NaCl and 10 mM ascorbic acid added.

ASAXS Measurements and Data Analysis

X-ray energy calibration

ASAXS experiments were performed at the BESSRC-CAT beamline 12-ID of the Advanced Photon Source[4] (APS). Before starting an ASAXS measurement series, the experimental energy was calibrated by scanning the X-ray energy in steps of 1 eV around the L-III absorption edge of gold with a 50 µm thick gold foil inserted into the beam path and recording the incident I_0 and transmitted I_t X-ray intensity. The position of the absorption edge was determined from the maximum of the numerical derivative of $log(I_0/I_t)$. The absorption edge position was referenced to the tabulated (http://henke.lbl.gov/optical_constants/) value for the L-III absorption edge of 11.919 keV by applying an additive (and small, typically 5-10 eV) offset to all data recorded during that measurement run. All energies reported in this work have been calibrated by this offset, unless otherwise noted.

ASAXS measurements

All ASAXS measurements were carried out at room temperature using a ~16 µl sample cell and sample holder[5], a sample to detector distance of 1.0 m, a CCD detector (Mar CCD165), and sample concentrations of 200 µM. For each gold-labeled DNA sample, 10 measurements of 2 s exposure time were obtained at different energies in the range from 11.719 keV

(200 eV below edge) to 11.969 keV (50 eV above edge). Control measurement using identical samples and exposure times at one energy exhibit no sign of radiation damage as repeat exposures are superimposable, within experimental error (**Supplementary Fig. S2**). Data were circularly averaged and buffer profiles recorded using identical procedures were subtracted for background correction. The scattering profiles recorded at different energies were interpolated to a common *s*-scale that comprised the lowest *s*-value of the lowest energy and the largest *s*-value of the highest energy with 1000 equally spaced increments in between, therefore avoiding extrapolation beyond the recorded *s*-value range for all energies.

Normalization of scattering profiles

We normalized the scattering profiles measured at different energies as follows. First, we obtained the values of $f_0 + f'(E)$ and of $f''(E)$ for gold from the Lawrence Berkeley X-ray data site (http://henke.lbl.gov/optical_constants/sf/au.nff). Second we interpolate the values for $f_0 + f'(E)$ and $f''(E)$ to 1 eV resolution. For the imaginary part $f''(E)$ we employed simply linear interpolation. To achieve the same resolution for the real part $f_0 + f'(E)$ of the gold atomic scattering factor, we fitted the following function to the tabulated values (**Supplementary Fig. S3a**):

$$f_0 + f'(E) = Re\left[a \cdot \ln\left({E_{edge}}^2 - E^2\right) + b \cdot \ln\left({E_{edge}}^2 + E^2\right)\right] \quad (4)$$

with E_{edge} = 11918.6 eV. The fitting parameters a and b were determined to be 1.915 and 2.033, respectively. Third, we calculated theoretical energy-dependent intensity scaling factors $int(E)$ for each measured energy using the interpolated atomic scattering factors as

$$int(E) = |f_{Au}(E)|^2 + 2 \cdot Re(f_{Au}(E)) \quad (5)$$

Forth, we computed the raw mean scattering intensity for each of the experimental scattering profiles (**Supplementary Fig. S3b**) by averaging the scattering intensity over all *s* values (**Supplementary Fig. S3c, red circles**). Fifth, the theoretical intensity factors $int(E)$ were multiplicatively scaled to match the raw intensities at the minimum intensity (which corresponds to the edge). Finally, all scattering profiles were normalized by multiplying by a factor equal to the ratio of the scaled theoretical intensity factor to the raw intensity (**Supplementary Fig. S3c, green circles**). Last, we applied an additive constant to the

scattering profiles to match the normalized intensity profiles at high angles ($s \geq 0.08$ Å^{-1}) to correct for fluorescence[6].

Matrix inversion to obtain partial scattering factors

We determine the gold label-gold label interference pattern $G_{Au\text{-}Au}(s)$ by matrix inversion of Eqn. 3 from the experimentally determined scattering profiles at the various energies (assembled in the matrix I) and from the matrix T that contains information about the energy and angle dependence of the scattering factors following the approach outlined by Pinfield and Scott[7]. The matrix equation is more explicitly visualized in **Supplementary Fig. 4a**.

The $(N_E \cdot N_s) \times (3 \cdot N_s)$ matrix T comprises the pre-computed $a_i(s)$, $b_i(s)$, and $c_i(s)$ that are all $N_s \times N_s$ square diagonal matrices containing label-label pair scattering factors, label-molecule pair scattering factors and intra-molecule pair scattering factors, respectively. The matrices $a_i(s)$ comprise the energy and angle dependent squared scattering factor of the gold label (**Supplementary Fig. 4b**). The individual matrices $a_i(s)$ for each energy E are constructed by multiplying the absolute value squared of the atomic scattering factor $|f_{Au}(E)|^2$ with a matrix that has the angle-dependent squared scattering factor of the gold label on the diagonal. To compute the scattering factor of the gold labels, we approximate the gold nano-crystal scattering by the scattering factor of a sphere with N_{Au} atoms and a radius R:

$$F_{Au}(s) = N_{Au} \cdot \frac{\sin x - x \cdot \cos x}{x^3} \qquad (6)$$

with $x = 2\pi \cdot R \cdot s$; for the gold nanocrystals used here[1] $N_{Au} = 78$ and $R = 7$Å. In practice, we obtain the best fits when using a distribution of spheres, distributed uniformly with a spread of ± 4% around $R = 7$ Å. The matrices $b_i(s)$ have the angle-dependent gold scattering factor $F_{Au}(s)$ on the diagonal and are multiplied by the real part of the gold atom scattering factor $Re(f_{Au}(E))$ for each energy (**Supplementary Fig. 4c**). Finally, the matrices $c_i(s)$ are constructed by multiplying the N_s x N_s identity matrix by the absolute value squared of the average atomic scattering factors of atoms in the DNA molecule at the respective energies, averaged over the atomic composition of our DNA constructs (**Supplementary Fig. 4d**). The atomic scattering factors for all relevant atoms and energies were obtained from (http://henke.lbl.gov/optical_constants/). Before matrix inversion, the matrix T is conditioned by dividing all entries of the matrices a, b, and c by their respective elements corresponding to the smallest scattering angle and highest energy. The vector of partial scattering vectors G

was computed by matrix inversion of Eqn. 3 using the matrix left division in Matlab, which determines the least square solution of the system of linear equations.

Inversion of the label-label interference pattern to obtain label-label distance distributions

We obtain the distance distributions *P(d)* by inverting the gold label-gold label structure factor $G_{Au\text{-}Au}(s)$ using the maximum entropy algorithm described previously[1, 2]. In brief, the gold label-gold label distance distribution *P(d)* is determined at discrete values *d* by decomposing the label-label interference pattern into a linear combination of scattering interference profiles $I_d(s)$, which correspond to pairs of gold nano-crystals at a distance *d:*

$$G_{Au-Au}(s) = \sum_{d=0}^{d_{max}} P(d) \cdot I_d(s) \qquad (7)$$

The sum runs over values of *d* from 1 to d_{max} in steps of 1 Å, where d_{max} was chosen to be approximately twice the mean label-label distance for each data set. The $I_d(s)$ are proportional to $sin(2\pi \cdot d \cdot s)/(2\pi \cdot d \cdot s)$ with a prefactor that corresponds to the scattering of a pair of gold nanocrystals at zero distance and was taken from Mathew-Fenn et al.[1].

In principle, the values of $P(d)$ at each value of *d* can be determined from Eqn. 7 using non-negative least squares. However, as was shown previously, a maximum entropy regularization procedure for the fit yields better results[1, 7] and was employed for the results presented in **Figure 3c**.

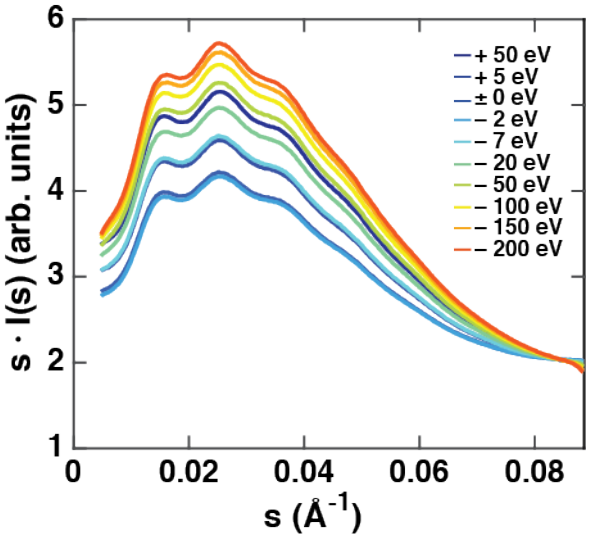
Analysis of label-label distance distributions

The mean and variance of the gold label-gold label distributions was determined from analysis of the main peaks using the *autopeak* routine in Matlab. All data are provided in Table S1. Errors were estimated to be 1% for the length and 10% for the variance from triplicate repeat measurements.

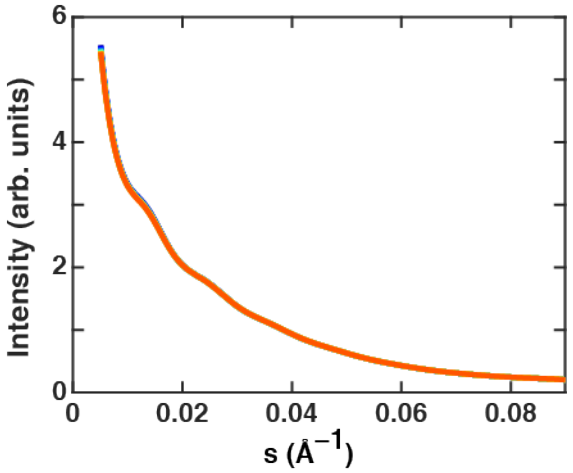
The mean of the label-label distances was fitted to the model described previously (Supplementary Figure 2 of Ref.[2]) with the rise per base pair and the radial displacement of the labels as free fitting parameters. The axial offset $axial_0$ and angle offset θ_0 were kept fixed at their previously determined values of 24 Å and 1.34 π, respectively. The rise per base pair was determined from the fit to be of 3.23 ± 0.1 Å and the radial displacement to be 7 ± 1 Å.

Table S1. DNA sequences and the mean and variances of the main peak in the gold label-label distance distributions obtained using the ASAXS method in this work and previously using single-energy SAXS.

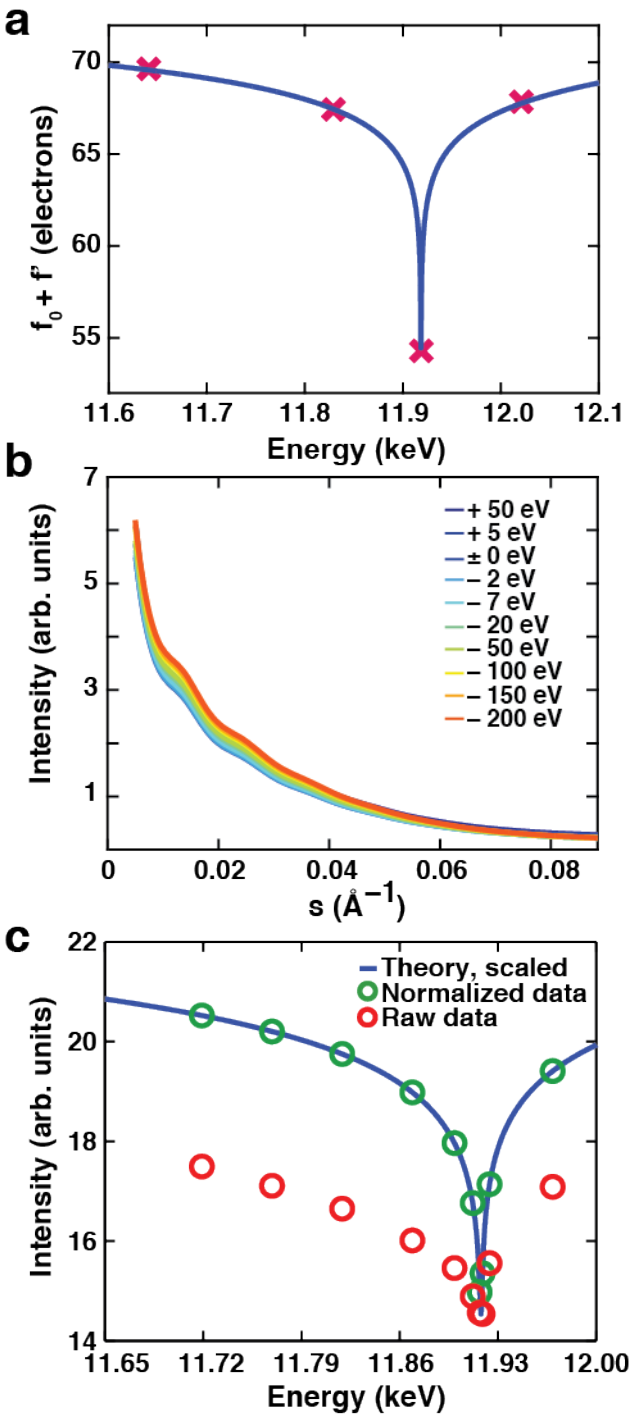
DNA construct	Sequence	Mean distance (Å) (This work)	Variance distance ($Å^2$) (This work)	Mean distance (Å) (Ref. [2])	Variance distance ($Å^2$) (Ref. [2])
10 bp	5'-GCATCTGGGC-3' CGTAGACCCG	54.3 ± 0.5	6.52 ± 0.7	55.7 ± 0.3	8.5 ± 0.6
20 bp	5'-CGACTCTACGGCATCTGCGC-3' GCTGAGATGCCGTAGACGCG	87.3 ± 0.9	25.8 ± 2.6	86.0 ± 0.4	21.6 ± 1.4
30 bp	5'-CGACTCTACGGAAGGTCTCGGACTACGCGC-3' GCTGAGATGCCTTCCAGAGCCTGATGCGCG	117.0 ± 1.2	52.0 ± 5.2	119.1 ± 0.6	41.1 ± 2.7



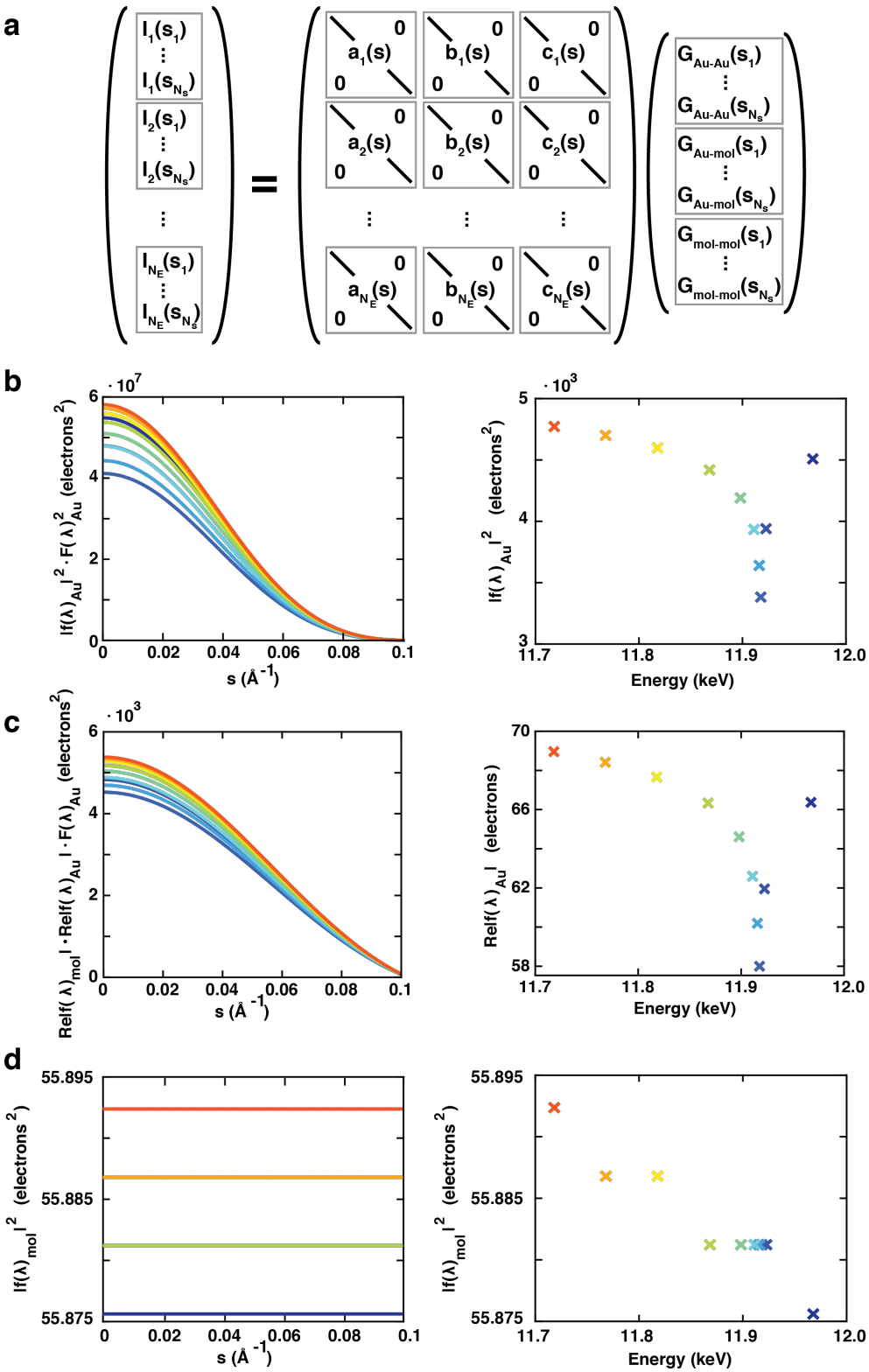
Supplementary Figure S1. Scattering intensities as a function of momentum transfer *s* for the 20 bp double labeled DNA construct at 10 different energies shown in Holtzer representation of $s \cdot I(s)$ as a function of *s* that emphasizes differences in the intermediate *s* range (same data as in **Fig. 3a**). Energies in the figure legend are relative to the gold L-III edge.



Supplementary Figure S2. Scattering intensity as a function of momentum transfer *s* for the 20 bp double gold labeled DNA construct. 10 subsequent exposures (color coded from blue to red) of 2 s each all at the same X-ray energy of 11.911 keV (8 eV below the gold L-III edge) are shown. Scattering profiles are identical within experimental error, indicating the absence of radiation damage.

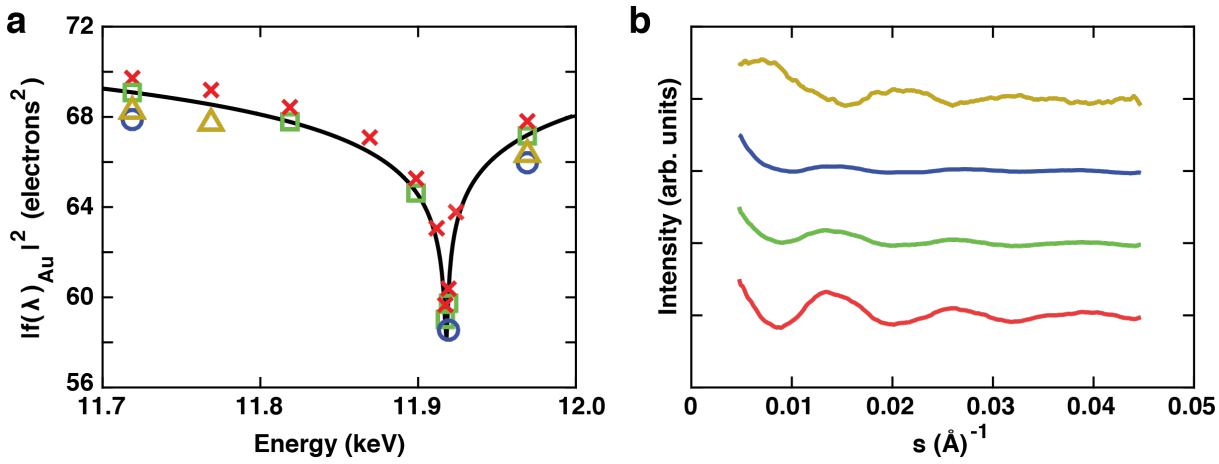


Supplementary Figure S3. Normalization of ASAXS profiles. a) Real part of the atomic scattering factors for gold close to the L-III edge. Crosses are the tabulated values obtained from http://henke.lbl.gov/optical_constants/sf/au.nff. The solid blue line as a fit of Eqn. 4 to the data (see Online Methods for details). **b)** Scattering intensities as a function of momentum transfer *s* for a 20 bp double labeled DNA sample at 10 different energies prior to normalization. The same data set after the application of the energy dependent normalization is shown in **Fig. 3a**. **c)** Normalization factors for the energy dependent normalization. Raw intensity values, obtained by averaging scattering intensities at each energy over all *s* values are show as red circles. The theoretical scattering intensity scaling curve (**Eqn. 5**) adjusted to the minimum value in the raw intensity data is shown as a blue line. Scaled mean intensity values are shown as blue circles. For details of the normalization procedure see Online Methods.

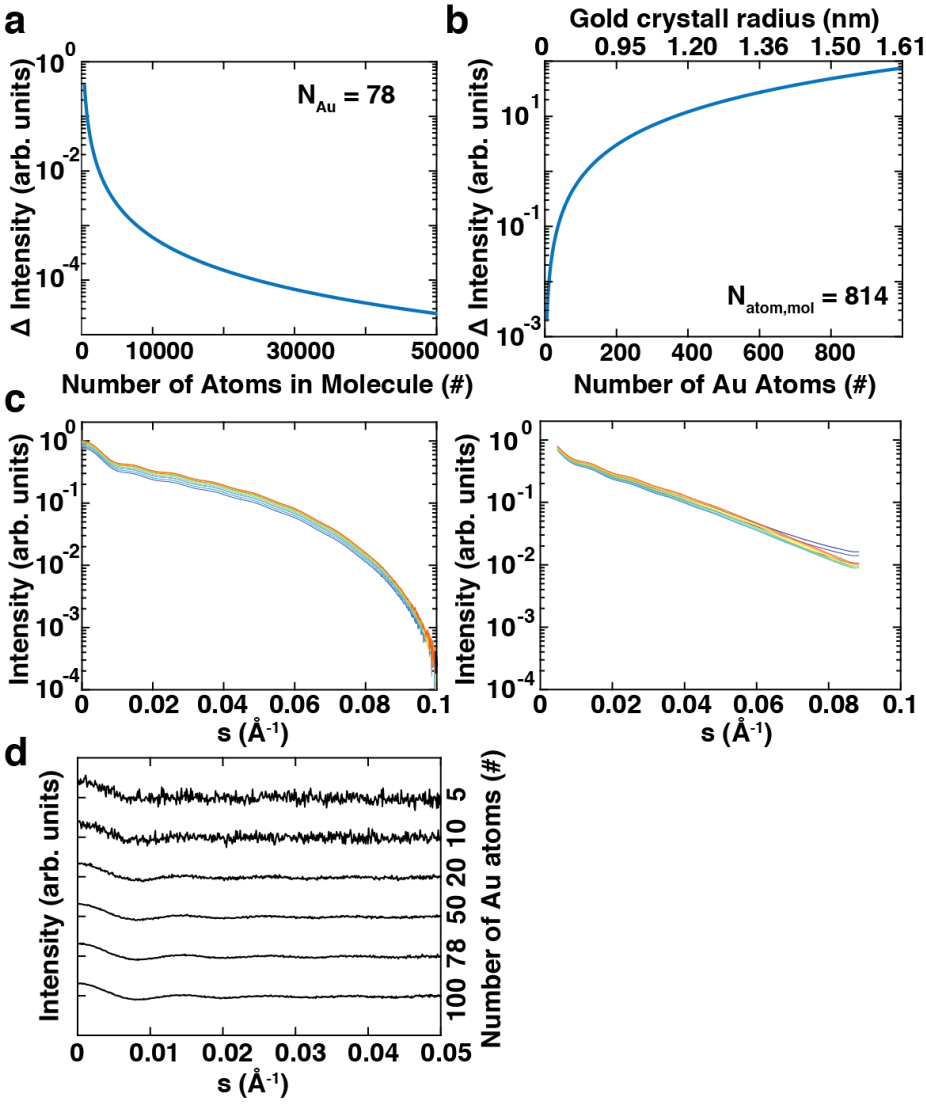


Supplementary Figure S4. Explicit form of the matrices $a_i(s)$, $b_i(s)$, and $c_i(s)$ used in the matrix inversion. a) Structure of the matrix equation (Eqn. 3 in the main text) connecting the experimental measured scattering intensities $I_i(s)$ to the partial structure factors $G_{Au\text{-}Au}$, $G_{Au\text{-}mol}$, and $G_{mol\text{-}mol}$ by multiplication with the pair scattering factors of the gold-gold terms a_i, pair scattering factors of the cross terms b_i, and pair scattering factors of the molecule-molecule terms c_i at energies E_i. **b)** Plots of the diagonal entries of the a_i matrices, i.e. the pair scattering factors of the gold-gold terms a_i as a function of s (left), which are the scattering functions for the spherical gold labels with a radius of 0.7 nm (Equation 6, squared) scaled by the absolute value squared of the atomic scattering factors of gold as function of energy (plotted on the right). **c)** Plots of the diagonal entries of the b_i matrices, i.e. the

pair scattering factors of the gold-molecule cross terms (left). The *s* dependency is caused by the spherical gold label scattering factor (Equation 6); the energy dependence is given by real part of the gold atomic scattering factors (right). **d)** Plots of the diagonal entries of the pair scattering factors for the molecule-molecule terms c_i (left). There is no dependence on *s,* therefore, the scattering factors are horizontal lines (left). The energy dependence is given by the average of the atomic scattering factors for DNA molecules as function of energy (right). Since the atomic scattering factors for the elements making up biological samples are almost constant in the energy range considered, the c_i show almost no energy dependence (compare the y-scales of the right plots in panels b, c, and d).



Supplementary Figure S5. Stability of the partial structure factor computation using scattering profiles at different energies. Using the data set shown in Fig. 3 of the main text as a starting point, we selected sub-sets of the scattering profiles at different energies and computed the gold-gold interference term by inversion of Eqn. 3 as described in the main text. **a)** Atomic scattering factor of gold as a function of energy (black line). The sets of energies selected for different calculations of the gold-gold interference term are shown as differently colored symbols. The symbols are slightly vertically offset for clarity. **b)** Gold-gold partial structure factors computed using sub-sets of the scattering profiles at different energies. The color code corresponds to the symbols in panel a. Profiles are vertically offset for clarity. Using 10 different energies (red crosses in panel a), i.e. the full data set, for inversion yields a well defined partial structure factor (red line in panel b), as discussed in the main text. Using only 6 energies (green squares in panel a) or 3 energies (blue circles in a) positioned below, on, and above the absorption edge still enables inversion of the matrix equation and gives rise to gold-gold partial structure factors with the expected oscillating pattern but with (much) reduced amplitudes (green and blue lines in panel b). Using scattering profiles at 3 energies away from the gold L-III absorption edge (dark yellow triangles in panel a) -and therefore with little change in scattering intensity as a function of energy- is insufficient for robust inversion of the matrix equation; the resulting gold-gold partial structure factor is very noisy and shows spurious oscillations (dark yellow line in panel b).



Supplementary Figure S6. Relative scattering intensities of the gold labels and molecules. The level of the anomalous signal from gold nanocrystals relative to the scattering from the molecule is evaluated by the relationship

$$\Delta I = \frac{N_{Au}^2 \Delta |f_{Au}(\lambda)|^2}{N_{mol}^2 <|f_{mol}|^2>} \qquad (8)$$

where N_{Au} is the number of gold atoms in the nanocrystals, N_{mol} is the number of atoms in the labeled macromolecule, $<|f_{mol}|^2>$ is the absolute value squared of the average atomic scattering factor of the macromolecule (averaged over the element composition of the molecule), and $\Delta|f_{Au}|^2$ is the change in the absolute value squared of the atomic scattering factor for gold, evaluated on and off edge.

a) Ratio between the energy dependent variations of the gold scattering term and the molecular scattering contribution (computed using Eqn. 8) as a function of the size of the macromolecule (in atoms). The number of gold atoms in the label was kept constant at 78, corresponding to the size of labels used in this study. **b)** Ratio between the energy dependent variations of the gold scattering term and the molecular scattering contribution (computed using Eqn. 8) as function of the number of gold atoms in the label. The corresponding approximate gold nanocrystal radius is indicated in the top axis. The number of atoms in the molecule was kept constant at 814 atoms, corresponding to the 20 bp DNA used in this work. Based on our experiments, we can estimate that a level of $\Delta I \sim 0.1$ is sufficient for reliable inversion; lower values of the intensity variation ΔI might still be sufficient for the computation of distance distributions for high quality measurements.

c) Simulated intensity profiles for double labeled 20 bp DNA molecules, assuming gold labels with 78 gold atoms, with random noise of 0.01% I_0 added to the scattering profile (left) and experimentally measured profiles for the double labeled 20 bp DNA (right). The simulations used DNA structures

generated with the make-na server (http://structure.usc.edu/make-na/server.html) and followed the approach of Pinfield and Scott[7]. The energy color code is the same as in Fig. 3. While random errors with a fixed magnitude of 0.01% I_0 do not completely describe the error encountered in the real experiments, this level of error is approximately representative of our experimental noise.
d) Simulated label-label partial structure factors for double-labeled 20 bp DNA with different gold label sizes as a function of s. Simulations were carried out as shown in panel c), for different label sizes; label-label partial structure factors were computed with the same routines as used in the analysis of the experimental data. The amplitude of the oscillating interference pattern is decreasing with decreasing label size (bottom to top), while the noise level is visibly increasing.

Supplementary References

1. Mathew-Fenn, R. S.; Das, R.; Silverman, J. A.; Walker, P. A.; Harbury, P. A. *PLoS One* **2008,** 3, (10), e3229.
2. Mathew-Fenn, R. S.; Das, R.; Harbury, P. A. *Science* **2008,** 322, (5900), 446-9.
3. Schaaff, T. G.; Knight, G.; Shafigullin, M. N.; Borkman, R. F.; Whetten, R. L. *The Journal of Physical Chemistry B* **1998,** 102, (52), 10643-10646.
4. Beno, M. A.; Jennings, G.; Engbretson, M.; Knapp, G. S.; Kurtz, C.; Zabransky, B.; Linton, J.; Seifert, S.; Wiley, C.; Montano, P. A. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **2001,** 467–468, Part 1, 690-693.
5. Lipfert, J.; Millett, I. S.; Seifert, S.; Doniach, S. *Review of Scientific Instruments* **2006,** 77, (4), 046108.
6. Stuhrmann, H. B.; Notbohm, H. *Proc Natl Acad Sci U S A* **1981,** 78, (10), 6216-20.
7. Pinfield, V. J.; Scott, D. J. *PLoS One* **2014,** 9, (4), e95664.