

Printing Proteins as Microarrays for High-Throughput Function Determination

Gavin MacBeath^{1*} and Stuart L. Schreiber²

as single-nucleotide polymorphism analysis, where single-mismatch resolution, sensitivity, cost, and ease of use are important factors. Moreover, the sensitivity of this system, which has yet to be totally optimized, points toward a potential method for detecting oligonucleotide targets without the need for target amplification schemes such as the polymerase chain reaction.

References and Notes

1. C. A. Mirkin, R. L. Letsinger, R. C. Mucic, J. J. Storhoff, *Nature* **382**, 607 (1996).
2. R. Elghanian, J. J. Storhoff, R. C. Mucic, R. L. Letsinger, C. A. Mirkin, *Science* **277**, 1078 (1997).
3. J. J. Storhoff, R. Elghanian, R. C. Mucic, C. A. Mirkin, R. L. Letsinger, *J. Am. Chem. Soc.* **120**, 1959 (1998).
4. L. A. Chrisey, G. U. Lee, C. E. O'Ferrall, *Nucleic Acids Res.* **24**, 3031 (1996).
5. Supplementary material is available at Science Online at www.sciencemag.org/feature/data/1051941.shl.
6. G. W. Hacker, in *Colloidal Gold: Principles, Methods, and Applications*, M. A. Hayat, Ed. (Academic Press, San Diego, CA, 1989), vol. 1, chap. 10.
7. I. Zehbe et al., *Am. J. Pathol.* **150**, 1553 (1997).
8. R. C. Mucic, thesis, Northwestern University, Evanston, IL (1999).
9. For a review on oligonucleotide arrays, see S. P. A. Fodor, *Science* **277**, 393 (1997).
10. For the experiments reported in Fig. 2, dissociation measurements were made from the surface of glass beads 250 to 300 μm in diameter (Polysciences, Warrington, PA) rather than planar substrates to increase the UV-visible and fluorescence signal intensity.
11. 5'-Cy3-labeled oligonucleotide probes were synthesized on an Expedite automated synthesizer (Millipore, Bedford, MA) using Cy3 phosphoramidite (Glen Research, Sterling, VA) as the label source. Arrays of spots 175 μm in diameter separated by 375 μm were patterned with a GMS 417 Microarrayer (Genetic Microsystems, Woburn, MA).
12. R. K. Saiki and H. A. Erlich, in *Mutation Detection*, R. G. H. Cotton, E. Edkins, S. Forrest, Eds. (Oxford Univ. Press, Oxford, 1998), chap. 7.
13. S. Ikuta, K. Takagi, R. B. Wallace, K. Itakura, *Nucleic Acids Res.* **15**, 797 (1987).
14. First, 20 μl of a 1 nM solution of synthetic target in 2 \times phosphate-buffered saline (PBS) [0.3 M NaCl and 10 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ buffer (pH 7)] was hybridized to the array for 4 hours at 10°C in a CoverWell PC20 hybridization chamber (Grace Bio-Labs, Bend, OR) and was then washed at 10°C with clean PBS buffer. Next, 20 μl of a 100 pM solution of either oligonucleotide-functionalized gold nanoparticles or 5'-Cy3-labeled probe in 2 \times PBS was hybridized to the array for 4 hours at 10°C in a fresh hybridization chamber. The array was then washed at the stringency temperature (shown in Fig. 3) with clean 2 \times PBS buffer for 2 min. Arrays labeled with nanoparticle probes were washed twice at room temperature with 2 \times PBN [0.3 M NaNO_3 and 10 mM $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ buffer (pH 7)], then submerged in Silver Enhancer Solution (Sigma) for 5 min at room temperature and washed with water.
15. H. T. Allawi and J. SantaLucia Jr., *Biochemistry* **36**, 10581 (1988).
16. The temperature ranges for the melting curves in Fig. 2 do not correspond exactly with the stringency temperatures associated with the oligonucleotide array experiments reported in Fig. 3. This is probably because the two sets of experiments are not identical with respect to the substrate.
17. J. E. Forman, I. D. Walton, D. Stern, R. P. Rava, M. O. Trulsson, in *Molecular Modeling of Nucleic Acids*, N. B. Leontis and J. SantaLucia Jr., Eds. [American Chemical Society (ACS) Symposium Series 682, ACS, Washington, DC, 1998], pp. 206–228.
18. C.A.M. and R.L.L. acknowledge the Army Research Office (DAAG55-97-1-0133) and the National Institute of General Medical Sciences (GM 57356) for support of this work.

Systematic efforts are currently under way to construct defined sets of cloned genes for high-throughput expression and purification of recombinant proteins. To facilitate subsequent studies of protein function, we have developed miniaturized assays that accommodate extremely low sample volumes and enable the rapid, simultaneous processing of thousands of proteins. A high-precision robot designed to manufacture complementary DNA microarrays was used to spot proteins onto chemically derivatized glass slides at extremely high spatial densities. The proteins attached covalently to the slide surface yet retained their ability to interact specifically with other proteins, or with small molecules, in solution. Three applications for protein microarrays were demonstrated: screening for protein-protein interactions, identifying the substrates of protein kinases, and identifying the protein targets of small molecules.

Historically, genome-wide screens for protein function have been carried out with random cDNA libraries. Most frequently, the libraries are prepared in phage vectors and the expressed proteins immobilized on a membrane by a plaque lift procedure. This method has been effective for a variety of applications (1–4), but it has several limitations. Most clones in the library do not encode proteins in the correct reading frame, and most proteins are not full-length. Bacterial expression of eukaryotic genes frequently fails to yield correctly folded proteins, and products derived from abundant transcripts are overrepresented. Moreover, because plaque lifts are not amenable to miniaturization on the micrometer scale, it is hard to imagine screening all the proteins of an organism hundreds or thousands of times by this approach.

With the advent of high-throughput molecular biology, it is now possible to prepare large, normalized collections of cloned genes. UniGene sets in the form of polymerase chain reaction products have been used extensively over the past decade to construct DNA microarrays for the study of transcriptional regulation (5). Recently, spatially segregated clones in expression vectors were used to study protein function in vivo using the yeast two-hybrid system (6) and in vitro using biochemical assays (7). We have built on these efforts by developing microarray-based methods to study protein function.

To accomplish these goals, it is necessary to immobilize proteins on a solid support in a way that preserves their folded conformations. One

group has described methods of arraying functionally active proteins, using microfabricated polyacrylamide gel pads to capture their samples and microelectrophoresis to accelerate diffusion (8). In contrast, we have immobilized proteins by covalently attaching them to the smooth, flat surface of glass microscope slides. One of our primary objectives in pursuing this approach was to make the technology easily accessible and compatible with standard instrumentation. We use a variety of chemically derivatized slides that can be printed and imaged by commercially available arrayers and scanners. For most applications, we use slides that have been treated with an aldehyde-containing silane reagent (9). The aldehydes react readily with primary amines on the proteins to form a Schiff's base linkage. Because typical proteins display many lysines on their surfaces as well as the generally more reactive α -amine at their NH_2 -termini, they can attach to the slide in a variety of orientations, permitting different sides of the protein to interact with other proteins or small molecules in solution.

To fabricate protein microarrays, we use a high-precision contact-printing robot (10) to deliver nanoliter volumes of protein samples to the slides, yielding spots about 150 to 200 μm in diameter (1600 spots per square centimeter). The proteins are printed in phosphate-buffered saline with 40% glycerol included to prevent evaporation of the nanodroplets. It is important that the proteins remain hydrated throughout this and subsequent steps to prevent denaturation. After a 3-hour incubation, the slides are immersed in a buffer containing bovine serum albumin (BSA). This step not only quenches the unreacted aldehydes on the slide, but also forms a molecular layer of BSA that reduces nonspecific binding of other proteins in subsequent steps.

Although appropriate for most applications, aldehyde slides cannot be used when peptides

¹Center for Genomics Research, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA.
²Howard Hughes Medical Institute (HHMI), Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: gavin_macbeath@harvard.edu

or very small proteins are printed, presumably because the BSA obscures the molecules of interest. For such applications, we use BSA-NHS (BSA-*N*-hydroxysuccinimide) slides that are fabricated by first attaching a molecular layer of BSA to the surface of glass slides and

then activating the BSA with *N,N'*-disuccinimidyl carbonate (11). The activated lysine, aspartate, and glutamate residues on the BSA react readily with surface amines on the printed proteins to form covalent urea or amide linkages. The slides are then quenched with glycine. In contrast to the aldehyde slides, proteins or peptides printed on BSA-NHS slides are displayed on top of the BSA monolayer, rendering them accessible to macromolecules in solution.

As a first application of protein microarrays, we have looked at protein-protein interactions. Until now, only the yeast two-hybrid system has been used to investigate such interactions systematically on a genome-wide scale (6). This *in vivo* method, although easy to implement and of great utility, has several limitations. Proteins that function as transcriptional activators yield false positives when expressed as DNA binding domain fusions. False negatives are encountered when proteins are displayed inappropriately or when the DNA binding domain fusions are produced in excess. Proteins that do not fold correctly in yeast are inaccessible, and posttranslational modifications (such as phosphorylation or glycosylation) cannot be controlled. Finally, it is impossible to control the environment (e.g., ion concentration, presence or absence of cofactors, temperature) during the experiment.

To determine whether microarrays could be used for these types of studies, we selected three pairs of proteins that are known to interact: protein G and immunoglobulin G (IgG) (12); p50 (of the nuclear factor NF- κ B complex) and the NF- κ B inhibitor I κ B α (13); and the FKBP12-rapamycin binding (FRB) domain of FKBP-rapamycin-associated protein (FRAP) and the human immunophilin FKBP12 (12 kD FK506-binding protein) (14). The first two interactions occur without special requirements, whereas the third interaction depends on the presence of the small molecule rapamycin (14). We arrayed the first protein of each pair in quadruplicate on five aldehyde slides and probed each slide with a different fluorescently labeled protein (11).

The slide in Fig. 1A was probed with BODIPY-FL-conjugated IgG, washed, and

scanned with an ArrayWoRx fluorescence slide scanner (15). As anticipated, only the spots containing protein G were visible, indicating that the immobilized protein is able to retain its functional properties on the glass surface. Similarly, only the p50-containing spots were visible on the slide probed with Cy3-I κ B α (Fig. 1B) (15). For Cy5-FKBP12, binding to FRB was observed only when rapamycin was added (Fig. 1, C and D). Because the three fluorophores used for these studies have nonoverlapping excitation and emission spectra, we were also able to detect these interactions simultaneously (Fig. 1E).

By varying the concentration of FRB (the protein being immobilized), we found that at concentrations above 1 mg/ml, the fluorescence of the spots began to saturate. Below this, fluorescence scaled linearly with decreasing concentrations of FRB. All proteins immobilized on the slides described here were spotted at 100 μ g/ml. Because only a few microliters of each protein are sufficient to fabricate thousands of microarrays, purified proteins may be readily obtained by high-throughput expression and purification, or even by *in vitro* transcription/translation (16).

Much lower concentrations are needed for the solution-phase protein. In the case of Cy5-FKBP12, fluorescence scaled linearly with protein concentration over four orders of magnitude (11). Specific binding could be detected using Cy5-FKBP12 concentrations as low as 150 pg/ml (\sim 12.5 pM). Concentrations in this range are accessible not only with purified proteins, but also with fluorescently labeled proteins from cell lysates. Thus, specific interactions, once defined, may potentially be exploited to quantify protein abundance and modification in whole cells or tissues.

At the spot density used for these studies, it was possible to fit more than 10,000 samples in about half the area of a standard (2.5 cm by 7.5 cm) slide. To investigate the feasibility of detecting a single specific interaction in this larger context, we prepared a slide containing 60 rows and 180 columns of spatially separated spots. Protein G was spotted 10,799

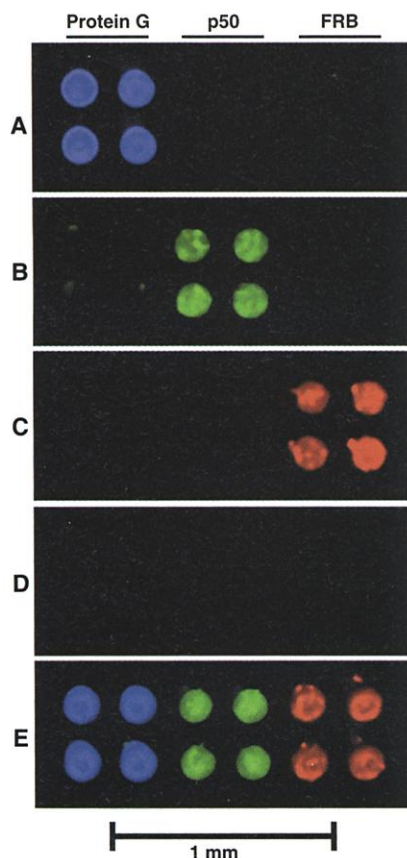
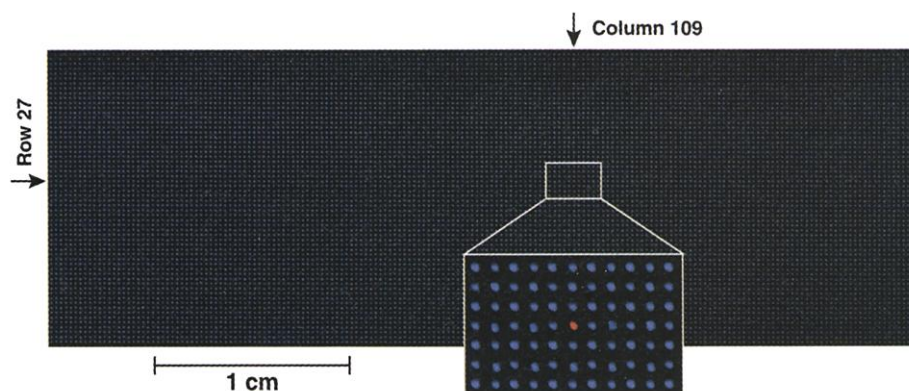


Fig. 1. Detecting protein-protein interactions on glass slides. (A) Slide probed with BODIPY-FL-IgG (0.5 μ g/ml). (B) Slide probed with Cy3-I κ B α (0.1 μ g/ml). (C) Slide probed with Cy5-FKBP12 (0.5 μ g/ml) and 100 nM rapamycin. (D) Slide probed with Cy5-FKBP12 (0.5 μ g/ml) and no rapamycin. (E) Slide probed with BODIPY-FL-IgG (0.5 μ g/ml), Cy3-I κ B α (0.1 μ g/ml), Cy5-FKBP12 (0.5 μ g/ml), and 100 nM rapamycin. In all panels, BODIPY-FL, Cy3, and Cy5 fluorescence were false-colored blue, green, and red, respectively.

Fig. 2. A single slide holding 10,800 spots. Protein G was printed 10,799 times. A single spot of FRB was printed in row 27, column 109. The slide was probed with BODIPY-FL-IgG (0.5 μ g/ml), Cy5-FKBP12 (0.5 μ g/ml), and 100 nM rapamycin. BODIPY-FL and Cy5 fluorescence were false-colored blue and red, respectively.



were then visualized using an automated light microscope (20) and individual frames were stitched together. As anticipated, only the specific substrates for each enzyme were phosphorylated (Fig. 3).

As the third and most demanding application, we sought to use protein microarrays to identify protein–small molecule interactions. With the advent of high-throughput, cell-based screening, more and more compounds are being identified on the basis of their biological activity. Once a “hit” is obtained, the daunting task of target identification remains. Several innovative techniques have been developed to address this bottleneck (4, 21–23), but they all suffer from the common limitations imposed by using random cDNA libraries. As an alternative, we sought to develop microarray-based assays that use purified, full-length, correctly folded proteins.

To test this approach, we chose three unrelated small molecules for which specific protein receptors are available: DIG, a derivative of the steroid digoxigenin that is recognized by a mouse monoclonal antibody (24); biotin, a common vitamin recognized by the bacterial protein streptavidin (25); and AP1497 (Fig. 4), a synthetic pipecolyl α -ketoamide designed to be recognized by FKBP12 (26). The proteins from all three pairs were spotted in quadruplicate on four aldehyde slides, and each slide was probed with a different small molecule. Rather than labeling the compounds directly, each ligand was coupled to BSA that had previously been labeled with a unique fluorophore (Alexa₄₈₈, Cy3, or Cy5) (15). As anticipated, fluorescence localized to the appropriate spots in all three cases (Fig. 5, A to C). Because the fluorophores used for these studies have non-overlapping excitation and emission spectra, we were also able to detect all three interactions simultaneously (Fig. 5D).

(11). This means that interactions in the micromolar range can easily be observed. The fact that the intensity of the fluorescence did not vary appreciably as the affinity of the interaction was lowered can be attributed to the multivalency of the BSA conjugates (avidity effects). In the context of small-molecule microarrays (27), we have previously shown that when these three compounds are immobilized on a glass surface and then probed with Cy5-labeled FKBP12 (a monomeric protein), the intensity of the fluorescence correlates very well with the affinity of the interaction. Thus, by controlling the valency of the probe, we can choose whether to observe differences in affinity or to favor the detection of low-affinity interactions. The combination of these two approaches may prove useful in the identification of both primary and secondary drug targets.

Although traditional biochemical methods have yielded invaluable insight into protein function on a case-by-case basis, they cannot realistically be applied to the study of every protein in a cell, tissue, or organism. If we hope to assign function on a broader level, we must turn to miniaturized assays that can be performed in a highly parallel format. It is certainly a daunting task to express and purify thousands of different proteins, and some

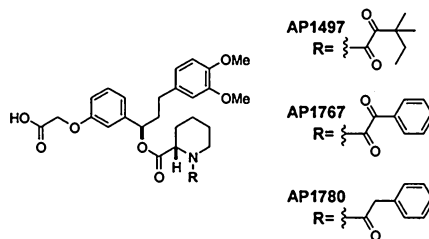
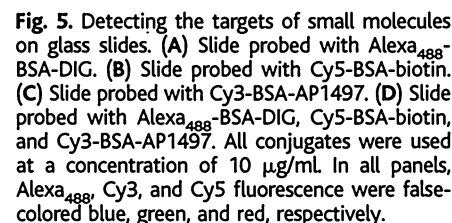


Fig. 4. Synthetic ligands for FKBP12. The compounds were coupled to BSA through their carboxyl groups (via a flexible linker).



The Global Spread of Malaria in a Future, Warmer World

David J. Rogers^{1*} and Sarah E. Randolph²

The frequent warnings that global climate change will allow *falciparum* malaria to spread into northern latitudes, including Europe and large parts of the United States, are based on biological transmission models driven principally by temperature. These models were assessed for their value in predicting present, and therefore future, malaria distribution. In an alternative statistical approach, the recorded present-day global distribution of *falciparum* malaria was used to establish the current multivariate climatic constraints. These results were applied to future climate scenarios to predict future distributions, which showed remarkably few changes, even under the most extreme scenarios.

Predictions of global climate change have stimulated forecasts that vector-borne diseases will spread into regions that are at present too cool for their persistence (1–5). For example, life-threatening cerebral malaria, caused by *Plasmodium falciparum* transmitted by anopheline mosquitoes, is predicted to reach the central or northern regions of Europe and large parts of North America (2, 4). *falciparum* malaria is the most severe form of the human disease, causing most of the ~1 million deaths worldwide among the ~273 million cases in 1998 (6). Despite these figures, the epidemiology of malaria, like many other vector-borne tropical diseases, remains inadequately understood. Only the most general of maps for its worldwide distribution are available (7), and its global transmission patterns cannot be modeled satisfactorily because crucial parameters and their relations with environmental factors have not yet been quantified. Most importantly, absolute mosquito abundance has not yet been related to multivariate climate.

Nevertheless, the problem of malaria has led to its being included in most predictions about the impact of climate change on the future distribution of vector-borne diseases (8). These studies, which draw on the forecasts of future climate from various global circulation models (GCMs) (9, 10), generally use only one or at most two climatic variables to make their predictions. Biological models for malaria distribution are based principally on the temperature dependence of mosquito blood-feeding intervals and longevity and the development period of the malaria parasite within the mosquito, each of which affects the rate of transmission (4, 11). Those models based on threshold values include a lower temperature threshold, below which all development of the malaria parasite ceases, and an upper limit of mosquito

lethality (2). In addition, the suitability (or unsuitability) of habitats for these vectors, which require a minimum atmospheric moisture, is defined by the ratio of rainfall to potential evapotranspiration (2). The output of such models, therefore, represents predicted areas where parasite development within the vector is fast enough to be completed before the vector dies, bounded by limits imposed by habitat suitability (2). The fit of these predictions to the current global malaria situation shows noticeable mismatches in certain places (12); false predictions of presence (e.g., over the eastern half of the United States) are accounted for by past control measures or by “peculiar vector biogeography,” whereas false predictions of absence are dismissed as model errors (2).

Refinements of these biological models (3–5) are based on modifications of an equation describing transmission potential, expressed as the basic reproduction number R_0 , which must equal at least 1 for disease persistence (13, 14). For an estimation of the correct value of R_0 from which to predict malaria distribution, absolute, not relative, estimates of all quantities in the equation are needed. Instead, by omitting certain unquantified but important parameters and rearranging the equation (15), a relative measure of “epidemic potential” (EP) [now “transmission potential” (5)] has been derived as the reciprocal of the vector/host ratio required for disease persistence. This predicts a more extensive present-day distribution of malaria than is currently observed (12). The ratio of future EP to present EP is then presented as indicating the relative degree of the future risk of malaria, but this is an inappropriate measure of changing risk because a high ratio may still leave $R_0 < 1$.

Until such biological approaches can give accurate descriptions of the current situation of global malaria, they cannot be used to give reliable predictions about the future. Instead, an alternative two-step statistical approach to mapping vector-borne diseases gave a better description of the present global distribution of *falciparum* malaria and predicted remarkably few future changes, even under the most ex-

References and Notes

1. R. A. Young and R. W. Davis, *Science* **222**, 778 (1983).
2. A. B. Sparks, N. G. Hoffman, S. J. McConnell, D. M. Fowlkes, B. K. Kay, *Nature Biotechnol.* **14**, 741 (1996).
3. R. Fukunaga and T. Hunter, *EMBO J.* **16**, 1921 (1997).
4. H. Tanaka, N. Ohshima, H. Hidaka, *Mol. Pharmacol.* **55**, 356 (1999).
5. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995).
6. P. Uetz et al., *Nature* **403**, 623 (2000).
7. M. R. Martzen et al., *Science* **286**, 1153 (1999).
8. P. Arenkov et al., *Anal. Biochem.* **278**, 123 (2000).
9. Aldehyde slides were purchased from TeleChem International (Cupertino, CA) under the trade name SuperAldehyde Substrates.
10. For Figs. 1, 3, and 5, proteins were spotted using a GMS 417 Arrayer (Affymetrix, Santa Clara, CA). For Fig. 2, proteins were spotted using a split pin arrayer constructed following directions on P. Brown's Web page (<http://cmgm.stanford.edu/pbrown/>).
11. For detailed protocols and additional data, see *Science* Online (www.sciencemag.org/feature/data/1053284.shl).
12. L. Björck and G. Kronvall, *J. Immunol.* **133**, 969 (1984).
13. P. A. Baeuerle and D. Baltimore, *Science* **242**, 540 (1988).
14. E. J. Brown et al., *Nature* **369**, 756 (1994).
15. Four different fluorophores were used in these studies. BODIPY-FL and Alexa488 were obtained from Molecular Probes (Eugene, OR) and have excitation/emission maxima of 503/512 nm and 499/520 nm, respectively. Cy3 and Cy5 were obtained from Amersham Pharmacia Biotech (Piscataway, NJ) and have excitation/emission maxima of 552/565 nm and 650/667 nm, respectively. Fluorescence was visualized with an ArrayWoRx fluorescence slide scanner (Applied Precision, Issaquah, WA) with appropriate excitation/emission filter sets for each dye.
16. K. Madin, T. Sawasaki, T. Ogasawara, Y. Endo, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 559 (2000).
17. B. E. Kemp, D. J. Graves, E. Benjamini, E. G. Krebs, *J. Biol. Chem.* **252**, 4888 (1977).
18. A. A. DePaoli-Roach, *J. Biol. Chem.* **259**, 12144 (1984).
19. R. Marais, J. Wynne, R. Treisman, *Cell* **73**, 381 (1993).
20. DeltaVision microscope (Applied Precision, Issaquah, WA).
21. M. J. Caterina et al., *Nature* **389**, 816 (1997).
22. E. J. Licitra and J. O. Liu, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12817 (1996).
23. P. P. Sche, K. M. McKenzie, J. D. White, D. J. Austin, *Chem. Biol.* **6**, 707 (1999).
24. Mouse anti-digoxigenin IgG clone 1.71.256 (Boehringer Mannheim).
25. I. Chalet and F. J. Wolf, *Arch. Biochem. Biophys.* **106**, 1 (1964).
26. D. A. Holt et al., *J. Am. Chem. Soc.* **115**, 9925 (1993).
27. G. MacBeath, A. N. Koehler, S. L. Schreiber, *J. Am. Chem. Soc.* **121**, 7967 (1999).
28. We thank R. Peters and T. Maniatis at Harvard University for samples of p50 and IxB α and D. Holt and T. Clackson at Ariad Pharmaceuticals Inc. for samples of AP1497, AP1767, and AP1780. We thank the Harvard Center for Genomics Research for support of the G.M. laboratory and the National Institute of General Medical Sciences for support of the S.L.S. laboratory. G.M. was also supported in part by a fellowship from the Cancer Research Institute. S.L.S. is an HHMI investigator.

¹Trypanosomiasis and Land-use in Africa Research Group, ²Oxford Tick Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

*To whom correspondence should be addressed. E-mail: david.rogers@zoology.ox.ac.uk