66. T. A. Stewart, P. K. Pattengale, P. Leder, *Cell* **38**, 627 (1984).
67. C. Schoenenberger *et al.*, *EMBO J.* **7**, 169 (1988).
68. U. Ruther, C. Garber, D. Komitowski, R. Muller, E. F. Wagner, *Nature* **325**, 412 (1987); U. Ruther, D. Komitowski, F. R. Schubert, E. F. Wagner, *Oncogene* **4**, 861 (1989).
69. R. L. Brinster *et al.*, *Cell* **37**, 367 (1984); R. D. Palmiter, H. Y. Chen, A. Messing, R. L. Brinster, *Nature* **316**, 457 (1985); T. A. Van Dyke *et al.*, *J. Virol.* **61**, 2029 (1987).
70. D. M. Ornitz *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 399 (1985).
71. S. Efrat and D. Hanahan, *Mol. Cell. Biol.* **7**, 192 (1987); S. Efrat, S. Baekkeskov, D. Lane, D. Hanahan, *EMBO J.* **6**, 2699 (1987).
72. K. A. Mahon *et al.*, *Science* **235**, 1622 (1987).
73. L. J. Field, *ibid.* **239**, 1029 (1988).
74. E. Sinn *et al.*, *Cell* **49**, 465 (1987).
75. C. J. Quaife *et al.*, *ibid.* **48**, 1023 (1987).
76. A. S. Tsukamoto *et al.*, *ibid.* **55**, 619 (1988).
77. W. J. Muller *et al.*, *ibid.* **54**, 105 (1988).
78. L. Bouchard, L. Lamarre, P. J. Tremblay, P. Jolicoeur, *ibid.* **57**, 931 (1989).
79. V. L. Bautch, S. Toda, J. A. Hassell, D. Hanahan, *ibid.* **51**, 529 (1987).
80. R. L. Williams, S. A. Courtneidge, E. F. Wagner, *ibid.* **52**, 121 (1988); R. L. Williams *et al.*, *ibid.* **57**, 1053 (1989); E. F. Wagner, R. L. Williams, U. Ruther, in *Cell to Cell Signals in Mammalian Development*, S. W. deLaat *et al.*, Eds. (Springer-Verlag, Berlin, 1989), pp. 301–310.
81. J. Vogel, S. H. Hinrichs, R. K. Reynolds, P. A. Luciw, G. Jay, *Nature* **335**, 606 (1988).
82. S. H. Hinrichs, M. Nerenberg, R. K. Reynolds, G. Khoury, G. Jay, *Science* **237**, 1340 (1987); M. Nerenberg, S. H. Hinrichs, R. K. Reynolds, G. Khoury, G. Jay, *ibid.*, p. 1324.
83. M. Lacey, S. Alpert, D. Hanahan, *Nature* **322**, 609 (1986); M. Sippola-Thiele, D. Hanahan, P. M. Howley, *Mol. Cell. Biol.* **9**, 925 (1989).
84. I thank I. Herskowitz, T. DeFranco, and M. Kirschner for many thoughtful comments on the manuscript, and L. Spector for manuscript preparation.

Research Article

Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda

WILLIAM D. HUSE, LAKSHMI SASTRY, SHEILA A. IVERSON, ANGRAY S. KANG, MICHELLE ALTING-MEES, DENNIS R. BURTON, STEPHEN J. BENKOVIC, RICHARD A. LERNER

A novel bacteriophage lambda vector system was used to express in *Escherichia coli* a combinatorial library of Fab fragments of the mouse antibody repertoire. The system allows rapid and easy identification of monoclonal Fab fragments in a form suitable for genetic manipulation. It was possible to generate, in 2 weeks, large numbers of monoclonal Fab fragments against a transition state analog hapten. The methods described may supersede present-day hybridoma technology and facilitate the production of catalytic and other antibodies.

MONOCLONAL ANTIBODIES HAVE BEEN GENERATED THAT catalyze chemical transformations ranging from simple acyl transfer reactions to the energetically demanding hydrolysis of the peptide bond in the presence of metal cofactors (*1, 2–11*). Initially, it was widely held that antibodies would be most useful for catalysis where their predominant role was to overcome entropic barriers that occur along the reaction pathway. The basis of this hypothesis was that the chance occurrence of amino acid side chains capable of acid base catalysis in proximity to the reaction center was unlikely. However, for some reactions, study of the *p*H rate profile has revealed the participation of monobasic residues. Other studies have focused on placing appropriate charges on the antigen to induce specific binding interactions by complementary charged amino acid side chains on the antibody (*9, 12, 13*). Such functionalities might participate as a general acid, base, or nucleophile in the reaction under study.

Apart from the validity of the design of the mechanism based antigen, the probability of finding antibodies where particular amino acid side chains participate in catalysis also depends on the number of different antibodies assayed. Because current methods of generating monoclonal antibodies do not provide for an adequate survey of the available repertoire, we have been devising methods to clone the antibody repertoire in *Escherichia coli* and have described the preparation of a highly diverse immunoglobulin gene library (*14*). Given the difficulty of expressing both heavy and light chains together, we initially considered the construction and expression of libraries restricted to fragments of the variable region of the immunoglobulin (Ig) heavy chain V_H (*14*). In fact, a recent report describes the construction of a plasmid expression library in *E. coli* in which V_H fragments with affinity for keyhole limpet hemocyanin (KLH) and lysozyme have been isolated (*15*). However, the use of isolated V_H fragments as antibody mimics may be limited because (i) the available crystal structures of antibody-antigen complexes show considerable contact between antigen and V_L (light chain

W. D. Huse, L. Sastry, S. A. Iverson, and R. A. Lerner are with the Departments of Molecular Biology and Chemistry, Research Institute of Scripps Clinic, La Jolla, CA 92037. D. R. Burton and A. S. Kang are at the Departments of Molecular Biology and Chemistry, Research Institute of Scripps Clinic, La Jolla, CA 92037 and at the Krebs Institute, Department of Molecular Biology and Biotechnology, The University, Sheffield, United Kingdom. M. Alting-Mees is at Stratagene Inc., La Jolla, CA 92037. S. J. Benkovic is at the Department of Chemistry, Pennsylvania State University, University Park, PA 16802 and at the Departments of Molecular Biology and Chemistry, Research Institute of Scripps Clinic, La Jolla, CA 92037

variable) domain as well as V_H (*16*). More explicitly, in the case of a series of antibodies to dextran, the V_L domain provides contacts critical to antigen binding (*17*). Thus, it is unlikely that the affinity of isolated V_H fragments will generally match that of intact antibodies. (ii) The absence of the V_L domain leaves a large hydrophobic patch on one face of the V_H fragment, which will almost certainly lead to increased nonspecificity relative to whole antibodies (*15*). In contrast, Fab fragments (antigen binding) have been studied for more than 30 years. They behave as whole antibodies in terms of antigen recognition, and their affinity and specificity are well defined. Furthermore, for Fab, the combinatorial properties of heavy and light chains serve as an important source of diversity.

In that individual Fab molecules can be expressed and assembled in *E. coli* (*18*), the route to mimicking the diversity of the antibody system in vitro should lie in solving the problem of expressing the repertoires of heavy and light chains in combination. Accordingly, we used a novel system to enable the construction of bacteriophage lambda (λ) libraries expressing a population of functional antibody fragments (Fab's) with a potential diversity equal to or exceeding that of the parent animal.

Criteria for vector construction. To obtain a vector system for generating the largest number of Fab fragments that could be screened directly, we constructed the expression libraries in bacteriophage λ for the following reasons. First, in vitro packaging of phage DNA is the most efficient method of reintroducing DNA into host cells. Second, it is possible to detect protein expression at the level of single-phage plaques. Finally, in our experience, screens of phage libraries diminish the usual difficulties with nonspecific binding. The alternative, plasmid cloning vectors are only advantageous in the analysis of clones after they have been identified. This advantage is not lost in our system because we use λzap II and are able to excise a plasmid (*19*) containing the heavy chain, light chain, or Fab expressing inserts.

The vectors for expression of V_H, V_L, Fv (fragment of the variable region), and Fab sequences are diagrammed in Figs. 1 and 2. They were constructed by a modification of λzap II (*19*) in which we inserted synthetic oligonucleotides into the multiple cloning site. The vectors were designed to be antisymmetric with respect to the Not I and Eco RI restriction sites that flank the cloning and expression sequences. This antisymmetry in the placement of restriction sites in a linear vector such as bacteriophage allows a library expressing light chains to be combined with one expressing heavy chains in order to construct combinatorial Fab expression libraries. The vector λLc1 is designed to serve as a cloning vector for light

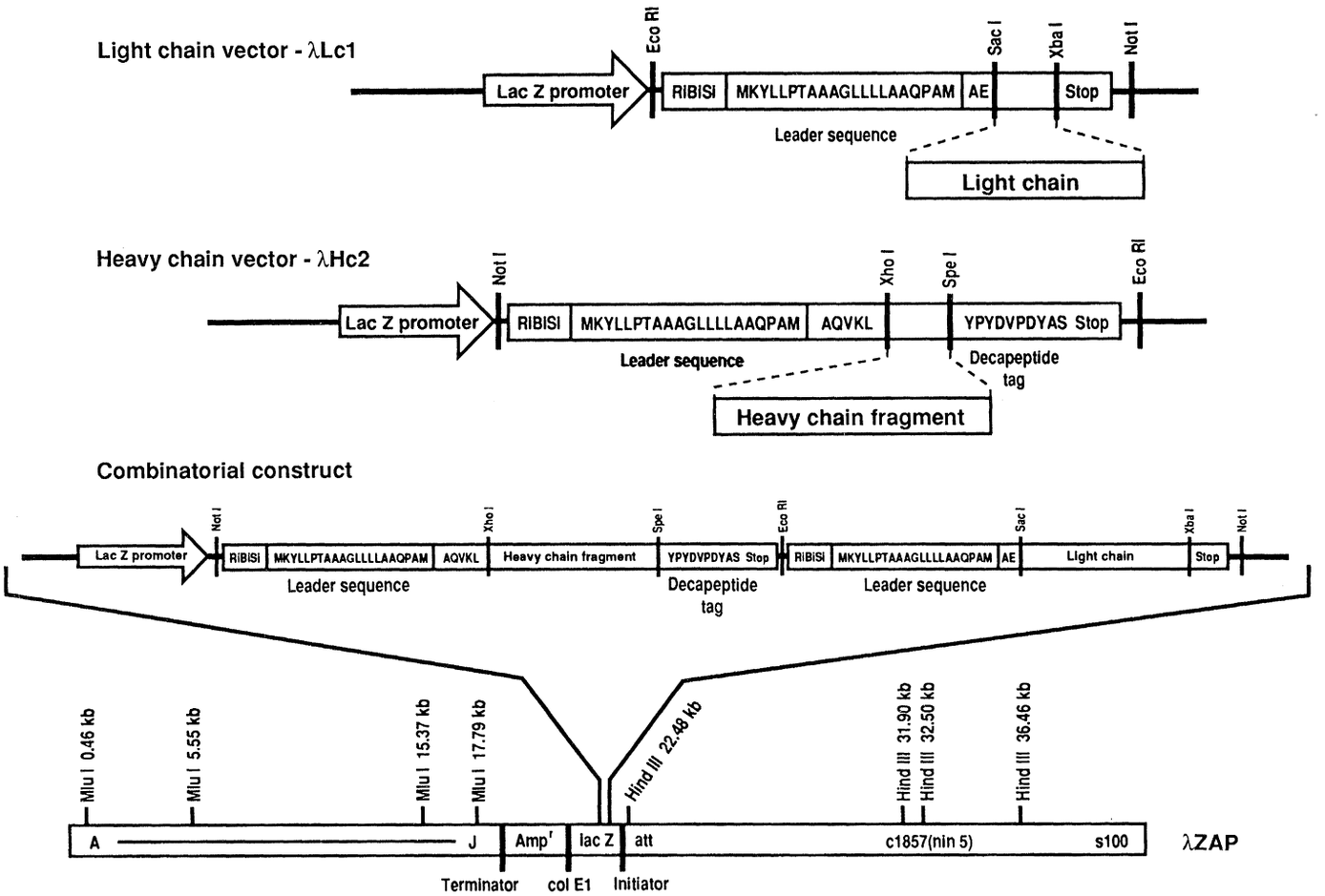


Fig. 1. Combinatorial bacteriophage λ vector system for expression of Fab antibody fragments. The λLc1 vector was constructed for the cloning of PCR amplified products of mRNA that code for light chain protein by inserting the nucleotide sequence shown depicted in Fig. 2A into the Sac I and Xho I sites of λzap II. The sequence was constructed from overlapping synthetic oligonucleotides varying in length from 25 to 50 nucleotides. The λHc2 vector was constructed for cloning PCR amplified products coding for heavy chain Fd sequences by inserting the nucleotide sequence (Fig. 2B) into the Not I and Xho I sites of λzap II. As with the light chain vector, the inserted sequence was constructed from overlapping synthetic oligonucleotides. The combinatorial constructs that can express Fab fragments are generated by cutting DNA isolated from light and heavy chain libraries at the antisymmetric Eco RI site of each vector, followed by re-ligation of the resulting arms. This generates constructs having random combination of light and heavy chains which can be expressed, upon induction with IPTG, from a dicistronic mRNA from the lac Z promoter.

chain fragments, and λHc2 is designed to serve as a cloning vector for heavy chain sequences in the initial step of library construction. These vectors are engineered to efficiently clone the products of PCR amplification with specific restriction sites incorporated at each end (*14, 15*). The sequence of the oligonucleotides used to construct these vectors include elements for construction, expression, and secretion of Fab fragments. These oligonucleotides introduce the antisymmetric Not I and Eco RI restriction sites; a leader peptide for the bacterial *pel* B gene, which has previously been successfully used in *E. coli* to secrete Fab fragments (*18*); a ribosome binding site at the optimal distance for expression of the cloned sequence; cloning sites for either the light or heavy chain PCR product; and, in λHc2, a decapeptide tag at the carboxyl terminus of the expressed heavy chain protein fragment. The sequence of the decapeptide tag was useful because of the availability of monoclonal antibodies to this peptide that were used for immunoaffinity purification of fusion proteins (*20*). The restriction endonuclease recognition sites included in the vectors were Sac I and Xba I in λLc1, and Xho I and Spe I in λHc2. The vectors were characterized by restriction digest analysis and DNA sequencing.

Choice of antigen and amplification of antibody fragments. We constructed the initial Fab expression library from mRNA isolated from a mouse that had been immunized with the KLH-coupled *p*-nitrophenyl phosphonamidate antigen **1** (NPN) (Fig. 3). This antigen was shown by Janda and co-workers (*7*) to be an effective one for the generation of catalytic antibodies. Also, the antibodies for the NPN reaction have been identified and therefore facilitate the implementation of assay systems. Finally, successful generation of catalytic antibodies generally requires binding to relatively small organic haptens, and it was necessary to test the suitability of our system for such molecules.

The PCR amplification of messenger RNA (mRNA) isolated from spleen cells or hybridomas with oligonucleotides that incorporate restriction sites into the ends of the amplified product can be used to clone and express heavy chain sequences (*14, 15*). This work is now extended to include the amplification of the Fd ($V_H - C_H1$)

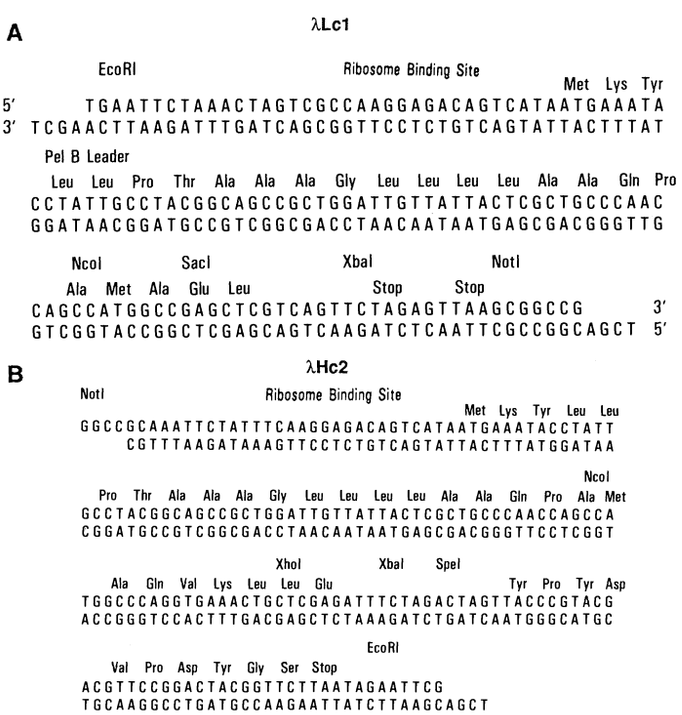


Fig. 2. (**A**) The nucleotide sequence inserted into λzap II to construct λLc1. (**B**) The nucleotide sequence inserted into λzap II to construct λHc2.

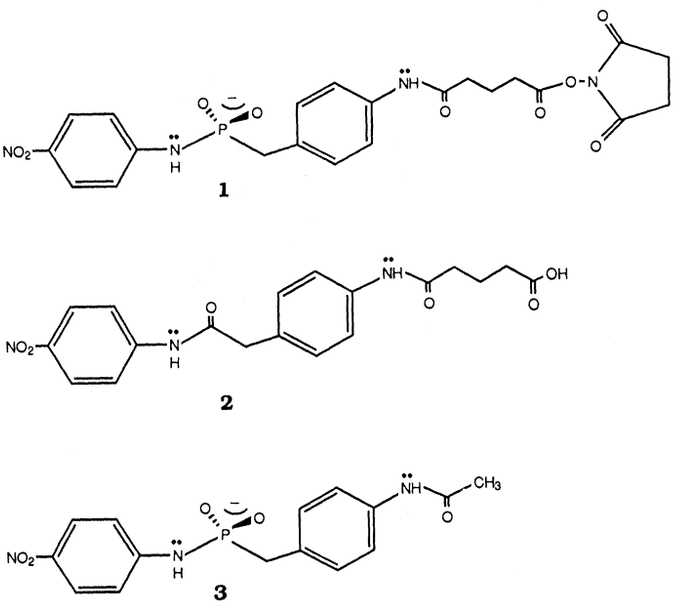


Fig. 3. The transition state analog **1**, which induces antibodies for hydrolyzing carboxamide substrate **2**. Compound **1** containing a glutaryl spacer and an *N*-hydroxysuccinimide–linker appendage is the form used to couple the hapten **1** to protein carriers KLH and BSA, while **3** is the inhibitor. The phosphonamidate functionality is a mimic of the stereoelectronic features of the transition state for hydrolysis of the amide bond.

and κ chain sequences (Fig. 4) from mouse spleen cells. The oligonucleotide primers used for these amplifications (Tables 1 and 2) are analogous to those that have been successfully used for amplification of V_H sequences (*14*). The set of 5′ primers for heavy chain amplification was identical to those used to amplify V_H, and those for light chain amplification were chosen similarly (*14, 21*). The 3′ primers of heavy (IgG1) and light (κ) chain sequences included the cysteines involved in disulfide bond formation between heavy and light chains. At this stage no primer was constructed to amplify light (λ) chains since they constitute only a small fraction of murine antibodies (*22*). Restriction endonuclease recognition sequences were incorporated into the primers to allow for the cloning of the amplified fragment into a λ phage vector in a predetermined reading frame for expression.

Library construction. We constructed a combinatorial library in two steps. In the first step, separate heavy and light chain libraries were constructed in λHc2 and λLc1, respectively (Fig. 1). In the second step, these two libraries were combined at the antisymmetric Eco RI sites present in each vector. This resulted in a library of clones each of which potentially coexpresses a heavy and a light chain. The actual combinations are random and do not necessarily reflect the combinations present in the B cell population in the parent animal. The λHc2 expression vector has been used to create a library of heavy chain sequences from DNA obtained by PCR amplification of mRNA isolated from the spleen of a 129 G_{IX}^+ mouse previously immunized with NPN conjugated to KLH. This primary library contains 1.3×10^6 plaque-forming units (pfu) and has been screened for the expression of the decapeptide tag to determine the percentage of clones expressing Fd sequences. The sequence for this peptide is only in frame for expression after the genes for an Fd (or V_H) fragment have been cloned into the vector. At least 80 percent of the clones in the library express Fd fragments when assayed by immunodetection of the decapeptide tag.

The light chain library was constructed in the same way as the heavy chain and shown to contain 2.5×10^6 members. Plaque screening, with an antibody to κ chain, indicated that 60 percent of

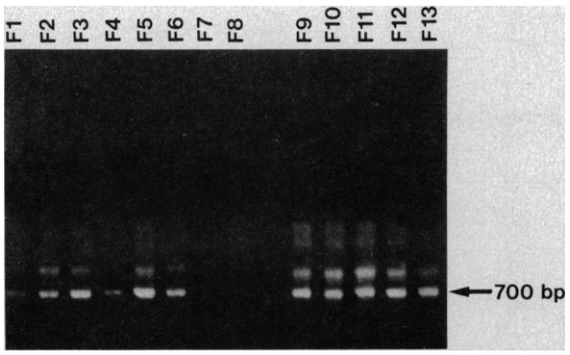


Fig. 4. PCR amplification of Fd and κ regions from the spleen mRNA of a mouse immunized with NPN. Amplification was performed as described (*14*) with RNA-cDNA hybrids obtained by the reverse transcription of the mRNA with primer specific for amplification of heavy chain sequences (12, Table 1) or light chain sequences (9, Table 2). Lanes F1 to F8 represent the product of heavy chain amplification reactions with one of each of the eight 5′ primers (primers 2 to 9, Table 1) and the 3′ primer (primer 12, Table 1). Light chain (κ) amplifications with the 5′ primers (primers 3 to 7, Table 2) and the appropriate 3′ primer (9, Table 2) are shown in lanes F9 through F13). A band of 700 base pairs is seen in all lanes indicating the successful amplification of Fd and κ regions.

HEAVY CHAIN PRIMERS

1) 5′-AGGT(CC/GA)(A/A)G CT(G/T) CTCGAGTC(T/A)GG-3′
2) 5′-AGGTCCAGCTGCTCGAGTCTGG-3′
3) 5′-AGGTCCAGCTGCTCGAGTCAGG-3′
4) 5′-AGGTCCAGCTTCTCGAGTCTGG-3′
5) 5′-AGGTCCAGCTTCTCGAGTCAGG-3′
6) 5′-AGGTCCAACTGCTCGAGTCTGG-3′
7) 5′-AGGTCCAACTGCTCGAGTCAGG-3′
8) 5′-AGGTCCAACTTCTCGAGTCTGG-3′
9) 5′-AGGTCCAACTTCTCGAGTCAGG-3′
10) 5′-AGGTIIAICTICTCGAGTC(T/A)GG-3′
11) 5′-CTATTAACTAGTAACGGTAACAGT-GGTGCCTTGCCCCA-3′
12) 5′- AGGCTTACTAGTACAATCCCTGG-GCACAAT-3′

Table 1. Primers used for amplification of heavy chain Fd fragments for construction of Fab's. Amplification was performed in eight separate reactions, each containing one of the 5′ primers (primers 2 to 9) and one of the 3′ primers (primer 12). The remaining 5′ primers that were used for amplification in a single reaction are either a degenerate primer (primer 1) or a primer that incorporates inosine at four degenerate positions (primer 10). The remaining 3′ primer (primer 11) has been used to construct Fv fragments. The underlined portion of the 5′ primers incorporates an Xho I site and that of the 3′ primer on Spe I restriction site.

Table 2. Primers used for amplification of κ light chain sequences for construction of Fab's. Amplification was performed in five separate reactions, each containing one of the 5′ primers (primers 3 to 7) and one of the 3′ primers (primer 9). The remaining 3′ primer (primer 8) has been used to construct Fv fragments. The underlined portion of the 5′ primers incorporate a Sac I restriction site and that of the 3′ primers an Xba I restriction site.

LIGHT CHAIN PRIMERS

1) 5′-CCAGTTCCGAGCTCGTTGTGACTCAGGAATCT-3′
2) 5′-CCAGTTCCGAGCTCGTGTTGACGCAGCCGCCC-3′
3) 5′-CCAGTTCCGAGCTCGTGCTCACCCAGTCTCCA-3′
4) 5′-CCAGTTCCGAGCTCCAGATGACCCAGTCTCCA-3′
5) 5′-CCAGATGTGAGCTCGTGATGACCCAGACTCCA-3′
6) 5′-CCAGATGTGAGCTCGTCATGACCCAGTCTCCA-3′
7) 5′-CCAGTTCCGAGCTCGTGATGACACAGTCTCCA-3′
8) 5′-GCAGCATTCTAGAGTTTCAGCTCCAGCTTGCC-3′
9) 5′-GCGCCGTCTAGAATTAACACTCATTCCTGTTGAA-3′

the library contained expressed light chain inserts. This relatively small percentage of inserts probably resulted from incomplete dephosphorylation of the vector after cleavage with Sac I and Xba I.

Once obtained, the two libraries were used to construct a combinatorial library by crossing them at the Eco RI site as follows. DNA was first purified from each library. The light chain library was cleaved with Mlu I restriction endonuclease, the resulting 5′ ends were dephosphorylated, and the product was digested with Eco RI. This process cleaved the left arm of the vector into several pieces, but the right arm containing the light chain sequences remained intact. The DNA of heavy chain library was cleaved with Hind III, dephosphorylated, and then cleaved with Eco RI; this process destroyed the right arm, but the left arm containing the heavy chain sequences remained intact. The DNA's so prepared were then mixed and ligated. After ligation, only clones that resulted from combination of a right arm of light chain–containing clones and a left arm of heavy chain–containing clones reconstituted a viable phage. After ligation and packaging, 2.5×10^7 clones were obtained. This is the combinatorial Fab expression library that was screened to identify clones having affinity for NPN. For determining the frequency of the phage clones that coexpress the light and heavy chain fragments, we screened duplicate lifts of the combinatorial library for light and heavy chain expression. In our examination of approximately 500 recombinant phage, approximately 60 percent coexpressed light and heavy chain proteins.

Antigen binding. All three libraries, the light chain, the heavy chain, and Fab were screened to determine whether they contained recombinant phage that expressed antibody fragments binding NPN. In a typical procedure, 30,000 phage were plated and duplicate lifts with nitrocellulose screened for binding to NPN coupled to ^{125}I-labeled bovine serum albumin (BSA) (Fig. 5). Duplicate screens of 90,000 recombinant phage from the light chain library and a similar number from the heavy chain library did not identify any clones that bound the antigen. In contrast, the screen of a similar number of clones from the Fab expression library identified many phage plaques that bound NPN (Fig. 5). This observation indicates that, under conditions where many heavy chains in combination with light chains bind to antigen, heavy or light chains alone do not. Therefore, in the case of NPN, we expect that there are many heavy and light chains that only bind antigen when they are combined with specific light and heavy chains, respectively. This result supports our decision to screen large combinatorial Fab expression libraries. To assess our ability to screen large numbers of clones and obtain a more quantitative estimate of the frequency of antigen binding clones in the combinatorial library, we screened one million phage plaques and identified approximately 100 clones that bound to antigen. For six clones, a region of the plate containing the positive phage plaques and approximately 20 surrounding them was "cored," replated, and screened with duplicate lifts (Fig. 5). As expected, the expression products of approximately 1 in 20 of the phage specifically bind to antigen. Phage which were believed to be negative on the initial screen did not give positives on replating.

To determine the specificity of the antigen-antibody interaction, antigen-binding was subjected to competition with free unlabeled antigen (Fig. 6). These studies showed that individual clones could be distinguished on the basis of antigen affinity. The concentration of free haptens required for complete inhibition of binding varied between 10 to $100 \times 10^{-9} M$, suggesting that the expressed Fab fragments had binding constants in the nanomolar range.

In preparation for characterization of the protein products, a plasmid containing the heavy and light chain genes was excised with helper phage (Fig. 7). Mapping of the excised plasmid demonstrated a restriction pattern consistent with incorporation of heavy and light chain sequences. The protein products of one of the clones was

analyzed by enzyme-linked immunosorbent assay (ELISA) and immunoblotting to establish the composition of the NPN binding protein. A bacterial supernatant after IPTG (isopropyl thiogalactoside) induction was concentrated and subjected to gel filtration. Fractions in the molecular size range 40 to 60 kD were pooled, concentrated, and subjected to a further gel filtration separation. ELISA analysis of the eluted fractions (Fig. 8) indicated that NPN binding was associated with a protein of a molecular size of about 50 kD, which contained both heavy and light chains. An immunoblot of a concentrated bacterial supernatant preparation under nonreducing conditions was developed with antibody to decapeptide. This revealed a 50-kD protein band. We have found that the antigen-binding protein can be purified to homogeneity from bacterial supernate in two steps involving affinity chromatography on protein G followed by gel filtration. SDS-PAGE analysis of the protein revealed a single band at ~50 kD under nonreducing conditions and a doublet at ~25 kD under reducing conditions. Taken together, these results are consistent with NPN-binding being a function of Fab fragments in which heavy and light chains are covalently linked by a disulfide bond.

Properties of the in vivo repertoire compared to the phage combinatorial library. Previously we constructed a highly diverse V_H library in *E. coli*. We have now combined heavy and light chain libraries to clone and express assembled and functional Fab fragments of immunoglobulin. A moderately restricted library was prepared because only a limited number of primers was used for polymerase chain reaction (PCR) amplification of Fd sequences. The library is expected to contain only clones expressing κ-γ1 sequences. However, this is not an inherent limitation of the method since the addition of more primers can amplify any antibody class or subclass. Despite this restriction we were able to isolate a large number of clones producing antigen binding proteins.

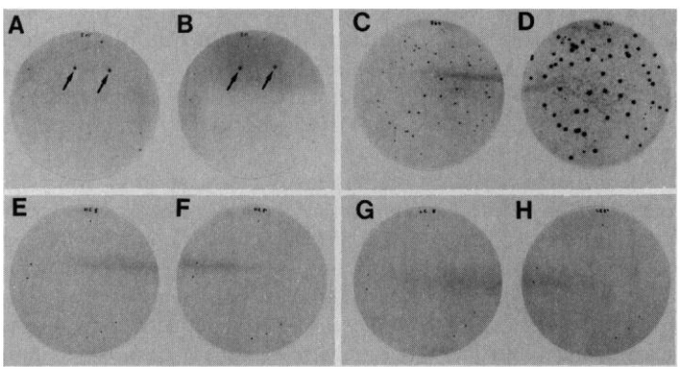


Fig. 5. Screening phage libraries for antigen-binding. Duplicate plaque lifts of Fab (filters **A** and **B**), heavy chain (filters **E** and **F**), and light chain (filters **G** and **H**) expression libraries were screened against ^{125}I-labeled BSA conjugated with NPN at a density of approximately 30,000 plaques per plate. Filters **C** and **D** illustrate the duplicate secondary screening of a cored positive from a primary filter A (arrows) as discussed in the text. Standard plaque lift methods were used in screening. Cells (XL1 blue) infected with phage were incubated on 150-mm plates for 4 hours at 37°C, protein expression was induced by overlay with nitrocellulose filters soaked in 1 m*M* IPTG, and the plates were incubated at 25°C for 8 hours. Duplicate filters were obtained during a second incubation under the same conditions. Filters were then blocked in a solution of 1 percent BSA in phosphate-buffered saline (PBS) for 1 hour before incubation (with rocking) at 25°C for 1 hour with a solution of ^{125}I-labeled BSA (at 0.1 μ*M*) conjugated to NPN (2×10^6 cpm/ml; approximately 15 NPN per BSA molecule), in 1 percent BSA in PBS. Background was reduced by preliminary centrifugation of stock ^{125}I-labeled BSA solution at 100,000*g* for 15 minutes and preliminary incubation of solutions with plaque lifts from plates containing bacteria infected with a phage having no insert. After labeling, filters were washed repeatedly with PBS containing 0.05 percent Tween 20 before the overnight development of autoradiographs.

A central issue is how our phage library compares with the in vivo antibody repertoire in terms of size, characteristics of diversity, and ease of access.

The size of the mammalian antibody repertoire is difficult to judge, but a figure of the order of 10^6 to 10^8 different antigen specificities is often quoted. With some of the reservations discussed below, a phage library of this size or larger can readily be constructed by a modification of the method described. Once an initial combinatorial library has been constructed, heavy and light chains can be shuffled to obtain libraries of exceptionally large numbers.

In principle, the diversity characteristics of the naive (unimmunized) in vivo repertoire and corresponding phage library are expected to be similar in that both involve a random combination of heavy and light chains. However, different factors act to restrict the diversity expressed by an in vivo repertoire and phage library. For example, a physiological modification such as tolerance will restrict the expression of certain antigenic specificities from the in vivo repertoire, but these specificities may still appear in the phage library. However, bias in the cloning process may introduce restrictions into the diversity of the phage library. For example, the representation of mRNA for sequences expressed by stimulated B cells can be expected to predominate over those of unstimulated cells because of higher levels of expression. In addition, the resting repertoire might overrepresent spontaneously activated B cells whose immunoglobulins have been suggested to be less specific. In any event, methods exist to selectively exclude such populations of cells. Also, the fortuitous presence of restriction sites in the variable gene similar to those used for cloning and combination will cause them to be eliminated. We can circumvent some of these difficulties by making minor changes, such as introducing amber mutations in the vector system. Different source tissues (for example, peripheral blood, bone marrow, or regional lymph nodes) and different PCR primers (for example, those to amplify different antibody classes), may result in libraries with different diversity characteristics.

Another difference between in vivo repertoire and phage library is

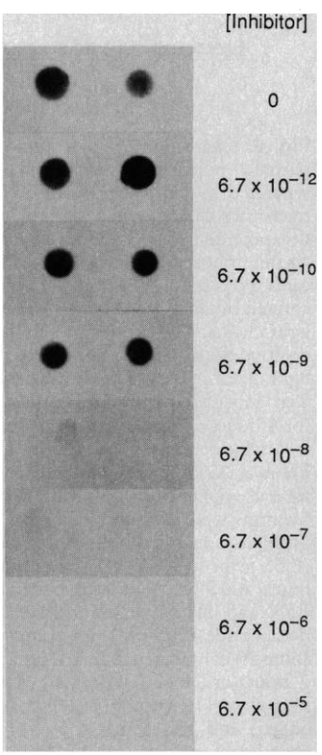


Fig. 6. Specificity of antigen binding shown by competitive inhibition. Filter lifts from positive plaques were exposed to ^{125}I-labeled BSA-NPN in the presence of increasing concentrations of the inhibitor NPN. A number of phages correlated with NPN-binding as in Fig. 5 were spotted in duplicate (about 100 particles per spot) directly onto a bacterial lawn. The plate was then overlaid with an IPTG-soaked filter and incubated for 19 hours at 25°C. The filters were then blocked in 1 percent BSA in PBS before incubation in ^{125}I-BSA–NPN as done previously with the inclusion of varying amounts of NPN in the labeling solution. Other conditions and procedures were as in Fig. 5. The results for a phage of moderate affinity are shown in duplicate in the figure. Similar results were obtained for four other phages with some differences in the effective inhibitor concentration ranges.

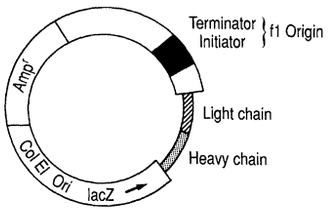


Fig. 7. A plasmid can be excised from λLc1, λHc2, and their combination because they are a modification of λzap II. M13mp8 was used as helper phage and the excised plasmid was infected into a F^+ derivative of MC1061. The excised plasmid contains the same constructs for antibody fragment expression as do the parent vectors (Fig. 1). These plasmid constructs are more conveniently analyzed for restriction pattern and protein expression of the λ phage clones identified and isolated on the basis of antigen binding. The plasmid also contains an f1 origin of replication which facilitates the preparation of single-stranded DNA for sequence analysis and in vitro mutagenesis.

that antibodies isolated from the repertoire may have benefited from affinity maturation as a result of somatic mutations after combination of heavy and light chains whereas the phage library randomly combines the matured heavy and light chains. Given a large enough phage library derived from a particular in vivo repertoire, the original matured heavy and light chains will be recombined. However, since one of the potential benefits of this technology is to obviate the need for immunization by the generation of a single highly diverse "generic" phage library, it would be useful to have

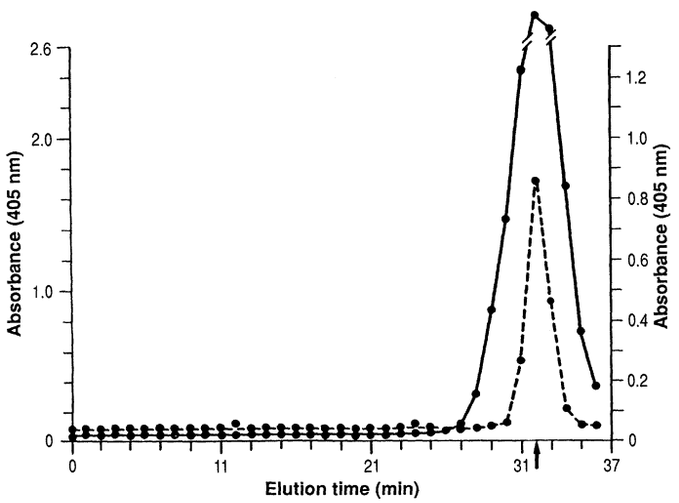


Fig. 8. Characterization of an antigen binding protein. The concentrated partially purified bacterial supernatant of an NPN binding clone was separated by gel filtration and samples from each fraction were applied to microtiter plates coated with BSA-NPN. Addition of either antibody to decapeptide (---) or antibody to κ chain (—, left-hand scale) conjugated with alkaline phosphatase was followed by color development. The arrow indicates the position of elution of a known Fab fragment. The results show that antigen binding is a property of a 50-kD protein containing both heavy and light chains. To permit protein characterization, a single plaque of a NPN-positive clone (Fig. 5) was picked, and the plasmid containing the heavy and light chain inserts (Fig. 7) was excised (*19*). Cultures (500 ml) in L broth were inoculated with 3 ml of a saturated culture of the clone and incubated for 4 hours at 37°C. Protein synthesis was induced by the addition of IPTG to a final concentration of 1 m*M*, and the cultures were incubated for 10 hours at 25°C. The supernatant from 200 ml of cells was concentrated to 2 ml and applied to a TSK-G4000 column. Samples (50 μl) from the eluted fractions were assayed by ELISA. Microtiter plates were coated with BSA-NPN at 1 μg/ml, 50-μl samples were mixed with 50 μl of PBS-Tween 20 (0.05 percent) BSA (0.1 percent) added, and the plates were incubated for 2 hours at 25°C. The plated material was then washed with PBS-Tween 20-BSA and 50 μl of appropriate concentrations of a rabbit antibody to decapeptide (*20*) or a goat antibody to mouse κ light chain (Southern Biotech) conjugated with alkaline phosphatase were added and incubated for 2 hours at 25°C. The plates were again washed, 50 μl of *p*-nitrophenyl phosphate (1 mg/ml in 0.1*M* tris, *p*H 9.5, containing 50 m*M* $MgCl_2$) was added, and the plates were incubated for 15 to 30 minutes and the absorbance was read at 405 nm.

methods to optimize sequences to compensate for the absence of somatic mutation and clonal selection. Three procedures are made readily available through the vector system presented. First, saturation mutagenesis may be performed on the complementarity-determining regions (CDR's) (*23*) and the resulting Fab's can be assayed for increased function. Second, a heavy or a light chain of a clone that binds antigen can be recombined with the entire light or heavy chain libraries, respectively, in a procedure identical to that used to construct the combinatorial library. Third, iterative cycles of the two above procedures can be performed to further optimize the affinity or catalytic properties of the immunoglobulin. The last two procedures are not permitted in B cell clonal selection, which suggests that the methods described here may actually increase our ability to identify optimal sequences.

Access is the third area where it is of interest to compare the in vivo antibody repertoire and phage library. In practical terms the phage library is much easier to access. The screening methods used have allowed one to survey the gene products of at least 50,000 clones per plate so that 10^6 to 10^7 antibodies can be readily examined in a day but the most powerful screening methods depend on selection. In the catalytic antibody system, this may be accomplished by incorporating into the antigen leaving groups necessary for replication of auxotrophic bacterial strains or toxic substituents susceptible to catalytic inactivation. Further advantages are related to the fact that the in vivo antibody repertoire can only be accessed via immunization, which is a selection on the basis of binding affinity. The phage library is not similarly restricted. For example, the only general method to identify antibodies with catalytic properties has been by preselection on the basis of affinity of the antibody to a transition state analog. Such restrictions do not apply to the in vitro library where catalysis can, in principle, be assayed directly. The ability to assay directly large numbers of antibodies for function may allow selection for catalysts in reactions where a mechanism is not well defined or synthesis of the transition state analog is difficult. Assaying for catalysis directly eliminates the bias of the screening procedure for reaction mechanisms limited to a particular synthetic analog; therefore, simultaneous exploration of multiple reaction pathways for a given chemical transformation are possible.

We have described procedures for the generation of Fab fragments that are clearly different in a number of important respects from antibodies. There is undoubtedly a loss of affinity in having monovalent Fab antigen binders, but it is possible to compensate for this by selection of suitably tight binders. For a number of applications such as diagnostics and biosensors, monovalent Fab fragments may be preferable. For applications requiring Fc effector functions, the technology already exists for extending the heavy chain gene and expressing the glycosylated whole antibody in mammalian cells.

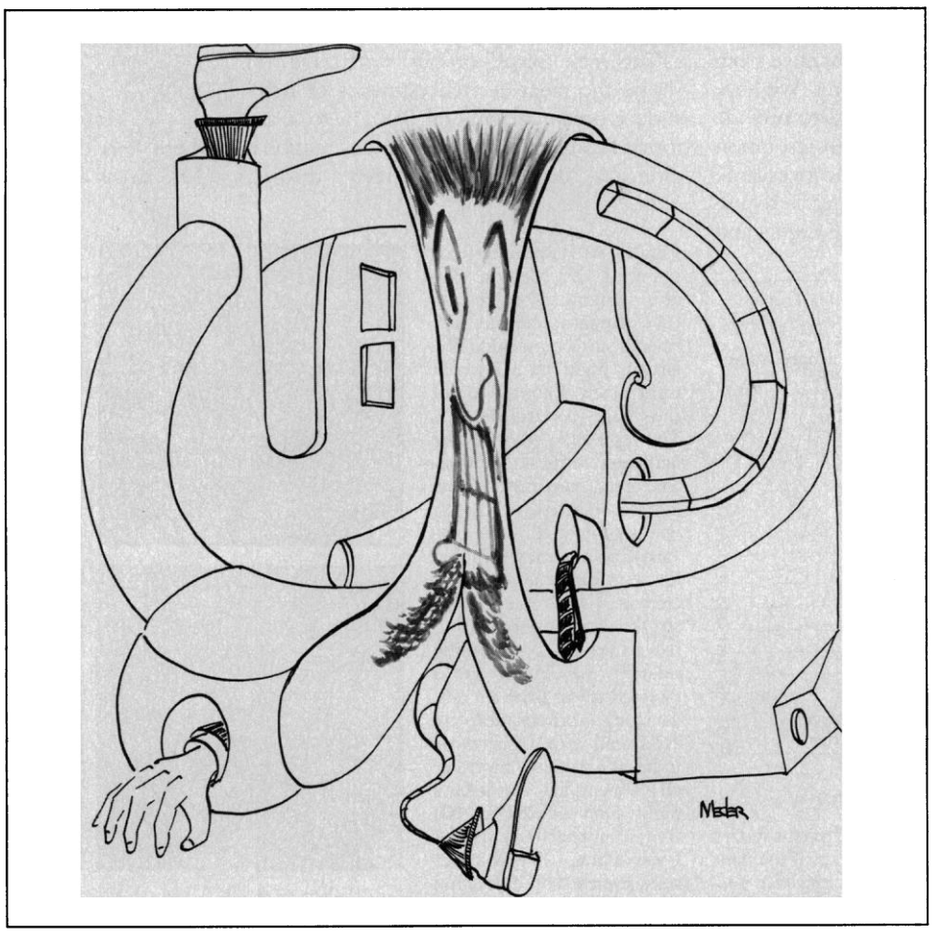
Our data show that it is now possible to construct and screen at least three orders of magnitude more clones with monospecificity than previously possible. The data also invite speculation concerning the production of antibodies without the use of live animals.

REFERENCES AND NOTES

1. R. A. Lerner, in *Proceedings of the XVIIIth Solvay Conference* on Chemistry, G. Van Binst, Ed. (Springer-Verlag, Berlin, 1983), pp. 43–49.
2. A. Tramontano, K. D. Janda, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6736, 1986; *Science* **234**, 1566 (1986); A. Tramontano, A. A. Ammann, R. A. Lerner, *J. Am. Chem. Soc.* **110**, 2282 (1988); S. J. Pollack and P. G. Schultz, *ibid.* **111**, 1929 (1989); S. J. Pollack, P. Hsiun, P. G. Schultz, *ibid.*, p. 5961.
3. S. J. Pollack, J. W. Jacobs, P. G. Schultz, *Science* **234**, 1570 (1986); J. Jacobs, P. G. Schultz, R. Sugasawara, M. J. Powell, *J. Am. Chem. Soc.* **109**, 2174 (1987).
4. S. J. Benkovic, A. D. Napper, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5355 (1988); K. D. Janda, R. A. Lerner, A. Tramontano, *J. Am. Chem. Soc.* **110**, 4835 (1988); K. D. Janda, S. J. Benkovic, R. A. Lerner, *Science* **244**, 437 (1989).
5. A. D. Napper, S. J. Benkovic, A. Tramontano, R. A. Lerner, *Science* **237**, 1041 (1987).

6. D. Hilvert, S. H. Carpenter, K. D. Nared, M.-T. M. Auditor, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4953 (1988); D. Y. Jackson *et al.*, *J. Am. Chem. Soc.* **110**, 4841 (1988).
7. K. D. Janda, D. Schloeder, S. J. Benkovic, R. A. Lerner, *Science* **241**, 1188 (1988). We thank K. D. Janda for providing hapten 1 and inhibitor 3.
8. A. Cochran, R. Sugasawara, P. G. Schultz, *J. Am. Chem. Soc.* **110**, 7888 (1988).
9. K. M. Shokat *et al.*, *Nature* **338**, 269 (1989).
10. B. L. Iverson and R. A. Lerner, *Science* **243**, 1184 (1989).
11. D. Hilvert *et al.*, *J. Am. Chem. Soc.*, in press.
12. R. A. Lerner and S. J. Benkovic, *Bioassays* **9**, 107 (1988); K. D. Janda, M. I. Weinhouse, D. M. Schloeder, R. A. Lerner, S. J. Benkovic, unpublished results.
13. B. L. Iverson, K. Cameron, K. Jahangiri, D. Pasternak, unpublished results.
14. L. Sastry *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5728 (1989).
15. E. S. Ward, D. Gussow, A. D. Griffiths, P. T. Jones, G. Winter, *Nature* **341**, 544 (1989); R. Orlandi, D. H. Gussow, P. T. Jones, G. Winter, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3833 (1989).
16. D. R. Davies, S. Sheriff, E. A. Padlan, *J. Biol. Chem.* **263**, 10541 (1988); J. N. Herron, X. He, M. L. Mason, E. W. Voss, Jr., A. B. Edmundson, *Proteins Struct. Funct. Genet.* **5**, 271 (1989); M. Whitlow, A. J. Howard, J. F. Wood, E. W. Voss, Jr., K. D. Hardman, unpublished results. In the structure of Whitlow *et al.*, the light chain in the fluorescein isothiocyanate conjugated Fab 4-4-20 crystal structure has major contributions to the binding of the hapten fluorescein. All of the hydrogen bonds to fluorescein are from the light chain. Three come from the light chain CDR 1 (residues His^{27D}, Tyr^{32}, and Arg^{34}, in Kabat's numbering system) and one from the light chain CDR 3 (residue Ser^{91}). Fluorescein is bound in a deep pocket in the Fab 4-4-40 structure, so that only 10 percent (54 $Å^2$ of 521 $Å^2$) of its solvent accessible surface area is exposed to solvent; 38 percent of this pocket is lined with light chain residues, based on calculations of solvent-accessible surface area. Aromatic residues His^{27D}, Tyr^{32}, and Trp^{96} from the light chain and Trp^{33}, Tyr^{53}, Tyr^{100D}, and Tyr^{100E} from the heavy chain form this deep pocket for the fluorescein.
17. S. C. Wallick, E. A. Kabat, S. L. Morrison, *Immunology* **142**, 1235 (1989).
18. M. Better, C. P. Chang, R. R. Robinson, A. H. Horwitz, *Science* **240**, 1041 (1988); A. Skerra and A. Plückthun, *ibid.*, p. 1038.
19. J. M. Short, J. M. Fernandez, J. A. Sorge, W. D. Huse, *Nucleic Acids Res.* **16**, 7583 (1988).
20. J. Field *et al.*, *Mol. Cell. Biol.* **8**, 2159 (1988).
21. R. Orlandi, D. H. Gussow, P. T. Jones, G. Winter, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3833 (1989).
22. The Fv (variable region) fragments may be constructed with a 3′ primer that is complementary to the mRNA in the J (joining) region (amino acid 128) and a set of 5′ primers that are complementary to the first strand cDNA in the conserved amino-terminal region of the processed protein.
23. Amino acid and nucleotide sequences of immunoglobulins used in this paper were taken from E. A. Kabat, T. T. Wu, M. Reid-Miller, H. M. Perry, K. S. Gottesman, *Sequences of Proteins of Immunological Interest* (US Public Health Service, National Institutes of Health, Bethesda, MD, 1987).
24. We thank D. Schloeder, D. A. McLeod, R. Samodal, B. Hay for technical assistance; E. A. Kabat, N. Klinman, K. D. Janda, and B. L. Iverson for advice and comments. Supported by an NIH postdoctoral fellowship (S.A.I), the SERC Protein Engineering Club (A.K.), and a Jenner Fellowship of the Lister Institute of Preventive Medicine (D.R.B.).

26 October 1989; accepted 14 November 1989



" Convinced of the correctness of his 17-dimensional model of the universe, theorist Martin Nowsky was not above testing it directly."