Beating the reaction limits of biosensor sensitivity with dynamic tracking of single binding events

Derin Sevenler, Jacob Trueb, and M. Selim Ünlü

Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215; Center for Engineering in Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02125; and Department of Mechanical Engineering, Boston University, Boston, MA 02215

Edited by David R. Walt, Brigham and Women’s Hospital–Harvard Medical School, Boston, MA, and accepted by Editorial Board Member John A. Rogers January 19, 2019 (received for review September 13, 2018)

The clinical need for ultrasensitive molecular analysis has motivated the development of several endpoint-assay technologies capable of single-molecule readout. These endpoint assays are now primarily limited by the affinity and specificity of the molecular-recognition agents for the analyte of interest. In contrast, a kinetic assay with single-molecule readout could distinguish between low-abundance, high-affinity (specific analyte) and high-abundance, low-affinity (nonspecific background) binding by measuring the duration of individual binding events at equilibrium. Here, we describe such a kinetic assay, in which individual binding events are detected and monitored during sample incubation. This method uses plasmonic gold nanorods and interferometric reflectance imaging to detect thousands of individual binding events across a multiplex solid-phase sensor with a large area approaching that of leading bead-based endpoint-assay technologies. A dynamic tracking procedure is used to measure the duration of each event. From this, the total rates of binding and debinding as well as the distribution of binding-event durations are determined. We observe a limit of detection of 19 fm for a proof-of-concept synthetic DNA analyte in a 12-plex assay format.

Significance

Low-abundance molecular biomarkers are a promising reservoir of diagnostic information for the early detection and precise treatment of a range of diseases. Yet, most leading assay technologies are currently limited by the affinity and specificity of the molecular-recognition agents they use. This binding “reaction limit” could be beaten by measuring the duration of individual molecular-binding events as they happen in real time. We developed a technology to measure the duration of thousands of individual binding events across a large sensor area. These duration measurements allowed specific analyte binding to be distinguished from nonspecific background binding. The sensor is large enough to accommodate 12 ultrasensitive tests at once.

Author contributions: D.S. and M.S.U. designed research; D.S. performed research; D.S. and J.T. contributed new reagents/analytic tools; D.S. and M.S.U. analyzed data; and D.S., J.T., and M.S.U. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.R.W. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

1To whom correspondence may be addressed. Email: dssevenler@mgh.harvard.edu or selim@bu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1815329116/-/DCSupplemental.
However, single-molecule kinetic measurements are technically demanding: without amplification reactions, specific binding events are more difficult to discern against a background of nonspecific interactions. Indeed, an exquisitely sensitive transduction mechanism is required to directly detect single binding events at all. A range of scientific apparatuses have been developed to investigate single-molecule binding kinetics. However, none of these techniques are useful for ultrasensitive clinical assays because the sensors are too small (14). To investigate nanoscale phenomena, these devices are themselves nanoscale: their active sensors are the size of single nanoparticles or nanowires (15–17), or else they require high-magnification and high-numerical aperture optics with a small field of view (0.001–0.01 mm²) (18–20). This is problematic because small sensors only have space for a small number of capture probes. Maximizing the number of probes is vital for ultrasensitivity: at low concentrations, the amount of captured analyte at equilibrium is proportional to the number of probe molecules. Single-molecule assay technologies therefore use large sensor areas packed with capture probes. For example, the Quanterix Simoa technology interrogates a range of scientific apparatuses have been developed to investigate single-molecule binding kinetics. However, none of these techniques are useful for ultrasensitive clinical assays because the sensors are too small (14). To investigate nanoscale phenomena, these devices are themselves nanoscale: their active sensors are the size of single nanoparticles or nanowires (15–17), or else they require high-magnification and high-numerical aperture optics with a small field of view (0.001–0.01 mm²) (18–20). This is problematic because small sensors only have space for a small number of capture probes. Maximizing the number of probes is vital for ultrasensitivity: at low concentrations, the amount of captured analyte at equilibrium is proportional to the number of probe molecules. Single-molecule assay technologies therefore use large sensor areas packed with capture probes. For example, the Quanterix Simoa technology interrogates ~25,000 beads, each 2.7 μm in diameter, corresponding to a total sensor area of 0.57 mm² (12).

Here, we describe a kinetic assay that measures the duration of individual binding events over time on a large sensor surface with a low-magnification objective, while retaining the advantages of kinetic analysis such as discrimination between specific and nonspecific events based on duration (i.e., affinity). In this study, we used a 20x, 0.45 NA objective and a 1.1-inch format camera, which yielded a sensor area of 0.38 mm², comparable with that of ultrasensitive endpoint methods. (This area could be further increased several-fold with different optical instrumentation and stage scanning.)

**Results**

**Detection of Individual Binding Events Across a Large Field of View.** We recently described the development of a “digital microarray” assay technology, which rapidly enumerates individual captured molecules across hundreds of microarray spots (21). This technology uses probe-conjugated gold nanorods as molecular labels and an interferometric reflectance imaging sensor (IRIS) to rapidly detect individual nanorods with a low-magnification (10–20x) objective. The large field of view enabled a similar throughput to commercial fluorescence readers while enhancing the limit of detection (LOD) and dynamic range by a factor of ~10,000. The assay reaction is based on the barcode assay developed by others (22) and is compatible with a range of different analytes.

We adapted the IRIS digital microarray platform for dynamic measurements by designing a perfusion chamber that consists of an IRIS chip, a patterned silicone gasket, and an antireflection-coated coverglass window (Fig. 1A). Two holes for the chamber inlet and outlet are drilled in the IRIS chips by wafer-scale laser micromachining. The assembly is held by a custom clamp fixture that makes fluidic connections to the inlet and outlet on the bottom of the chip (SI Appendix, Fig. S1).

To demonstrate dynamic detection of single molecules, a synthetic ssDNA oligonucleotide was used as a proof-of-concept analyte. The analyte was preincubated with cDNA-conjugated gold nanorods nominally 25 nm × 70 nm for 90 min. The concentration of nanorod labels was kept constant for all experiments at 14 pM, while the concentration of the analyte varied from 10 pM to 10 fM. After preincubation, the mixture was perfused over IRIS chips with DNA microarrays of complementary and noncomplementary probes and nanorod-labeled oligos hybridized to the complementary spots (Fig. 1B).

Images were acquired every 30 s with the IRIS instrument during perfusion. Nanorods on the IRIS chip were visible as faint

![Fig. 1. Dynamic measurements of single binding events across a large microarray. (A) Rendered image of IRIS chip perfusion chamber for dynamic measurements of molecular interactions. A DNA or antibody microarray is printed on the IRIS chip. Then, the chamber is formed by layering a patterned adhesive gasket and antireflection-coated coverglass viewing window. The IRIS chip has two through-holes for sample perfusion. The entire disposable costs about $5 USD. (B) Nucleic acid assay with IRIS. DNA-conjugated gold nanorods are preincubated with the sample solution and hybridized with complementary nucleic acids. The mixture is flowed over the chip. Complementary nucleic acid strands tether nanorods to the cDNA microarray spot. (C) Schematic of dynamic detection of single nanorods with IRIS. Images are simulated. Nanorods on the chip surface are observed as diffraction-limited spots and automatically detected using purpose-built software. (D) Plots of total nanoparticle binding to six complementary (red) and six noncomplementary (green) DNA spots over time, as measured with dynamic tracking, for one representative experiment where the analyte concentration is 316 fM.](Image)
diffraction-limited blobs in the images, which were detected in each frame independently using custom software (Fig. 1C). The binding rates were then measured determined using a custom dynamic tracking algorithm described in the following section (Fig. 1D and Movie S1).

IRIS detects individual nanorods based on light scattering. Since water has a higher refractive index than air, the polarizability and scattering cross section of the nanorods were reduced compared with dry chips. Additionally, the image suffered from spherical aberrations caused by the air-coverglass interface. Although nanorods were detectable with a 10×, 0.3 NA objective, visibility was much improved with a 20×, 0.45 NA coverglass corrected objective. The resulting field of view of 0.38 mm² (725 μm × 530 μm) could comfortably accommodate 12 microarray spots, each ∼80 μm in diameter.

Dynamic Tracking of Binding Events over Time. Under sufficiently high flow rates, the initial rate of binding of analyte is proportional to the bulk analyte concentration. One may estimate the analyte concentration by plotting the number of bound nanorods over time and measuring the initial slope (“naive counting”; Fig. 2A). However, this approach has several problems. The first is the fundamental issue related to finite probe affinity mentioned earlier. Ultralow analyte concentrations will reach equilibrium with very few (or even fewer than one) bound analyte molecules. In those cases, the initial slope will not be measurable even with perfect error-free readout. The second issue is that some unbound nanorods are visible in each frame as they transiently diffuse through the detection volume. These transient particles result in false positives that increase the overall noise floor of the sensor.

To address these issues, we developed a postprocessing algorithm that uses the spatial positions of particles to track them individually over the course of the experiment (Fig. 2B). First, particles are detected in each video frame independently of other frames, and their positions in the image are recorded. Second, these positions are compared with those of particles in the next frame of the video (Fig. 2C). Particles in the same position in both frames are “matched,” indicating that they are in fact the same particle. Particle matching includes a clustering algorithm that is robust to small translations between frames. For a video with N time points, this results in N − 1 lists of matches. Third, these lists are compiled into a single master catalog, which tracks the contiguous series of frames in which each particle was observed (Fig. 2D). This is essentially a table that lists when each particle bound, where it bound, and when (if ever) it debound from the surface. Finally, this catalog is filtered to reduce false positives and false negatives. False negatives occur when a particle is mistakenly not detected in a single frame but was detected in the same position in previous and subsequent frames. This erroneously results in two entries in the catalog. These gaps are repaired by identifying whether the binding of each particle corresponds to the exact same place as the debinding of another particle two frames prior and then merging the two catalog entries (e.g., particles 3 and 7; Fig. 2D). False positives are caused by particles visible in just one frame and are simply removed (e.g., particle 4; Fig. 2D).

This catalog can then be used to plot the cumulative number of new binding events over time. For low-analyte concentrations, most binding sites remain empty at equilibrium, and the rate of new binding events will be constant and proportional to the bulk concentration. At an analyte concentration of 316 fM, for example, the sensor reached equilibrium with about 80 bound nanorods after 1 h (naive counting; Fig. 3A). However, the cumulative number of new binding events continued to increase linearly even after equilibrium was reached (“total binding”; Fig. 3B). Note that the total number of binding and debinding events

![Fig. 2.](image-url) (A) Dynamic tracking improves sensitivity when the debinding rate is nonzero. Assuming first-order reaction kinetics, equilibrium is reached when the rate of new analyte binding is balanced by the rate of debinding from the surface. Naïvely counting the instantaneous number of bound analyte provides no additional information once equilibrium has been reached. Dynamic tracking distinguishes the binding of new particles from the debinding of old ones. The cumulative number of binding events continues to increase, even at equilibrium. (B–D) Diagram of the multistep dynamic tracking algorithm, described in the text.
are both cumulative measurements and are therefore monotonically increasing over time.

We compared measured binding rates with the predicted rates of transport of analyte-bound nanorods to the spots and found the measured binding rates were well below the theoretical upper limit predicted by mass transport (SI Appendix). We also experimentally characterized the maximum expected binding rates with a series of positive control experiments, in which the nanorods irreversibly bound directly to the microarray spots without any intermediate analyte (SI Appendix). In these control experiments, the rate of debinding was negligible, so the instantaneous and cumulative binding rates were nearly identical and linearly proportional to nanorod concentration, as expected (SI Appendix, Figs. S4–S5). We also evaluated whether replicate spots downstream of one another would experience reduced binding due to sample depletion. We observed no difference in binding rates between upstream and downstream spots at any concentration, supporting the notion that the system was not transport limited (SI Appendix, Fig. S6).

Measurements of Nanorod Dwell Times. The duration or “dwell time” of each binding event may be measured using dynamic tracking. Note that the dwell time can only be measured for particles that debind before the end of the experiment. Taken together, these dwell times allow the off rate to be determined (Fig. 3C). We found that our experimental results were best explained by a biexponential fit of the form $N(t) = A_1 e^{-k_1 t} + A_2 e^{-k_2 t}$. A histogram of dwell times across all complementary spots was generated for each experiment and fitted independently (SI Appendix, Fig. S3). Ignoring experiments that were too brief or contained too few binding events to have meaningful statistics, the fitting parameters consistent across experiments regardless of analyte concentration (SI Appendix, Fig. S3). The average values were $k_1 = 0.53 \text{ min}^{-1}$ and $k_2 = 0.082 \text{ min}^{-1}$ and $A_1/A_2 \approx 25$, corresponding with primary and secondary dissociation time constants $\tau_1 = 1.9 \text{ min}$ and $\tau_2 = 12 \text{ min}$. At first, we hypothesized this heterogeneity consisted of nanorods tethered by either one versus two or more analyte molecules. However, the relative weights between the two terms $A_1/A_2$ was similar across a large range of concentrations, and the ratio did not tend to decrease with lower analyte concentrations. Since the total nanorod concentration was kept constant at 14 pM, the relative number of nanorods with two bound analytes versus one bound analyte would have decreased with decreasing analyte concentration.

Instead, we hypothesize that the biexponential distribution in dwell times was caused by the asymmetry of the nanorods themselves as they bind to the surface. The binding energy is likely greater if the rod is tethered to the surface by one end, rather than by the middle. First, there is electrostatic repulsion between the DNA-functionalized nanorods and the DNA-coated chip surface. A side-tethered nanorod is constrained in a manner that brings the centroid closer to the chip and brings a larger surface area adjacent to the chip. Second, the end-tethered nanorod has a larger number of conformational degrees of freedom (DOF) (three rotational DOF) than a sideside-tethered one (one rotational DOF), resulting in a lower entropic penalty to binding. Since the rods are functionalized uniformly across their surfaces, nanorods are more likely to capture analyte to their sides rather than ends during pre-incubation and therefore be side-tethered. This is supported by the observation that the faster dissociation rate was also more prevalent across all concentrations.

It may be important to note that we cannot differentiate between dissociation of the analyte from the surface probe and from the nanorod label. In the model system used here, both surface–analyte and analyte–label duplexes are 25 bp long and have similar GC content. Therefore, their affinities should be similar, and they should be responsible for nanorod dissociation at roughly similar rates.

Detection Below the Critical Concentration. A standard curve was measured by performing identical experiments with a range of analyte concentrations between 10 fM and 10 pM. Nanorod binding to three replicate complementary spots were independently analyzed using both dynamic tracking and naïve counting (Fig. 4). A modified first-order reaction model was used to capture both the linear behavior at low-analyte concentrations and the saturation behavior at high concentrations, where most nanorods had captured at least one analyte (SI Appendix). The LOD was calculated as three SDs above the mean signal from the blank sample. As expected, the LOD for dynamic tracking (LOD = 19 fM) far surpassed that of naïve counting (LOD = 687 fM), by about 36-fold.

Notably, dynamic tracking had an LOD 3.6 times lower than the critical concentration of this assay ($e^*$; Fig. 4A). Equivalently, dynamic tracking was able to detect the presence of the analyte...
even when the average duration of binding events was shorter than the time between them. At the critical concentration, equilibrium is reached with one bound analyte molecule on average. This dissociation rate is simply the weighted average of the two dissociation constants found earlier: 

$$k_{off} = \frac{(A_1k_{1} + A_2k_{2})}{A_1A_2}$$

0.52 min$^{-1} = 31$ particles per hour. The critical concentration was determined by taking intercept of the dynamic tracking regression line with this binding rate: $c^* = 55$ fM.

We observed a saturation in binding rate at concentrations above 2–3 pM (Fig. 4A). We hypothesized this could be the result of using a lower concentration of nanorods than expected. We performed spectrophotometry to estimate the concentration of the conjugated nanorods as well as the nanorod stock solution from the manufacturer. We compared the absorbances to that predicted by numerical simulations and estimated that the nanorod concentration was indeed several times lower than we expected (SI Appendix, Fig. S7).

**Discussion**

A range of endpoint-assay technologies have been developed that have single-molecule readout. For these assays, the limiting factor is the affinity and specificity of the molecular-recognition agents rather than readout sensitivity. The number of capture probes (for example, the number of functionalized beads) can almost always be increased until sensitivity is limited by non-specific binding rather than insufficient numbers of probes (23), but further improvement must come through careful optimization of washes and reactions. Protocol optimization are particularly challenging for multiplexed test development, since the optimal wash conditions (duration, ion content, surfactants, pH, and so on) are often different for different probe–analyte complexes.

In response to these limitations, we have introduced a kinetic-assay technology that measures the duration of individual binding events across a large sensor area. In this work, we distinguished specific binding events from the background without even a single wash step, which could be used to further improve specificity. Notably, kinetic analysis alleviates the need for a globally optimal wash protocol and therefore makes multiplexed tests straightforward.

“Solid-phase” surface sensors are sometimes criticized for having poor mass-transport kinetics, compared with bead-based assays. We alleviated this effect by using a high flow rate, which makes the depletion layer very thin (SI Appendix). For longer experiments or smaller sample volumes, peristaltic pumping and recirculation may be used.

The unexpected biexponential histogram of dwell times suggested two different conformations of immobilized nanorods, each with different binding free energy: end-tethered and side-tethered. This variability complicates the probe–analyte affinity measurement and could be problematic for heterogeneous samples. This could be further tested and perhaps mitigated by preferentially functionalizing just the ends of nanorods (24, 25).

Because these proof-of-concept experiments were conducted with synthetic analytes and pure buffer solutions, direct comparisons with more mature assay technologies cannot yet be made. However, future applications of this technique are envisioned in response to current clinical needs for multiplexed quantification of microRNAs for cancer diagnostics, mRNAs for phenotypic identification of drug susceptibility, and circulating protein biomarkers of viral infection.

**Materials and Methods**

Additional materials and methods are available in SI Appendix, Supplemental Materials and Methods.

**Perfusion Chamber Assembly.** No. 1 coverslips, 25.4 mm × 12.7 mm, with a broadband antireflection coating on one side were purchased from Abra BioTechologies. Custom patterned silicone gaskets were purchased from Grace BioLabs. Silicone gaskets were 25.4 mm × 12.7 mm, 0.15-mm thickness, with pressure-sensitive adhesive on one side. In preparation, gaskets were adhered to the noncoated side of the coverslip and stored with protective tape in place.
The perfusion chamber was assembled by aligning the gasket-window assembly to the IRIS chip, loading it into the clamp fixture, removing the protective tape, and engaging the clamp to form seals between the chip and the gasket as well as with the sample inlet and outlet. The volume of the chamber was ~8 μL.

IRIS Digital Microarray Instrument for Dynamic Detection. The operating principle of IRIS is thoroughly described elsewhere (26). Briefly, the IRIS instrument consists of a reflectance microscope with a high-power LED for illumination (M660L4 LED with FB650-10 bandpass filter; Thorlabs) and a monochrome machine vision camera (Grasshopper GS3-13U2356M-C; Point Gray Research). The digital microarray implementation of IRIS is optimized for rapidly detecting individual gold nanorods based on their anisotropic light-scattering properties. The design, optimization, and implementation of the optical system has been described in detail elsewhere (21). For dry IRIS chips, this system can detect single gold nanorods with a 10x, 0.3 NA objective. For dynamic experiments the system was entirely the same except that a 20x, 0.45 NA coverslip-corrected air immersion objective (Nikon CFI S Plan Fluor ELWD 20x) was used. The higher light-collection efficiency compensated for the decreased intensity of nanorod light scattering due to immersion of the rods in water, and the collar allowed correction of spherical aberrations from the coverslip-air interface.

Assay Protocol. The assay protocol was identical for all experiments, except that the concentration of the analyte ssDNA was varied. First, the DNA–nanorod conjugates and synthetic DNA oligos were premixed in a “hybridization buffer” consisting of 10 mM phosphate (pH 7.4), 600 mM Na+, 0.1% Tween-20, and 1 mM EDTA; 100 μL of nanorods stored at 140 °C were mixed with 900 μL of hybridization buffer containing the analyte ssDNA. The final nanorod concentration was 14 pM for all experiments. The mixture was vortexed briefly and sonicated for 10 s before storing at room temperature in a microcentrifuge tube. After 90 min, the sample was aspirated with a 1-mL needle-tipped syringe. The needle tip was removed, and the syringe was connected to a Luer fitting on the end of the inlet tube. The outlet waste tube was left in conical vial. The syringe was mounted in a syringe pump, and the sample was dispensed at 10 μL/min for up to 90 min. The instrument was refocused as soon as the liquid sample filled the chamber. Video acquisition began 1–3 min after the sample first contacted the chip surface.

4. Alsdurf H, Hill PC, Matteelli A, Getahun H, Menzies D (2016) The cascade of care in the gasket as well as with the sample inlet and outlet. The volume of the chamber was ~8 μL.

**Image Acquisition.** Image acquisition was automated using the Micromanager (27) microscope control software with custom scripts. Scripts have been made available online at https://www.github.com/derinsevenler/IRIS-API. Time points were taken every 30 s at each time point, and a z-stack of nine images was acquired with a step size of 2 μm (i.e., a span of 16 μm). At each z-position, four images were acquired and averaged pixelwise before saving to reduce shot noise.

**Image Processing and Particle Detection.** The video data from each experiment consisted of an image hyperstack of 180 (x) × 9 (y) × 12.4 MP (z) pixels region of the video was cropped around each microarray spot. Nanorods in each region and timepoint were detected independently. The particle detection method described here is a refinement of methods described earlier (21, 28), and has three steps: preprocessing, key point detection, and key point filtering. First, a sparse pseudomedian filter is applied to each frame of the z-stack (made available online at https://www.github.com/derinsevenler/IRIS-API: filter:fast:filters:start) to estimate the image background. True median filtering is effective for removing punctate features but computationally expensive for larger kernels. We found the sparse pseudomedian algorithm preferable due to its speed. Next, the normalized intensity image was calculated by pixelwise division of the original frame from the background. Finally, the normalized intensity range (NIR) image was measured by projecting the maximum difference (i.e., maximum − minimum) at each pixel of the normalized intensity stack. Although not every nanorod is visible in every normalized intensity image in the stack, each particle is clearly visible in the resulting NIR image.

Key points in the NIR images were detected by applying a global threshold to binarize the image. Blobs in the binary image (i.e., regions brighter than the threshold) were enumerated and then filtered based on size and shape. Specifically, a minimum area, maximum area, and minimum area-to-perimeter ratio were specified. The detection threshold and key point-filtering parameters were manually selected and then kept constant for all experiments.

**ACKNOWLEDGMENTS.** The authors are grateful to Fulya Ezik Kanik, Joseph Choong and Abdul Bhuiya for feedback on the image acquisition and analysis software and perfusion chamber design; to Celalettin Yurdakul for help designing the chip photolithography pattern; to Matthew Geib for help preparing IRIS chips and spotting; and to Ahmad Khaliif for fruitful discussions about these results.  