Absolute Intramolecular Distance Measurements with Angstrom-Resolution Using Anomalous Small-Angle X-ray Scattering

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

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

[‡]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, United States

[§]X-ray Science Division, Argonne National Laboratory, Argonne, Illinois 60439, United States

[‖]Department of Biochemistry and [⊥]Departments of Applied Physics and Physics, Stanford University, Stanford, California 94305, United States

S *Supporting Information*

ABSTRACT: Accurate determination of molecular distances is fundamental to understanding the structure, dynamics, and conformational ensembles of biological macromolecules. Here we present a method to determine the full distance distribution between small (~7 Å radius) gold labels attached to macromolecules with very high-precision (≤1 Å) and on an absolute distance scale. Our method uses anomalous small-angle X-ray scattering close to a gold absorption edge to separate the gold–gold interference pattern from other scattering contributions. Results for 10–30 bp DNA constructs achieve excellent signal-to-noise and are in good agreement with previous results obtained by single-energy SAXS measurements without requiring the preparation and measurement of single labeled and unlabeled samples. The use of small gold labels in combination with ASAXS read out provides an attractive approach to determining molecular distance distributions that will be applicable to a broad range of macromolecular systems.
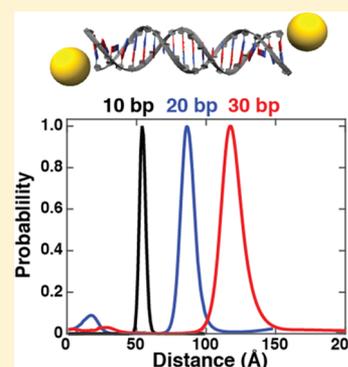
KEYWORDS: *Intramolecular distances, molecular ruler, anomalous small-angle X-ray scattering, SAXS, gold nanocrystals, DNA*

Measurements of molecular distances are key to dissecting the structure, dynamics, and functions of biological macromolecules. While (Förster) fluorescence energy transfer (FRET) and electron (EPR) or nuclear magnetic resonance (NMR) techniques have provided invaluable details by measuring intramolecular distances, they suffer from a limited range (typically <10 nm) and difficulties in converting the measured signal into absolute distances or, better yet, complete distance distributions.[1−5] SAXS measurements employing gold nanoclusters as labels attached to DNA molecules have demonstrated their ability to provide information about the entire label–label distance distribution for a considerable range of distances ranging from 5 up to 40 nm.[1,2,6−9] The approach has provided a detailed view of DNA structure and flexibility,[1,8,9] revealed conformational changes of DNA upon protein binding,[7] and probed the conformational landscape of a complex RNA motif in response to solution conditions and protein binding.[10]

The distance distributions are obtained by inverting the gold label–label scattering interference term. According to the Debye formula this interference pattern is given by

$$I_{Au-Au}(s) = \sum_{d=0}^{d_{max}} P(d) |f_{Au}(s)|^2 \frac{\sin(2\pi s d)}{(2\pi s d)} \quad (1)$$

where $f_{Au}(s)$ denotes the scattering factor for a gold nanocrystal and s is the magnitude of the momentum transfer $s = 2\sin(\theta)/\lambda$ with λ as the X-ray wavelength and 2θ being the total scattering angle; d is the distance between the gold nanocrystals and $P(d)$ is the center-of-mass distance distribution. Unfortunately, the scattering from a double-labeled macromolecule does not only contain the label–label interference term (eq 1) but also scattering contributions from the gold label and macromolecule alone as well as a gold–macromolecule interference term. So far, two approaches have been employed to separate the label–label interference term from the other contributions (intragold label, gold label–macromolecule, and intramacromolecule) to the measured scattering profile. In a first approach, each of the single-labeled and the unlabeled samples are measured in addition to the double-labeled macromolecule; from addition and subtraction of the appropriate profiles, the interference term can be determined.[1,2] While this approach has provided unprecedented insights into the conformational ensembles of nucleic acids,[1,8−10] it requires preparation and measurements of several differently labeled

samples for each measured distance distribution. Preparing separate samples of the single- and double-labeled macromolecule can be challenging or impossible if the sample cannot be assembled reliably from individually labeled components. In addition, the differencing procedure requires chemically nonequivalent molecules to adopt the same conformational ensembles and the concentrations or scattering intensities of the various samples need to be carefully matched to achieve the desired separation of scattering terms. A second approach relies on using relatively large (∼50 Å diameter) gold particles and neglecting the DNA and gold−DNA scattering terms.[6,7] A drawback of this approach is the large size of the labels, which might perturb the conformational ensemble and limit its resolution. In addition, neglecting the macromolecule scattering as well as the gold−macromolecule term is problematic or unfeasible for large and strongly scattering macromolecules.

Here we present an alternative strategy to separating the gold label−label interference term and instead determine the intramolecular distance distribution based on the physics of anomalous small-angle X-ray scattering[11] (ASAXS). Our approach is based on recording scattering profiles of the double-labeled sample at different X-ray energies, scanning through a gold absorption edge. Close to an absorption edge, atomic scattering factors change rapidly with X-ray energy and take on the complex form

$$f(E) = f_0 + f'(E) + if''(E) \qquad (2)$$

Away from the edge, the anomalous terms f' and f'' are negligible and the energy independent term f_0 dominates.[11] Because the gold absorption edges are well separated from the elements that make up biological macromolecules (C, H, N, O, P), tuning through a gold absorption edge provides a way to significantly alter the gold scattering compared to the scattering contribution from the macromolecule. The dependence on X-ray energy permits, therefore, separation of the gold−gold scattering terms from macromolecule and gold−macromolecule terms.[12] We demonstrate that our ASAXS approach enables the accurate determination of gold label−label distance distributions from measurements of double-stranded DNA constructs double-labeled with small gold clusters without the need to prepare and measure multiple molecular constructs or to use large gold labels.

We employed 10, 20, and 30 bp double-stranded DNA constructs double-labeled with 7 Å radius thio-glucose functionalized gold particles attached via thiol-chemistry as described previously[1,2,13] (see Supporting Information). ASAXS data were recorded at beamline 12-ID of the Advanced Photon Source[14,15] (Figure 1 and Supporting Information). Prior to measurements of DNA constructs, the X-ray energy was calibrated by inserting a thin (50 μm) gold foil into the X-ray beam and measuring the incident and transmitted X-ray intensity as a function of energy (Figure 2). The data clearly indicate the position of the gold L-III edge at ∼11.9 keV and allow us to reference all measurements to the tabulated absorption data (http://henke.lbl.gov/optical_constants/; see Supporting Information). Similar measurements using the thioglucose passivated gold nanocrystals in solution instead of the gold foil confirm that the absorption edge of the gold clusters is identical within experimental error to bulk gold (Figure 2).

Experimental scattering profiles for double gold-labeled DNA constructs recorded at different energies ranging from 200 eV below to 50 eV above the gold L-III edge exhibit clear and systematic changes with X-ray energy (Figure 3a and Figures
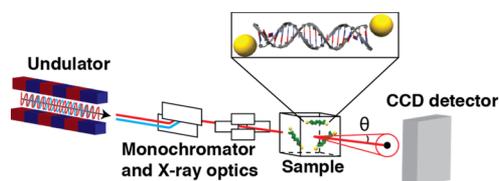

Figure 1. Schematic of the ASAXS measurement on double-labeled DNA molecules. The incident beam is shown along with the monochromator to select particular X-ray energies. Double gold-labeled DNA molecules are placed into the X-ray beam in a sample cell. The direct beam is blocked by a beam stop and scattered photons are detected using a CCD detector. The inset shows a 20 bp DNA molecule with two gold labels; the linkers and gold functionalization are omitted for clarity.
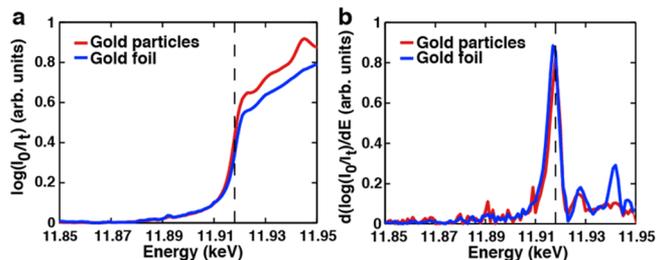

Figure 2. Energy calibration of the gold L-III absorption edge. (a) Normalized logarithm of the incident X-ray intensity divided by the transmitted intensity for a gold foil (blue) and for gold nanoparticles (red) in solution. The data show a sharp increase at the absorption edge. (b) Numerical derivative of the data in (a), revealing the position of the edge as a maximum in the derivative. Data are referenced (see Supporting Information) to the tabulated value of the gold L-III absorption edge at 11.919 keV (shown as dashed vertical lines).

S1−S3; see Supporting Information for details of the experimental and normalization procedures). To extract the gold label-gold label interference pattern from the scattering profiles recorded at different energies, we used a matrix inversion approach[12] based on the following relation (for a more detailed visualization see Figure S4)

$$\begin{pmatrix} I_1(s) \\ I_2(s) \\ \vdots \\ I_{N_E}(s) \end{pmatrix} = \begin{pmatrix} a_1(s) & b_1(s) & c_1(s) \\ a_2(s) & b_2(s) & c_2(s) \\ \vdots & \vdots & \vdots \\ a_{N_E}(s) & b_{N_E}(s) & c_{N_E}(s) \end{pmatrix} \begin{pmatrix} G_{Au-Au}(s) \\ G_{Au-mol}(s) \\ G_{mol-mol}(s) \end{pmatrix}$$

$$I = TG \qquad (3)$$

Here the $I_i(s)$ are the scattering profiles recorded at energies i and the matrix I has dimensions $(N_E \cdot N_s) \times 1$ (where N_E is the number of energies and N_s is the number of s channels). The $(N_E \cdot N_s) \times (3N_s)$ matrix T comprises precomputed matrices $a_i(s)$, $b_i(s)$, and $c_i(s)$ that are all $N_s \times N_s$ diagonal square matrices containing the energy-dependent label−label pair scattering factors, label−molecule pair scattering factors, and intramolecule pair scattering factors, respectively (see Supporting Information). The s-dependence of the $a_i(s)$ is given by the squared scattering factors for the gold labels that are approximated as spheres with a radius of 7 Å; similarly, the s-dependence of the $b_i(s)$ is given by the (nonsquared) scattering factor for the gold labels (Figure S4b,c). The $c_i(s)$ have no explicit dependence on s, that is, they are constant for different values of s (Figure S4d). Importantly, this means that no
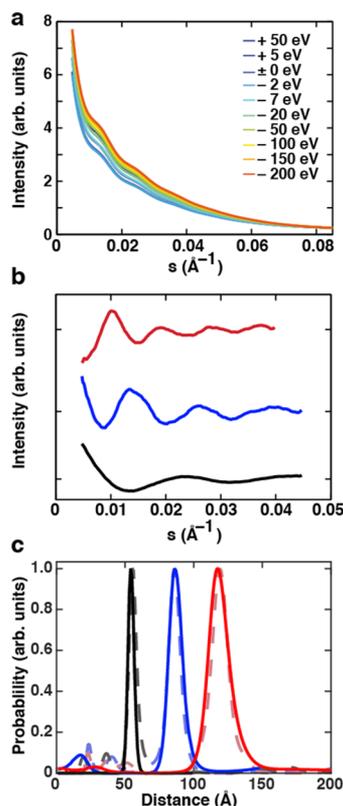
Figure 3. ASAXS scattering data for double-labeled gold samples and distance distributions. (a) Scattering intensities as a function of momentum transfer s for the 20 bp double-labeled DNA construct at 10 different energies. Energies in the figure legend are relative to the gold L-III edge. (b) Gold label−gold label interference patterns for 10 bp (black), 20 bp (blue), and 30 bp (red) DNA constructs obtained as described in the main text. Profiles are vertically offset for clarity. (c) Gold label−gold label distance distributions computed from the data in (b) (solid lines; same color code) by regularized Fourier transformation (see Supporting Information). For comparison, the distance distributions of the same samples obtained in ref 1 are shown as dashed lines.

assumption about the shape of the molecule scattering term is required for our analysis. The energy dependence of the $a_i(s)$, $b_i(s)$, and $c_i(s)$ is given by the energy dependence of the atomic scattering factors for gold and for the atoms in the macromolecule, respectively. We note that this means that we have assumed knowledge of the average element composition of the labeled macromolecule, as is typically the case. However, given the very minor energy dependence of the atomic scattering factors for the elements that make up biological macromolecules in the energy range used (Figure S4d, right panel), the energy dependence of the molecular term could be neglected without loss of resolution.

Finally, G represents the $(3N_s) \times 1$ vector of partial structure factors corresponding to the gold label−gold label (G_{Au-Au}), gold label−molecule (G_{Au-mol}) and molecule−molecule ($G_{mol-mol}$) terms. The vector of partial scattering factors G is obtained by (least-squares) matrix inversion of eq 3 as $G = T^{-1}$ I. The gold label−label structure factor $G_{Au-Au}(s)$ was further processed[12] by subtracting a constant offset such that the mean level of oscillation approaches zero and by truncating at high s values as the signal-to-noise ratio decreases with increasing s.

We note that in principle scattering profiles at 3 different energies would be sufficient to determine G; however, using

data at more (10 in our case) energies overdetermines the matrix equation (eq 3) and improves the signal.[16−18] It is difficult to give general guidelines on how many energies are required, as the results not only depend on the number of energies included, but also on the signal-to-noise of the measurements and on the positioning of the energies relative to the absorption edge. Nonetheless, model calculations on truncated data sets suggest that for our system going from 10 to 6 appropriately chosen energies leads to a loss of signal but still enables computation of meaningful gold−gold partial structure factors (Figure S5). Further reducing the number of energies or incorrectly selecting the position of the energies relative to the edge leads to a further reduction or even complete loss of signal (Figure S5). Our current choice of 10 energies is a compromise of balancing the need to avoid radiation damage (which would be an issue if significantly more exposures would be recorded for the same sample) while achieving a good signal-to-noise ratio for the computed gold−gold partial structure factors.

The gold label−gold label structure factors $G_{Au-Au}(s)$ show characteristic oscillation patterns (Figure 3b) and contain information about the label−label distance distributions. The distance distributions are obtained from the $G_{Au-Au}(s)$ by regularized Fourier transformation using a maximum entropy algorithm (Supporting Information). The distance distribution $P(d)$ for the different DNA lengths all exhibit prominent approximately Gaussian peaks (Figure 3c, solid lines). The center positions of the main peaks increase with increasing DNA length (Figures 3c and 4a and Table S1) and are well fit by a model that takes into account the off-center attachment of the labels (Supporting Information) yielding a rise per base pair of 3.23 ± 0.1 Å (Figure 4a). The mean positions of the peaks
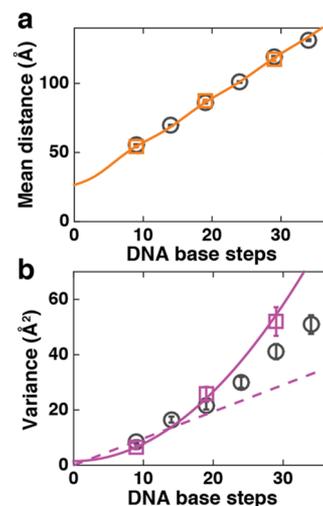

Figure 4. Distance parameters for double-stranded DNA obtained from ASAXS measurements. (a) Mean label−label distances obtained from the main peak of the distance distributions as a function of DNA base steps. Data obtained from ASAXS measurements described in this work (orange squares) and the best fit of a model (see Supporting Information) including the DNA length and label positions (orange line). Error bars are smaller than symbols. (b) Variances of the main peaks of the label−label distance distributions from ASAXS analysis (magenta squares). The data are well described by a quadratic dependence ($\chi^2 = 0.61$; solid line) and incompatible with a linear dependence ($\chi^2 = 13.6$; dashed line). For comparison the previously determined values[1] for the same constructs are shown as black circles.

determined in independent repeat measurements are within $\lesssim 1$ Å (Table S1), highlighting the exquisite precision of the method.

The variance of the main peaks increases rapidly with DNA length and the dependence of the variance of the label−label distances on DNA length is well described ($\chi^2 = 0.61$) by a quadratic dependence with a constant offset (Figure 4b, solid line). The constant offset was fit to be 1.5 Å^2 and prefactor of the quadratic term to be 0.063 Å^2/bp^2; this observation is consistent with a model that assumes cooperative stretching of the helix[1,19] where each base pair step contributes 0.25 Å to the standard deviation. Alternatively, fitting the dependence of the variance on DNA length to a linear dependence that includes a constant term yields fits that would imply large negative values of the variance at zero DNA length, which is unphysical. If we instead fit a linear dependence without an offset, we obtain a poor fit (Figure 4b, dashed line; $\chi^2 = 13.6$). If the complete label−label variance is attributed to the stretch modulus of the DNA alone, the slope of the linear fit implies a value for the stretch modulus of $S_{apparent} \sim 140$ pN, considerably smaller than the value found in single-molecule stretching experiments.[20−22]

The distance distributions obtained using the ASAXS approach (Figure 3c, solid lines) are in excellent agreement with the results of prior measurements employing single-energy SAXS measurements and subtraction of single-labeled and unlabeled DNA contributions for the same DNA constructs and labels[1] (Figure 3c, dashed lines and Figure 4). Therefore, our results provide a clear confirmation, through an independent experimental approach, of the surprising findings obtained previously using SAXS at a single X-ray energy.

In summary, we have used ASAXS measurements at different energies around the gold L-III edge to determine the full label−label distance distribution for gold nanoparticles attached to the ends of 10−30 bp long DNA constructs. Our results demonstrate that by using small (~ 7 Å radius; ~ 80 gold atoms) nanoparticles, the label−label term can be reliably obtained from ASAXS measurements, which was not possible in previous measurements using single atom labels.[23−26] Model calculations suggest that for molecules in the size range investigated here, even smaller gold labels than were used in this study, down to ~ 20 atoms, would be sufficient to obtain interpretable gold−gold interference patterns (Figure S6). Significantly larger molecules will require larger gold labels; to maintain an approximately constant level of relative anomalous scattering signal, the number of gold atoms in the labels should be increased proportionally to the number of atoms in the labeled molecule (Figure S6).

Our ASAXS method has the advantage that only the double-labeled sample needs to be prepared and measured and that it does not rely on intrinsic assumptions about the macro-molecular scattering contributions being negligible; these properties will be particularly advantageous for macromolecular samples where selective labeling at only one position is difficult to achieve and/or that are strongly scattering. These advantages make our method reliable and experimentally attractive, yet it retains the full ability to provide absolute distances with angstrom resolution and precise distance distributions to evaluate the flexibility of macromolecular systems. While this proof-of-concept work uses gold-labeled DNA samples, we fully expect our method to be equally applicable to labeled proteins and to protein−nucleic acid complexes; similarly, other labels, such as silver and platinum nanoparticles, can provide equally attractive and orthogonal labeling options and might permit to measure several distinct distance distributions for one sample.[12] In conclusion, ASAXS provides a powerful new approach to determining intramolecular distance distributions for labeled biological macromolecules that we anticipate to provide new and quantitative insights into the structure, dynamics, and interactions of biological macromolecules.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nano-lett.6b01160.

Supplementary Methods, Table S1, and Supplementary Figures S1−S3. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: Jan.Lipfert@lmu.de. Phone: +49-89-2180-2005.
*E-mail: harbury@stanford.edu. Phone: +1-650-723-6161.

Author Contributions

R.S.M., P.A.B.H., S.D., and J.L. designed the study, R.S.M., S.S., and J.L. performed the experiments, T.Z. and J.L. analyzed the data, and all authors contributed to writing the manuscript.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Mathew-Fenn, R. S.; Das, R.; Harbury, P. A. *Science* **2008**, *322* (5900), 446−9.

(2) Mathew-Fenn, R. S.; Das, R.; Silverman, J. A.; Walker, P. A.; Harbury, P. A. *PLoS One* **2008**, *3* (10), e3229.

(3) Muller, S. M.; Galliardt, H.; Schneider, J.; Barisas, B. G.; Seidel, T. *Front. Plant Sci.* **2013**, *4*, 413.

(4) Yang, Z.; Jimenez-Oses, G.; Lopez, C. J.; Bridges, M. D.; Houk, K. N.; Hubbell, W. L. *J. Am. Chem. Soc.* **2014**, *136* (43), 15356−65.

(5) Krainer, G.; Hartmann, A.; Schlierf, M. *Nano Lett.* **2015**, *15* (9), 5826−9.

(6) Mastroianni, A. J.; Sivak, D. A.; Geissler, P. L.; Alivisatos, A. P. *Biophys. J.* **2009**, *97* (5), 1408−17.

(7) Hura, G. L.; Tsai, C. L.; Claridge, S. A.; Mendillo, M. L.; Smith, J. M.; Williams, G. J.; Mastroianni, A. J.; Alivisatos, A. P.; Putnam, C. D.; Kolodner, R. D.; Tainer, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (43), 17308−13.

(8) Shi, X.; Herschlag, D.; Harbury, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (16), E1444−51.

(9) Shi, X.; Beauchamp, K. A.; Harbury, P. B.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (15), E1473−80.

(10) Shi, X.; Huang, L.; Lilley, D. M.; Harbury, P. B.; Herschlag, D. *Nat. Chem. Biol.* **2016**, *12* (3), 146−152.

(11) Stuhrmann, H. B. *Q. Rev. Biophys.* **1981**, *14* (3), 433−60.

(12) Pinfield, V. J.; Scott, D. J. *PLoS One* **2014**, *9* (4), e95664.

(13) Schaaff, T. G.; Knight, G.; Shafigullin, M. N.; Borkman, R. F.; Whetten, R. L. *J. Phys. Chem. B* **1998**, *102* (52), 10643−10646.

(14) Beno, M. A.; Jennings, G.; Engbretson, M.; Knapp, G. S.; Kurtz, C.; Zabransky, B.; Linton, J.; Seifert, S.; Wiley, C.; Montano, P. A. *Nucl. Instrum. Methods Phys. Res., Sect. A* **2001**, *467−468*, 690−693.

(15) Lipfert, J.; Millett, I. S.; Seifert, S.; Doniach, S. *Rev. Sci. Instrum.* **2006**, *77* (4), 046108.

(16) Munro, R. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1982**, *25* (8), 5037.

(17) Sztucki, M.; Di Cola, E.; Narayanan, T. *J. Appl. Crystallogr.* **2010**, *43* (6), 1479−1487.

(18) Saito, M.; Waseda, Y. *J. Synchrotron Radiat.* **2000**, *7* (3), 152−159.

(19) Schurr, J. M. *J. Phys. Chem. B* **2015**, *119* (21), 6389−400.

(20) Smith, S. B.; Cui, Y.; Bustamante, C. *Science* **1996**, *271* (5250), 795−9.

(21) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (12), 6185−90.

(22) Wenner, J. R.; Williams, M. C.; Rouzina, I.; Bloomfield, V. A. *Biophys. J.* **2002**, *82* (6), 3160−9.

(23) Stuhrmann, H. B.; Notbohm, H. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78* (10), 6216−20.

(24) Miake-Lye, R. C.; Doniach, S.; Hodgson, K. O. *Biophys. J.* **1983**, *41* (3), 287−92.

(25) Das, R.; Mills, T. T.; Kwok, L. W.; Maskel, G. S.; Millett, I. S.; Doniach, S.; Finkelstein, K. D.; Herschlag, D.; Pollack, L. *Phys. Rev. Lett.* **2003**, *90* (18), 188103.

(26) Pabit, S. A.; Finkelstein, K. D.; Pollack, L. *Methods Enzymol.* **2009**, *469*, 391−410.