Using protein-DNA chimeras to detect and count small numbers of molecules

Ian Burbulis[1], Kumiko Yamaguchi[1], Andrew Gordon[1], Robert Carlson[2] & Roger Brent[1]

We describe general methods to detect and quantify small numbers of specific molecules. We redirected self-splicing protein inteins to create 'tadpoles', chimeric molecules comprised of a protein head covalently coupled to an oligonucleotide tail. We made different classes of tadpoles that bind specific targets, including *Bacillus anthracis* protective antigen and the enzyme cofactor biotin. We measured the amount of bound target by quantifying DNA tails by T7 RNA polymerase runoff transcription and real-time polymerase chain reaction (PCR) evaluated by rigorous statistical methods. These assays had a dynamic range of detection of more than 11 orders of magnitude and distinguished numbers of molecules that differed by as little as 10%. At their low limit, these assays were used to detect as few as 6,400 protective antigen molecules, 600 biotin molecules and 150 biotinylated protein molecules. In crudely fractionated human serum, the assays were used to detect as few as 32,000 protective antigen molecules. Tadpoles thus enable sensitive detection and precise quantification of molecules other than DNA and RNA.

Numerous questions in biology depend on detection and quantification of very small numbers of specific molecules mixed among large numbers of other molecules. For example, understanding how genetically identical cells behave differently will require quantification of differences in the amounts and modification states of small amounts of regulatory proteins. Similarly, early detection of many diseases from clinical samples will be facilitated by the ability to detect and quantify small numbers of signature molecules. The mainstay methods for quantification of biological molecules other than nucleic acids remain enzyme-linked immunoabsorption assays[1] (ELISAs) and their predecessors, radioimmune assays[2] (RIAs), developed in 1971 and 1959, respectively. One limitation to the sensitivity and accuracy of these assays has been the means used to quantify the target bound by the antibody.

One approach to target quantification that promises great sensitivity has been to bring specific DNA sequences to the target and then detect and count the bound sequences by nucleic acid hybridization or amplification approaches. The idea of detecting target binding by DNA amplification dates from 1992 (ref. 3), as does the use of DNA tags to detect small numbers of target bound antibodies[4]. These experiments demonstrated the application of the extraordinary sensitivity of PCR detection to non-nucleic acid molecules. Since these initial developments, other investigators have described detection of DNA–linked antibodies bound to targets by PCR[5–8], T7 runoff transcription[9] and hybridization[10]. For a number of reasons, including in many cases the molecular heterogeneity of the antibodies and the lack of precise chemical control of the site(s) of DNA attachment, assays based on DNA–linked antibodies have not gained wide use. We envisioned that the ability to effect well-controlled synthesis of defined protein-DNA conjugates in large quantities might enable wider development of target-binding and amplification measurement methods.

Here, we devised the means to covalently link chemically homogeneous affinity proteins and DNA molecules by controlled site-specific chemistry. The chemistry is catalyzed by inteins, protein domains that excise from polypeptides and join the remaining amino and carboxyl moieties to form a mature protein[11]. After excision, the amino terminal moiety acquires a single thioester residue at its C terminus. This thioester can couple with synthetic peptides in *trans*[12]. It has been used to couple single-chain antibodies to fluorescent labels[13] and has recently been shown to join proteins to cysteine-conjugated polyamide nucleic acids[14].

We used a collection of modified inteins to generate thioester-tagged recombinant antibodies as well as other affinity proteins and coupled them to synthetic oligonucleotides. The resulting chimeric tadpole molecules were composed of a protein head and a DNA tail. We adapted absorption and crosslinking, PCR, T7 amplification and statistical methods to use tadpoles to detect and count small numbers of different molecules, such as biotin and native proteins, including the 'protective antigen' (PA) subunit of the *B. anthracis* toxin. These methods enabled detection and quantification of non-nucleic acid molecules with great precision, great dynamic range and, in some cases, sensitivity near the limits of PCR.

RESULTS

Tadpole synthesis

To explore the diversity of protein structures suitable for the tadpole heads, we directed the synthesis of three classes of affinity protein moieties in *Escherichia coli*: streptavidin, a protein that specifically binds the vitamin biotin[15], single-chain antibodies

[1]The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, California 94704, USA. [2]Present address: University of Washington, Department of Electrical Engineering, Box 352500, Seattle, Washington 98195, USA. Correspondence should be addressed to I.B. (burbulis@molsci.org).

PUBLISHED ONLINE 21 DECEMBER 2004; DOI:10.1038/NMETH729

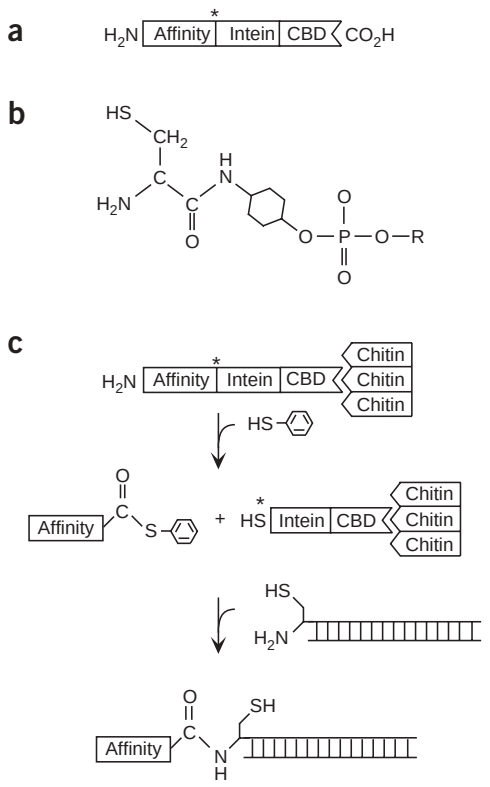

Figure 1 | Design and synthesis of chimeric detector molecules. (**a**) An illustration of the typical three-part fusion proteins containing, from N to C terminus, an affinity domain with 6-His sequence, an intein and the CBD, except for the streptavidin fusion, which lacked the 6-His tag. An asterisk denotes the active-site cysteine. (**b**) Cysteine-like residue incorporated at the 5′ terminus of the forward strand of the DNA tail (represented by 'R') that serves to covalently couple with the reactive thioester on the affinity protein. (**c**) Thioester-dependent ligation of protein to DNA. Chitin-purified recombinant intein fusion proteins are reacted with 0.2% thiophenol, 1 mM TCEP and modified DNA *in situ*. Thiophenol helps promote the hydrolysis of the peptide backbone to yield a reactive thioester at the C terminus of the affinity protein. Next, the thiol moiety on the synthetic DNA tail displaces the thiophenol[12], which then undergoes an N-S acyl shift[11] resulting in an amide linkage between protein and DNA.

containing disulfide linkages[16] and thioredoxin (TrxA) peptide aptamers[17] (**Supplementary Table 1** online). We expressed these affinity moieties as fusion proteins that comprised, from N to C terminus, an affinity protein, an intein and the 52-residue chitin binding domain (CBD) from *Bacillus circulans* WL-12 Chitinase A1[18] (**Fig. 1a**). For the DNA components, we synthesized double-stranded (ds) oligonucleotides containing a cysteine-like residue[19] at one 5′ end (**Fig. 1b** and **Supplementary Table 2** online). We purified the three-part fusion proteins on chitin columns[20] and activated intein excision and DNA coupling by filling the columns with buffer containing the reducing agent, thiophenol and cysteine-modified DNA (**Fig. 1c**). In the subsequent reaction, the thiophenol allows the intein to hydrolyze the peptide backbone, excising the affinity protein from the fusion and placing a reactive thioester at its C terminus. The thiol moiety on the cysteine in the synthetic DNA tail displaces the thiophenol[12] and undergoes an N-S acyl shift[11], resulting in an amide bond between the protein and DNA. We allowed the cleavage-coupling reactions to run overnight and purified the protein-DNA chimeras from the column eluate. The resulting tadpole molecules have an affinity protein head covalently linked by an amide to a DNA tail (**Fig. 1c**).

Tadpole characterization

The first tadpoles we created contained the 'core' domain of streptavidin[21] linked to a 149-mer DNA tail. We purified them by anion-exchange and hydrophobic-interaction chromatography to apparent homogeneity (effectively >95% purity) as judged by the presence of a single SYBR gold (Molecular Probes) stained band on native acrylamide gels. A number of observations were consistent with the idea that these molecules comprised a single affinity protein linked to a single DNA moiety (**Fig. 2**). On native gels stained for DNA, digestion with protease K caused a shift in the mobility from that expected for a single protein-DNA chimera to that of the free DNA tail. On size exclusion columns, the tadpoles' apparent molecular weight was as expected for a molecule containing a single protein and a single DNA moiety. As assayed from the change in mobility on native gels in the presence of biotin (and by ELISAs, data not shown), both the anti-biotin tadpole and the free protein head recognized biotin, which indicates the attachment of DNA did not interfere with the binding activity of the protein domain. Parental fusion proteins not treated with thiophenol or Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) did not form detectable amounts of this species, consistent with the idea that linkage occurs at the C-terminal thioester resulting from the intein reaction (**Fig. 2a**).

Biotin quantification

We used these tadpoles to detect and count small numbers of biotin molecules *in vitro*. We determined the biotin content of biotinylated bovine serum albumin (BSA; Pierce) and carbonic anhydrase preparations (data not shown) using 2-(4′-hydroxyazobenzene)-benzoic acid (HABA) in spectrophotometric assays[22] as described in Methods. We diluted biotinylated BSA into mixtures that contained an excess of BSA coupled to two other dyes into wells on a standard ELISA dish. We blocked the wells with casein and calf thymus DNA and added to them anti-biotin tadpoles. We washed the wells and quantified tadpoles that remained bound by loading the wells with T7 polymerase and ribonucleotide triphosphates and allowing run-off transcripts to accumulate[23]. We quantified the amount of RNA by measuring the fluorescence intensity of a stained band on denaturing polyacrylamide gels. The signals for biotinylated BSA were linear over six orders of magnitude, whereas controls containing only NDA- and Texas Red–coupled proteins gave no signal above background (**Fig. 3a**). In parallel experiments, we quantified biotin immobilized on carbonic anhydrase with equivalent sensitivity as that for biotinylated BSA (data not shown), indicating that the measurement of biotin was independent of the protein substrate. These results show that a tadpole can detect and quantify small numbers of an organic molecule against a background of very large numbers of molecules of similar mass.

To determine the smallest number of biotin molecules detectable using these tadpoles, we adapted real-time PCR[24]. We serially diluted biotinylated BSA into BSA, covalently coupled proteins in the dilution mixtures to epoxy-modified magnetic beads[25,26] (Dynal Biotech) and blocked the beads as above. We incubated the beads with tadpoles, washed the beads to remove non-bound

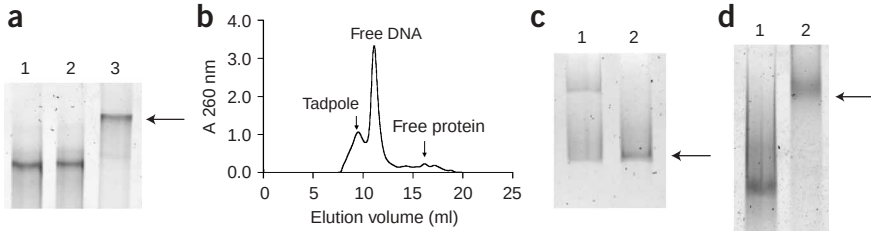

Figure 2 | Characterization of anti-biotin tadpole. (**a**) To monitor the coupling of DNA to the protein head we resolved sample before and after conjugation on a native 12% Tris-borate-EDTA polyacrylamide gel. Lane 1 is the DNA tail alone, lane 2 is DNA tail added to the streptavidin-intein fusion protein without thiophenol activation and lane 3 is the DNA tail added to the intein fusion protein in the presence of thiophenol and TCEP. (**b**) Superdex-200 size exclusion chromatography of the streptavidin tadpole with absorbance monitored at 260 nm. The arrows indicate the conjugated tadpole, free DNA and free protein in the elution. The mobility of the streptavidin tadpole is consistent with one DNA tail attached to a single protein. (**c**) Confirmation that the streptavidin tadpole is composed of protein and DNA. Samples were digested with proteinase K and resolved on a gel as above. Lane 1 is mock-treated tadpole and lane 2 is tadpole digested with proteinase K. The arrow indicates the increase in electrophoretic mobility from a low-mobility tadpole conjugate to a size consistent with a free DNA tail. (**d**) Confirmation that the streptavidin tadpole bind biotinylated BSA *in vitro*. Lane 1 is the tadpole alone and lane 2 is the tadpole in the presence of biotinylated BSA. The arrow indicates the reduced electrophoretic mobility in the presence of biotinylated BSA as compared to tadpole alone, and this is consistent with the formation of a complex between the tadpole and the biotinylated BSA.

species and counted the number of target-tadpole complexes by including the beads in real-time PCRs (**Supplementary Table 3** online). We quantified bound tadpoles by comparing the reaction cycle at which the accumulated fluorescent signal surpassed an arbitrary threshold to a calibration curve generated from known numbers of tadpoles (**Fig. 3b**).

We found that standard linear regression methods typically used to determine numbers of template molecules in real-time PCR experiments were inadequate to determine the actual limit of detection for these tadpole-based assays. The small amount of tadpole binding that occurred on the beads in the absence of target ligand resulted in a non-linear signal output at the transition between background binding to the beads and specific binding to the ligand. At these concentrations, the use of linear regression resulted in less precise measurements and frequently exaggerated the sensitivity of the measurement.

To better quantify the data, we constructed a 'confidence belt' for the measurements according to published methods[27,28]. We defined the limits (area) of our confidence belt as two standard deviations of the PCR cycle measurement (**Fig. 4**). For this work and for other applications of this method, we assumed that when testing an unknown sample, researchers would perform measurements in triplicate. Given the boundary limits of our confidence belt, we evaluated the estimated number of target molecules present in a sample that will result in a mean number of PCR cycles falling within the specified limits in 95 out of 100 measurements. As long as measurement results are distributed normally (**Supplementary Fig. 1** online), this approach is valid across the entire dynamic range of the assay, including the lowest values, as well as the linear range.

Using this technique to evaluate the PCR data, we detected as few as 600 biotin molecules (or 150 biotinylated BSA molecules) with >95% confidence (**Fig. 4**). In our hands, the detection limit for this tadpole-based assay was 10^9-fold lower than for a conventional ELISA (**Supplementary Fig. 2** online); by combining PCR and T7 polymerase data, the dynamic range of these measurements spanned 11 orders of magnitude. In the linear range, the smallest difference between the numbers of molecules these assays could distinguish was consistently below 27%. The resolution of these measurements is relatively high; for example, these methods can distinguish between 50,000 and 64,000 molecules with >97% confidence.

Synthesis of tadpoles to quantify native proteins

We then made tadpoles to quantify native proteins. First, we made tadpoles that contained TrxA peptide aptamers that bound human E2F1[29] and CDK2[17]. In bead-based real-time PCR assays performed as above, these tadpoles allowed quantification of

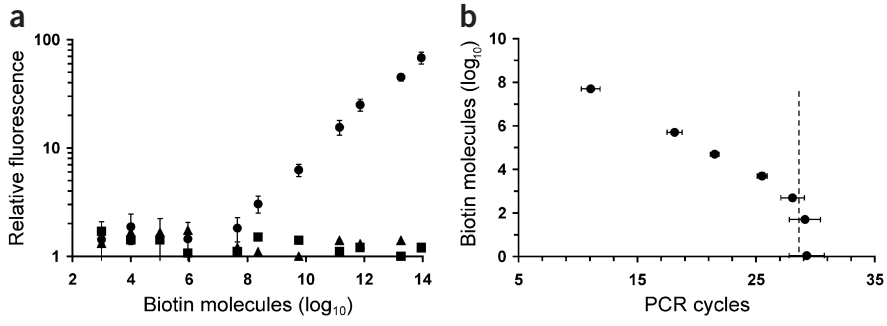

Figure 3 | Quantification of biotinylated BSA diluted in an excess of nonspecific organic molecules. (**a**) The fluorescence signal of SYBR green-stained RNA, is graphed as a function biotin quantity. The circles, triangles and squares represent biotinylated, NDA-labeled and Texas Red-labeled BSA, respectively. Error bars represent 2 s.d. The measurements were performed in triplicate. (**b**) Quantification of biotin using real-time PCR. The *x*-axis shows the PCR cycle at which the change in fluorescence surpassed a designated threshold. The *y*-axis shows the number of molecules of target-bound tadpole tails determined by reference to a calibration curve using known numbers of free tadpoles. Circles represent the mean of five replicate measurements and the error bars represent 2 s.d. The dashed line indicates the lower limit of detection for this assay: the smallest number of molecules from a single triplicate measurement that could be distinguished from zero with >97% confidence.

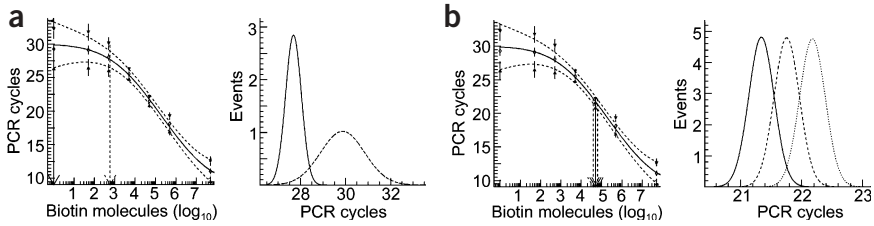

Figure 4 | Statistical analysis of biotin quantification. (**a**) Limit of detection of biotin measurement. Shown in the left graph is the plot of data with the mean PCR cycles/*N* biotin molecules as the solid line and the confidence belt, corresponding to 2 s.d., as dashed lines. These dashed lines define the confidence belt we used to calculate the limit of detection and precision of measurement. The solid arrow extending from the curve to the *x*-axis represents the signal measured in an assay with zero biotin molecules. The dashed arrow extending from the curve to the *x*-axis represents the intercept corresponding to the limit of detection illustrated in the right plot. The right plot illustrates the PCR cycle distribution, corresponding to the arrow in the left graph, representing a mean value (solid line) that is distinguished from the zero distribution (dashed line), given our confidence belt. (**b**) Precision of biotin measurement in the linear range of our assay. The left graph is the same as in **a**, except that the arrow corresponds to the intercept where the distributions in the right graph were evaluated. The right plot illustrates the PCR cycle distributions in the linear range of the assay where the standard deviation is low.

recombinant E2F1 and CDK2 over a 10^6-fold range, but presumably due to the relatively weak K_d (~1 μM) of the TrxA aptamers, the limit of detection in these experiments was only marginally lower than in ELISAs using existing antibodies (data not shown). Because of these findings, we chose to explore the use of single-chain variable fragments[16] (scFvs) from mammalian antibodies as affinity reagents. We made tadpoles to quantify PA, a protein subunit of anthrax toxin[30]. The tadpole heads were scFvs that bound PA with subnanomolar affinity[31], and the tails were 82-nucleotide dsDNAs (**Supplementary Table 2** online). The use of scFvs as tadpole heads presented substantial challenges (see Methods) because scFvs require disulfide bonds to form correctly, but the intein-mediated protein-DNA ligation requires reducing chemistry.

We expressed and purified soluble, full-length three-part fusion protein in the cytosol of *E. coli* Origami B (DE3), which carries lesions in *trxB, gor* and an unspecified suppressor mutation[32]. For this protein, we used an intein from *Mycobacterium xenopi gyrA* chosen after pilot experiments revealed that it gave efficient fusion protein hydrolysis even for the large scFv moiety. We treated the purified fusion with thiophenol, TCEP and the 82-mer to catalyze the excision-ligation reaction. We eluted the tadpole from the column and allowed the scFv disulfide bonds to reform by slowly dialyzing away the TCEP. We characterized the anti-PA tadpole by the methods used to evaluate the anti-biotin tadpole and measured its affinity for PA by evanescent wave experiments. The whole tadpole bound PA with a K_d of 4.5 nM, and the free scFv tadpole head bound with a K_d of 0.46 nM, a good match to the published affinity[31]. Reduced binding by the complete tadpole compared to the free scFv was largely due to slower association with the target (**Supplementary Table 4** online), possibly caused by the presence of the DNA tail; consistent with this idea, the off rate for the complete tadpole was also twofold slower.

Quantification of PA protein

We diluted protective antigen and coupled dilution mixtures to carboxylate-modified magnetic beads, as above. We blocked beads, probed with tadpoles for one hour, washed and quantified bound tadpoles by real-time PCR, as above (**Fig. 5a**). These measurements had a 10^9-fold dynamic range and a 10^6-fold linear range. We evaluated the PCR data for these PA assays as we did for the biotin measurements, using a confidence-belt[27,28] consisting of two standard deviations. At the low limit, the smallest number of molecules that could be distinguished with confidence from zero was 6,400 (**Fig. 6a**). In the linear range, the smallest difference in numbers of molecules these assays distinguished was about 10%, enabling for example 5.0×10^8 molecules to be distinguished from 5.5×10^8 molecules with >95% confidence (**Fig. 6b**). In our hands, this tadpole-based assay was thus considerably (10^9-fold) more sensitive than conventional ELISAs using goat anti-PA IgG (List Biologicals) or purified anti-PA scFv (**Supplementary Fig. 3** online).

Quantification of PA from human serum

We quantified PA in human serum. Unfractionated serum coupled to carboxylate beads bound tadpoles nonspecifically, regardless of the specificity of the protein head (data not shown). To circumvent this effect, we diluted PA into human serum and crudely fractionated it by running each dilution mixture on a 2-ml hydrophobic resin column (T-gel; Pierce). We eluted bound protein at low salt, coupled 2 μg of it to beads, blocked the beads, probed with tadpole,

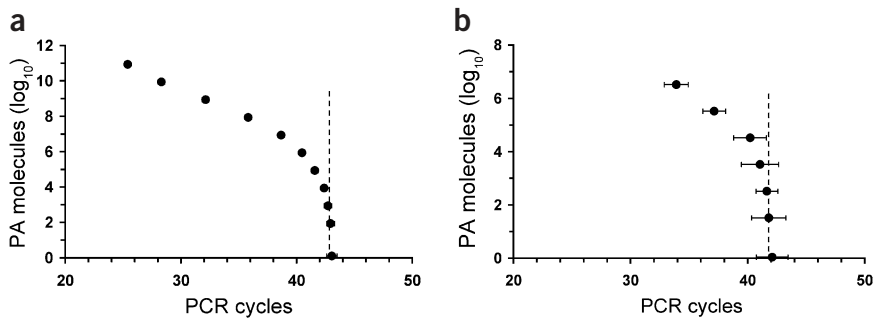

Figure 5 | Quantification of recombinant PA diluted in BSA or human serum. (**a**) Dilution in BSA. Circles represent the mean of ten separate measurements and the error bars represent 2 s.d. The *x*- and *y*-axes indicate the PCR cycles and numbers of PA molecules respectively, and the dashed line indicates the limit of detection, as in **Figure 3**. (**b**) Dilution in human serum. The *x*- and *y*-axes indicate the PCR cycles and numbers of PA molecules respectively, and the dashed line indicates the limit of detection. The circles represent the mean of eight separate measurements and the error bars represent 2 s.d.

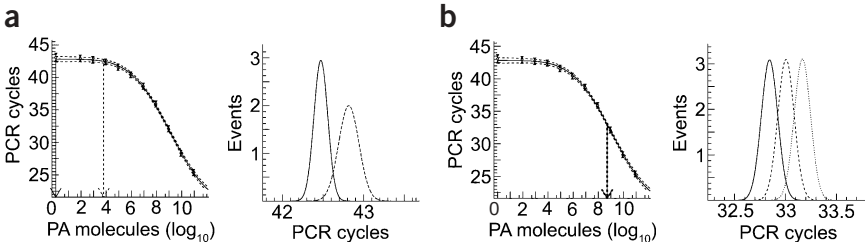

Figure 6 | Statistical analysis of PA quantification. (**a**) Limit of detection of PA measurement. The left plot shows the measurement data and error boundaries corresponding to 2 s.d. The dashed arrow represents the intercept corresponding to the limit of detection, as depicted in the right graph. The right plot illustrates the PCR cycle distribution at the limit of detection (solid line) compared to the distribution corresponding to zero molecules (dashed line). (**b**) Precision of PA measurement in the linear range. The left plot is the same as in **a** except that the arrow represents the intercept where the distributions in the right graph are evaluated. The right plot illustrates calculated PCR cycle distributions in the linear range of the assay where the standard deviation is low.

washed and quantified PA-bound tadpole as above. From serum, the precision of the measurement was lower than that in buffered BSA (the smallest difference in the numbers of molecules distinguished ranged from 55% to 98%), and the limit of detection corresponded to about 32,000 molecules (**Figs. 5b**). Tadpole-based detection from serum was thus about 10 times less sensitive than tadpole based detection in buffered BSA but about 200 times more sensitive than a corresponding ELISA of PA in serum using goat anti-PA (data not shown).

DISCUSSION

The intein-mediated chemistry used to assemble tadpoles from molecularly homeogeneous synthetic DNA and recombinant protein subunits is exceedingly efficient, precise and well controlled. In our hands, the chemistry allowed generation of milligram quantities. The resulting molecules were molecularly homogeneous: they had a precise (1:1) protein DNA stoichiometry with and the DNA was attached to the protein at a site chosen to not interfere with protein function. The molecules were made from a recombinant protein and synthetic DNA, allowing, as opposed to antibodies made in animals, new affinity proteins to be selected quickly *in vitro*, fusion proteins containing these affinity moieties to be expressed and purified rapidly and synthesis of the tadpoles to be scaled up rapidly. The resulting tadpoles are chemically stable for long periods at −20 °C.

We presented methods to measure the amounts of non-nucleic acid target molecules present in samples with a sensitivity approaching that for detecting nucleic acid targets by PCR. We adapted a statistical treatment[27,28], normally used to analyze low-incidence events, to evaluate real-time PCR data. This statistical treatment enabled us to distinguish the mean of a measured distribution from the mean of other measured distributions or of distribution from samples containing zero target molecules. This statistical treatment is valid throughout the linear and non-linear ranges of the data curve. Use of calibration standards internal to each PCR may further improve measurement precision, and the use of second–affinity reagent capture methods[33] ('sandwich methods') and methods dependent on binding by two different tadpoles (such as proximity-dependent ligation assays[5]) should further improve sensitivity, perhaps allowing detection and quantification of extremely small amounts of molecules from crude mixtures.

The antibodies now used in ELISAs are typically generated by the immune systems of vertebrate animals. These antibodies recognize limited sets of protein epitopes. But tadpole heads can be chosen from the entire range of possible affinity protein scaffolds, and affinity molecules based on those scaffolds can then be selected from combinatorial libraries *in vitro*. It should therefore be possible to make tadpoles against molecules including proteins, carbohydrates, fatty acids and organic chemicals that the vertebrate immune system does not normally recognize.

The assays are accurate and fast, requiring as little as three hours to measure the number of molecules present. The materials and skills needed to synthesize tadpoles are widely available, and the infrastructure, including real-time PCR machines for quantitative measurements, is widely deployed. Thus, these reagents might be applied to problems in biology, medicine, environmental monitoring and industry.

We also note that a reaction mix consisting of target, tadpole and polymerase can be considered a 'molecular amplifier'; viewed in this manner, tadpole-containing 'devices' may have utility beyond detection. Moreover, hybridized by their DNA tails, or oligomerized by their protein heads, tadpoles may also be useful in assembly and self-assembly of other nanoscale structures of defined geometry and function.

METHODS

Fusion-protein expression plasmids. Plasmids directed the synthesis of three-part fusion proteins that contained, from N to C terminus, an affinity domain with 6-His sequence, an intein and the CBD, except for the streptavidin fusion, which lacked the 6-His-tag. PCR products encoding affinity proteins were cloned into *Eco*RI and *Hin*dIII sites in the parent vectors pTYB1, pTWIN1 or pTWIN2[20]. The plasmids created, inteins used, *E. coli* expression host strains and DNA tail sequences are listed in **Supplementary Table 1** online. In initial attempts to express the scFv fusion proteins under oxidizing conditions, we targeted fusion proteins containing two model scFvs, 1H and 14B7[31], for export to the periplasm. Under these expression conditions, the proteins were hydrolyzed at the affinity protein-intein junction, even after changing residues at the hydrolysis junction[34] and co-expression of the chaperonin Skp1p[35–37] (data not shown).

DNA synthesis. We synthesized oligonucleotides on a controlled-pore glass (CPG) matrix with 2,500 Å pore size (Glen Research) using conventional phosphoramidite-based chemistries[38,39]. We incorporated a cysteine-like residue into the 5′ terminus of the forward strand of each dsDNA tail during synthesis. After synthesis, we treated the oligonucleotide-CPG matrix with 20% piperidine in dimethylformamide (DMF), washed with 10 ml of DMF, washed with 10 ml of acetonitrile and released oligonucleotides

from the CPG with 30% ammonium hydroxide. We deprotected samples at 65 °C for 1 h and purified DNA tails by anion-exchange chromatography. The DNA tail for the streptavidin tadpole (**Supplementary Table 2** online) contained a T7 gene 10 promoter that directed transcription away from the affinity protein head; other tails did not contain the T7 promoter. We activated the oligonucleotide for covalent coupling by reduction with 1 mM TCEP for 30 min prior to incubation with the activated thioester on the recombinant proteins.

Intein-mediated protein-DNA ligation. We expressed recombinant intein fusion proteins in strains listed in **Supplementary Table 1** online and purified these fusions from the soluble fraction of *E. coli* lysates using a chitin column. After thorough washing, we passed the cysteine-modified DNA over the column in the presence of 0.2% thiophenol and 1 mM TCEP. We incubated the equilibrated column at 22 °C for 16 to 20 h, collected the column eluate and exchanged the reaction buffer to phosphate-buffered saline (PBS) or Tris-buffered saline (TBS) using NAP-25 columns (Pharmacia). To allow the tadpoles containing scFv domains to refold into an active form, we exchanged into buffer containing 1 mM TCEP and allowed refolding by slowly dialyzing out the TCEP. We purified tadpoles to apparent homogeneity (as judged by observing a single band on SYBR gold–stained polyacrylamide gels) by serial ion-exchange, hydrophobic and affinity chromatography.

Characterization of tadpoles. To monitor the coupling of DNA to the protein head we resolved samples from before and after conjugation in native 12% Tris-borate-EDTA polyacrylamide gels, stained DNA with SYBR gold, imaged with a STORM 860 scanning fluorimeter, and displayed the image using ImageQuant software (Molecular Dynamics). To verify the composition of the tadpoles we resolved conjugates in a Superdex-200 size exclusion column (Pharmacia) and monitored absorbance at 260 nm. To confirm that tadpoles contain a protein head, we digested with 0.1 μg of tadpoles with proteinase K and examined the sample by PAGE as above. To test ligand-binding activities, we incubated 0.1 μg of tadpole and 0.5 μg of specific ligand for 1 h at 22 °C and resolved reaction mixtures on native 12% polyacrylamide gels as above.

Quantification of biotinylated BSA diluted in an excess of nonspecific organic molecules. We verified the biotin content of our biotinylated BSA (Pierce) and carbonic anhydrase preparations using 2-(4′-hydroxyazobenzene)-benzoic acid (HABA) in soluble spectrophotometric assays essentially as described[22]. To quantify the amount of biotin *in vitro*, we diluted 100 μg of biotinylated BSA (containing 4 biotins per BSA) into NDA- and Texas red-coupled BSA and bound to polystyrene microtiter plates. We blocked the plates with 0.2% casein, 0.01% Tween and 10 μg/ml calf thymus DNA in PBS, incubated with 100 nM streptavidin tadpole, washed away unbound species and amplified bound tadpole with 100 U of T7 RNA polymerase and 100 μM of each NTP for 4 h at 37 °C. We resolved synthesized RNA on a denaturing gel, stained with SYBR gold[40] (Molecular Probes) and scanned the gel with 100 micron resolution at a photomultiplier tube voltage of 1,000 mV on a STORM 860. We computed signal values using ImageQuant software and graphed relative signal as a function of biotinylated BSA.

Quantification of biotin using real-time PCR. We covalently attached biotinylated BSA to glycidyl ester[41] (epoxy-modified) M-270 magnetic beads (Dynal Biotech) in PBS at pH 8.5 for 3 h at 30 °C, and unreacted epoxy groups were quenched with 50 mM ethanolamine for 2 h. By measuring free protein concentration[42] before and after coupling and by direct ELISA of the beads using mouse anti-biotin (Pierce) and goat anti-mouse HRP conjugate (Pierce), we determined that each bead immobilized approximately 50 fg of biotinylated BSA (data not shown). We serially diluted these biotinylated BSA-labeled beads into dilutions of unlabeled beads to create samples containing a total of 1.0×10^7 beads. The samples were blocked with 0.2% casein, 0.01% Tween and 10 μg/ml calf thymus DNA in PBS in 100 μl at 22 °C for 1 h and probed with 100 nM anti-biotin tadpole in 100 μl at 22 °C for 20 min. We quickly washed beads 5× with 500 μl blocking buffer for each wash at 22 °C and quantified the amount of bound tadpole by real-time PCR[24] using specific primers (**Supplementary Table 3** online) at a concentration of 0.2 μM with SYBR green[43] in 50 μl reactions.

Quantification of recombinant PA diluted into BSA. We diluted recombinant PA into 0.1 mg/ml BSA and covalently-coupled 10 μg of total protein from the dilutions to 1 μm diameter carboxylate-modified 'MyOne' magnetic beads (Dynal Biotech) using 1-ethyl-3-(3-dimethylaminopropyl) carbodimide[25] (EDC) and N-hydroxy-succinimide[26] (NHS) for 2 h at 22 °C and quenched with 50 mM ethanolamine as above. Under these conditions, ~10 fg of total protein was immobilized to each bead. We confirmed the numbers of PA molecules on the beads by Bradford analysis[42] of the protein solutions before and after coupling and by direct ELISA of the beads using goat anti-PA (List Biologicals) and donkey anti-goat HRP conjugates (Pierce) (data not shown). We blocked beads with 0.2% casein, 0.01% Tween and 10 μg/ml calf thymus DNA in 10 mM Tris pH 8, 0.2 M NaCl, probed samples with 100 nM anti-PA tadpole, washed using blocking buffer and heated to 95 °C for 5 minutes to release bound tadpoles. We quantified tadpole tail by real-time PCR using primers that hybridized to the oligonucleotide tail (**Supplementary Table 3** online) and SYBR green fluorescent dye.

Quantification of recombinant PA diluted into human serum. We diluted PA into 100 μl aliquots of EDTA-treated human serum and fractionated each dilution over 2 ml of T-gel (Pierce). We eluted bound protein with low salt buffer and covalently-coupled 2 μg of protein from the eluate of each column to carboxylate-modified 'MyOne' magnetic beads as above. We assayed these beads for PA using the anti-PA tadpole as above.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

The authors thank R. Fu for technical assistance with early real-time PCR experiments, L. Lok for invaluable advice on statistical analysis, O. Resnekov for suggesting T-gel for fractionating serum, C. Cantor for the gene encoding the streptavidin core subunit, J. Maynard for plasmids encoding anti-PA scFvs, purified scFvs and antibodies, and A. Coleman-Lerner, K. Benjamin, A. Gordon, L. Lok, A. Mendelsohn, O. Resnekov, E. Serra and R. Yu for useful discussions and comments on the manuscript. Some support for this work came from grants from the National Science Foundation and Defense Advanced Research Projects Agency, but main support came from the Center for Genomic Experimentation and Computation, a National Institutes of Health Center of Excellence in Genomic Science (P50 HG02370).

AUTHORS' CONTRIBUTIONS

I.B.B. conceived the intein-mediated protein-DNA ligation synthesis and performed the experiments described here. K.Y. helped with nucleic acid constructions, protein expression, and in troubleshooting real-time PCR. A.G. provided the statistical methods used to quantify real-time PCR data. R.C. articulated the importance of controlled covalent linkage of nucleic acids to affinity molecules for biological measurements and provided input into initial experiments. R.B. input into the experiments and their quantification. R.B. and I.B. wrote the manuscript and guarantee the integrity of its findings.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Received 6 November; accepted 24 November 2004
Published online at http://www.nature.com/naturemethods/

1. Engvall, E. & Perlman, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871–874 (1971).
2. Yalow, R.S. & Berson, S.A. Assay of plasma insulin in human subjects by immunological methods. *Nature* **184** Suppl. 21, 1648–1649 (1959).
3. Brenner, S. & Lerner, R.A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA* **89**, 5381–5383 (1992).
4. Sano, T., Smith, C.L. & Cantor, C.R. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* **258**, 120–122 (1992).
5. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.* **20**, 473–477 (2002).
6. Hendrickson, E.R., Truby, T.M., Joerger, R.D., Majarian, W.R. & Ebersole, R.C. High sensitivity multianalyte immunoassay using covalent DNA-labeled antibodies and polymerase chain reaction. *Nucleic Acids Res.* **23**, 522–529 (1995).
7. McKie, A., Samuel, D., Cohen, B. & Saunders, N.A. A quantitative immuno-PCR assay for the detection of mumps-specific IgG. *J. Immunol. Methods* **270**, 135–141 (2002).
8. Niemeyer, C.M. *et al.* Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR. *Nucleic Acids Res.* **27**, 4553–4561 (1999).
9. Zhang, H.T., Kacharmina, J.E., Miyashiro, K., Greene, M.I. & Eberwine, J. Protein quantification from complex protein mixtures using a proteomics methodology with single-cell resolution. *Proc. Natl. Acad. Sci. USA* **98**, 5497–5502 (2001).
10. Nam, J.M., Thaxton, C.S. & Mirkin, C.A. Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science* **301**, 1884–1886 (2003).
11. Perler, F.B. *et al.* Intervening sequences in an Archaea DNA polymerase gene. *Proc. Natl. Acad. Sci. USA* **89**, 5577–5581 (1992).
12. Muir, T., Sondhi, D. & Cole, P. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. USA* **95**, 6705–6710 (1998).
13. Sydor, J.R., Mariano, M., Sideris, S. & Nock, S. Establishment of intein-mediated protein ligation under denaturing conditions: C-terminal labeling of a single-chain antibody for biochip screening. *Bioconjug. Chem.* **13**, 707–712 (2002).
14. Lovrinovic, M. *et al.* Synthesis of protein-nucleic acid conjugates by expressed protein ligation. *Chem. Commun. (Camb.)* **7**, 822–823 (2003).
15. Green, N.M. & Melamed, M.D. Optical rotatory dispersion, circular dichroism and far-ultraviolet spectra of avidin and streptavidin. *Biochem. J.* **100**, 614–621 (1966).
16. Huston, J.S. *et al.* Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **85**, 5879–5883 (1988).
17. Colas, P. *et al.* Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548–550 (1996).
18. Watanabe, T. *et al.* The roles of the C-terminal domain and type III domains of chitinase A1 from Bacillus circulans WL-12 in chitin degradation. *J. Bacteriol.* **176**, 4465–4472 (1994).
19. Stetsenko, D.A. & Gait, M.J. New phosphoramidite reagents for the synthesis of oligonucleotides containing a cysteine residue useful in peptide conjugation. *Nucleosides Nucleotides Nucleic Acids* **19**, 1751–1764 (2000).
20. Chong, S. *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271–281 (1997).
21. Sano, T., Pandori, M.W., Chen, X., Smith, C.L. & Cantor, C.R. Recombinant core streptavidins. A minimum-sized core streptavidin has enhanced structural stability and higher accessibility to biotinylated macromolecules. *J. Biol. Chem.* **270**, 28204–28209 (1995).
22. Green, N.M. A Spectrophotometric Assay for Avidin and Biotin Based on Binding of Dyes by Avidin. *Biochem. J.* **94**, 23C–24C (1965).
23. Van Gelder, R.N. *et al.* Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87**, 1663–1667 (1990).
24. Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology* **11**, 1026–1030 (1993).
25. Timkovich, R. Detection of the stable addition of carbodiimide to proteins. *Anal. Biochem.* **79**, 135–143 (1977).
26. Staros, J.V., Wright, R.W. & Swingle, D.M. Enhancement by N-hydroxysulfosuccinimide of water-soluble carbodiimide- mediated coupling reactions. *Anal. Biochem.* **156**, 220–222 (1986).
27. Neyman, J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. London* **A236**, 333–380 (1937).
28. Feldman, G.J. & Cousins, R.D. Unified approach to the classical statistical analysis of small signals. *Phys. Rev.* **57**, 3873–3889 (1998).
29. Fabbrizio, E. *et al.* Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* **18**, 4357–4363 (1999).
30. Leppla, S.H. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. *Proc. Natl. Acad. Sci. USA* **79**, 3162–3166 (1982).
31. Maynard, J.A. *et al.* Protection against anthrax toxin by recombinant antibody fragments correlates with antigen affinity. *Nat. Biotechnol.* **20**, 597–601 (2002).
32. Prinz, W.A., Aslund, F., Holmgren, A. & Beckwith, J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661–15667 (1997).
33. Salmon, S.E., Mackey, G. & Fudenberg, H.H. Sandwich" solid phase radioimmunoassay for the quantitative deterimination of human immunoglobulins. *J. Immunol.* **103**, 129–137 (1969).
34. Chen, L., Benner, J. & Perler, F.B. Protein splicing in the absence of an intein penultimate histidine. *J. Biol. Chem.* **275**, 20431–20435 (2000).
35. Schafer, U., Beck, K. & Muller, M. Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J. Biol. Chem.* **274**, 24567–24574 (1999).
36. Hayhurst, A. & Harris, W.J. *Escherichia coli* skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments. *Protein Expr. Purif.* **15**, 336–343 (1999).
37. Bothmann, H. & Pluckthun, A. Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nat. Biotechnol.* **16**, 376–380 (1998).
38. Caruthers, M.H. *et al.* New chemical methods for synthesizing polynucleotides. *Nucleic Acids Symp. Ser.* **7**, 215–223 (1980).
39. Caruthers, M.H. *et al.* Deoxyoligonucleotide synthesis via the phosphoramidite method. *Gene Amplif. Anal.* **3**, 1–26 (1983).
40. Tuma, R.S. *et al.* Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Anal. Biochem.* **268**, 278–288 (1999).
41. Zacharieva, E.I., Georgieva, M.P., Kabaivanov, V.S. & Popov, D.V. Preparation and investigation of cross-linked copolymers containing antitumour agents. *Biomaterials* **4**, 197–200 (1983).
42. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
43. Garner, D.L., Johnson, L.A., Yue, S.T., Roth, B.L. & Haugland, R.P. Dual DNA staining assessment of bovine sperm viability using SYBR-14 and propidium iodide. *J. Androl.* **15**, 620–629 (1994).