LETTERS

nature biotechnology

Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells

Sai T Reddy[1,2], Xin Ge[1], Aleksandr E Miklos[3,4], Randall A Hughes[3,4], Seung Hyun Kang[1], Kam Hon Hoi[2], Constantine Chrysostomou[1], Scott P Hunicke-Smith[3], Brent L Iverson[3,5], Philip W Tucker[3,6], Andrew D Ellington[3–5] & George Georgiou[1–3,6]

© 2010 Nature America, Inc. All rights reserved.

Isolation of antigen-specific monoclonal antibodies (mAbs) and antibody fragments relies on high-throughput screening of immortalized B cells[1,2] or recombinant antibody libraries[3–6]. We bypassed the screening step by using high-throughput DNA sequencing and bioinformatic analysis to mine antibody variable region (V)-gene repertoires from bone marrow plasma cells (BMPC) of immunized mice. BMPCs, which cannot be immortalized, produce the vast majority of circulating antibodies. We found that the V-gene repertoire of BMPCs becomes highly polarized after immunization, with the most abundant sequences represented at frequencies between ~1% and >10% of the total repertoire. We paired the most abundant variable heavy (V_H) and variable light (V_L) genes based on their relative frequencies, reconstructed them using automated gene synthesis, and expressed recombinant antibodies in bacteria or mammalian cells. Antibodies generated in this manner from six mice, each immunized with one of three antigens were overwhelmingly antigen specific (21/27 or 78%). Those generated from a mouse with high serum titers had nanomolar binding affinities.

The ability of the mammalian humoral immune response to generate a vastly diverse antibody repertoire in response to an antigen has been exploited for a range of biotechnology applications in diagnostics, therapy and basic research[7]. Since the development of the hybridoma technology by Kohler and Milstein 35 years ago[1], several methods for the generation of mAbs have been developed. Such methods include B-cell immortalization through genetic reprogramming by means of Epstein-Barr virus[8] or retrovirus-mediated gene transfer[2], cloning of V genes by single-cell PCR[9,10] and approaches for *in vitro* discovery that involve the display and screening of recombinant antibody libraries[3–6,11]. Both *in vitro* and *in vivo* methods for antibody discovery are critically dependent on high-throughput screening to determine antigen specificity. Recently, B-cell analysis has been expedited by soft lithography and microengraving techniques that allow for high-throughput identification of antigen-specific B cells[12,13]. However, this is at the cost of considerable technical complexity due to the need to amplify V genes and expand B cells. Similarly, the success of *in vitro* antibody discovery techniques depends on a range of screening parameters, which include the nature of the display platform, antigen concentration, binding avidity during enrichment, the number of rounds of screening (by panning or sorting), and the design and diversity of synthetic antibody libraries[7,14,15].

We have developed a simple and rapid method for antibody isolation without the need for screening. We exploited high-throughput DNA sequencing to analyze the V_L and V_H gene repertoires derived from the mRNA transcripts of fully differentiated mature B cells, antibody-secreting BMPCs, from immunized mice. After bioinformatic analysis, several abundant V_L and V_H gene sequences could be identified within the repertoire of each immunized mouse. V_L and V_H genes were paired according to their relative frequencies within the repertoire. Antibody genes were rapidly synthesized by oligonucleotide and PCR assembly by automated liquid-handling robots. Recombinant antibodies were expressed in bacterial and mammalian systems as single-chain variable fragments (scFv) and full-length IgG, respectively (**Fig. 1**). Finally, we confirmed that the resulting antibodies were overwhelmingly antigen specific (21/27 or 78%), thus confirming that our approach enables rapid and direct isolation of mAbs without screening.

B-cell maturation terminates with the formation of plasma cells, which represent <1% of all lymphoid cells but are responsible for the overwhelming majority of antibodies in circulation[16,17]. The bone marrow constitutes the major compartment where plasma cells reside and produce antibodies for prolonged periods of time, whereas plasma cells present in secondary lymphoid organs are often short lived. In mice, a stable and highly enriched antigen-specific BMPC population of ~10^5 cells (10–20% of all BMPCs) appears 6 d after secondary immunization and persists for prolonged periods[18]. In contrast, the increase in size of the splenic plasma cell population is highly transient, peaking at day 6 and rapidly declining to <10^4 cells by day 11. Notably, BMPCs are long lived and thus responsible for making the stable circulating population of antibodies in serum, which in turn is likely to play a dominant role in pathogen neutralization and other protective humoral immune responses[16].

To examine the dynamics of the V-gene repertoires in BMPCs, especially early after challenge with antigens, we immunized pairs

[1]Department of Chemical Engineering, University of Texas at Austin, Austin, Texas, USA. [2]Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas, USA. [3]Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, USA. [4]Applied Research Laboratories, University of Texas at Austin, Austin, Texas, USA. [5]Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas, USA. [6]Section of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, Texas, USA. Correspondence should be addressed to G.G. (gg@che.utexas.edu).

Received 11 June; accepted 2 August; published online 29 August 2010; doi:10.1038/nbt.1673

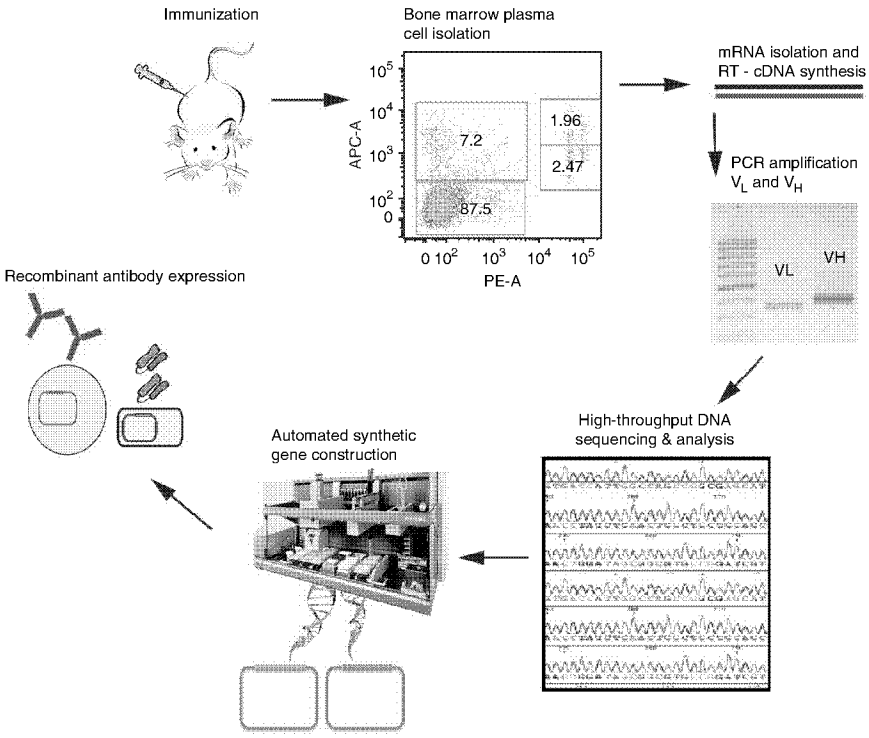

Figure 1 Isolation of monoclonal antibodies by mining the antibody variable (V)-gene repertoires of bone marrow plasma cells. Immunized mice are euthanized and CD45R$^-$ CD138$^+$ plasma cells are isolated. After mRNA isolation and first-strand cDNA synthesis, variable light (V_L) and variable heavy (V_H) gene DNA is generated. High-throughput 454 DNA sequencing and bioinformatic analysis is performed to determine the V_L and V_H repertoire. The most abundant V_L and V_H genes are identified and the sequences paired by using a simple relative-frequency rule. The respective antibody genes are synthesized using automated, robotically assisted gene synthesis. Finally, antigen-specific antibody single chain variable fragments or full-length IgGs are expressed in bacteria or mammalian cells, respectively. APC-A, CD45R-allophycocyanin-area; PE-A, CD138-R-phycoerythrin-area.

of mice with chicken egg ovalbumin, human complement serine protease (C1s), human B-cell regulator of IgH transcription (Bright) or adjuvant only. Antigen was co-injected with complete Freund's adjuvant followed by a secondary booster immunization in incomplete Freund's adjuvant. Mice were euthanized 6 d after secondary immunization and BMPCs (CD45R$^-$ CD138$^+$) were isolated to high purity (**Supplementary Fig. 1**). Total RNA was extracted and reverse transcribed for synthesis of first-strand cDNA. Well-characterized[11], degenerate V-gene primer mixes were used for second-strand amplifications, resulting in V_L and V_H PCR products of high purity (**Supplementary Fig. 2**), which were then submitted for high-throughput DNA sequencing of long reads using the Roche 454-GS FLX technology.

Unlike recent high-throughput sequencing analyses that explored V-gene repertoire diversity in zebrafish[19], humans[20,21] or synthetic libraries[22], our goals were to (i) identify highly expressed V genes whose products were likely to be antigen specific and (ii) determine the relative V-gene transcript abundance in the BMPC repertoires of immunized mice. These two tasks do not require exhaustive coverage of the V-gene repertoire; we have found that obtaining ~5,000 V-gene sequences per BMPC sample is sufficient to provide the information needed for antibody discovery, thus minimizing DNA sequencing costs. 454 reads were first processed by multiple sequence and signal filters, and then subjected to a simple and rapid bioinformatic analysis that relied on homologies to conserved framework regions within V genes to identify the most common complementarity determining region 3 (CDR3) sequences (**Supplementary Fig. 3**). This approach correctly identified ~94% of V_H and ~92% of V_L sequences in the Kabat database (**Supplementary Table 1**). Of a total of 415,018 reads, 23.2% contained CDR3 of V_H (CDRH3) and 26.6% contained CDR3 of V_L (CDRL3) sequences (**Supplementary Table 2**), representing 6,681–16,743 and 7,112–21,241 CDRH3 and CDRL3 sequences reads per mouse, respectively. For each mouse, frequency distributions of the CDR3s were calculated. Sequencing of the same samples, from separate cDNA library preparations by different facilities, gave quantitatively similar rankings for the abundances of CDR3 sequences. The same rank order frequencies are observed for all of the highly expressed CDR3s (**Supplementary Table 3**). This is important, because as discussed below, our approach for antibody discovery exploits the rank-order frequency of V genes to identify the most highly expressed clones. V-gene sequences containing a particular CDR3 were accepted as full length if they covered all three CDRs. Pairwise identities and frequencies were calculated by multiple sequence alignments, followed by germline analysis (**Supplementary Fig. 3** and Online Methods). A graphical user interface application was developed to enhance data analysis and visualization of the results (**Supplementary Fig. 4** and **Supplementary Data**).

Analysis of the BMPC repertoires from the six mice each immunized with one of the three antigens led to several interesting observations. First, in all immunized mice, including those receiving the same antigen, >92% of the CDRH3 sequences were unique to an individual mouse. The CDRL3 repertoires were less diverse, and in some instances, BMPCs from mice immunized with different antigens expressed high levels of the same CDRL3 (data not shown). A lower degree of V_L diversity, especially in early responses (as was the case here), is consistent with CDRL3 being derived from a single-gene recombination event (V-J), as opposed to two recombination events (V-D-J) for CDRH3. Second, and most importantly, ~10–20% of the total repertoire of all immunized mice were on average composed of only four CDRH3 sequences (**Supplementary Table 4**). For example, in the two mice immunized with C1s, the frequencies of the most abundant CDRH3s were 7.93% and 10.99% of the total repertoire. Third, as expected for early responses, the most highly abundant CDR3s were assembled from a diverse array of germline V-gene segments, with an average somatic mutation rate of only two and five amino acid substitutions for V_L and V_H, respectively (**Supplementary Tables 5** and **6**). Not surprisingly, certain germline V-gene families were represented preferentially in mice responding to particular antigens. For example, in mice immunized with C1s, 17.2% and 36.4% of the entire V_H-gene repertoire was composed of members of the *IGHV1* family, whereas the V_H-gene repertoire in mice injected only with adjuvant were dominated by sequences from the *IGHV5* or *IGHV6* families (**Supplementary Fig. 5**).

In most instances, the V genes encoding a highly abundant CDR3 were dominated by one sequence; the second most abundant V-gene sequence (somatic variant) was present at 10% the level and differed

© 2010 Nature America, Inc. All rights reserved.

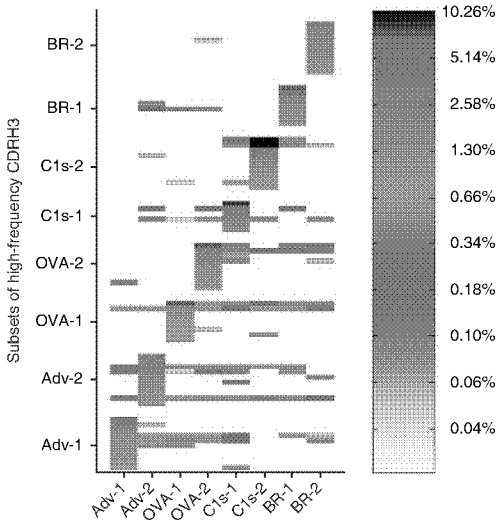

Figure 2 Comparison of high-frequency CDRH3s reveals unique V_H genes in each of eight mice immunized with one of three antigens or an adjuvant control. Heat map showing the distribution of highly represented CDRH3s in mice injected with Adjuvant (Adv), ovalbumin (OVA), C1s and Bright (BR). The *y* axis represents the ten highest frequency CDRH3 sequences identified in each mouse. The *x* axis compares the frequency of these prevalent CDRH3 sequences across all other mice. White, sequences found at frequencies that are not statistically significant (0.00–0.03%). Black, sequences found at a frequency of >10%.

from the dominant sequence by one or two amino acids. However, there were some instances where abundant CDRH3s were encoded by several V genes that were represented at comparable frequencies (**Supplementary Fig. 6** and **Supplementary Table 7**). Notably, the V_H repertoires were quite distinct even among genetically identical littermates immunized with the same antigen on the same day. For mice immunized with C1s or Bright, each mouse developed a distinct and diverse set of abundant CDRH3 sequences (**Fig. 2** and **Supplementary Table 8**). This suggests that each mouse generates its own unique and highly expressed V_H-gene repertoire, which may allow for the discovery of a panel of diverse antibodies. One exception, however, was that a few CDRH3 sequences abundant in both ovalbumin-immunized mice were also present at high frequency in other mice, suggesting that the corresponding antibodies may be polyspecific. Not surprisingly, some CDRH3 sequences from animals that received adjuvant only, found at moderate levels, were also present in immunized mice (**Fig. 2**). Antibodies encoding these sequences were probably specific to adjuvant or to common natural antigens. CDRL3 diversity was lower, with several promiscuous sequences represented at high frequency in several mice (**Supplementary Table 9**). Fourth, even though the BMPC V_H repertoires were largely composed of sequences unique to each mouse, principal component analysis of CDRH3s shared between mice revealed distinct clustering of the data for each cohort (that is, same cage and litter) immunized at the same time but with different antigens (**Supplementary Fig. 7**). This signature likely reflects environmental factors, such as the antigenic history of the animal groups, and suggests that V-gene repertoire analysis may provide valuable diagnostic information.

It should be noted that a few copies (typically <5) of the most abundant CDRH3 sequences raised to a given antigen were observed at very low levels (typically <0.1%) in the CDRH3 repertoires of mice receiving other antigens. As several of the respective V genes were shown to encode antigen-specific antibodies (see below), we believe that the presence of these sequences in mice immunized with other antigens might originate from low levels of cross-sample contamination, a conclusion supported by the biased distributions of common CDRH3 sequences within the same cohort (**Supplementary Fig. 8**). Because of the high sensitivity of 454 DNA sequencing, even with the utmost care it is not possible to completely rule out low-level contamination (sequence noise) during library preparation/multiplex sequencing. Although an important consideration for studies aiming to compare unbiased repertoires[19,20], sequence noise does not affect the methodology described herein, as the most abundant V genes in the BMPC repertoire are 20- to >100-fold more abundant than the sequence noise level.

Manual screening of small combinatorial libraries of scFvs in *Escherichia coli* using BMPC V genes led to a low yield of antigen-specific clones (less than four positive clones per 96-well plate; data not shown). Upon further analysis, most of these scFvs displayed low apparent affinity by enzyme-linked immunosorbent assay (ELISA) and/or poor expression and aggregation. We reasoned that this was a consequence of combinatorial pairing; even if a V_L and a V_H gene represented 5% of the cDNA pool, assuming no PCR biases in scFv assembly, the probability of correct pairing is only 0.25%. Discovery of positive clones would thus require an extensive amount of screening. To overcome these problems, and to avoid screening altogether, we hypothesized that V_L and V_H genes represented at approximately the same frequency likely arise from the same plasma cell and, hence, are naturally paired. To test this hypothesis, we synthesized the four or five most abundant full-length V_L and V_H genes from each mouse (excluding V_H sequences that were cross-represented in adjuvant-only mice), which accounted for a minimum of 0.5% of the repertoire, expressed the recombinant antibodies, and tested for antigen binding. Synthetic genes were constructed by robotically assisted, high-throughput DNA synthesis (Online Methods). Briefly, gene fragments (200–500 nucleotides long) were generated using inside-out nucleation PCR reactions. The design of these fragments and relevant overlaps was automated using customized software to facilitate robotic synthesis and assembly (**Supplementary Data**). Alignment and 'padding' of the sequences at either end yielded genes of identical length and permitted the use of a generic overlapping assembly strategy that ensured the greatest oligonucleotide reuse (**Supplementary Fig. 9**). In this manner, up to 48 V_L and 48 V_H genes could be synthesized and validated for the correct open reading frame by one researcher within 1 week, at a reagent cost of <$2,000. In most cases, V_L and V_H pairing was determined by rank ordering of CDR3 frequency within the repertoire. In cases where two V_L or V_H genes were found at very similar frequencies, we constructed multiple V_L-V_H combinations. Paired V genes were then expressed as scFv fragments in *E. coli*. ELISA analysis of bacterial lysates indicated that the resulting antibodies were overwhelmingly antigen specific (~78%): we obtained 21/27 antigen-specific antibodies from six mice immunized with three different protein antigens (**Table 1**). To further evaluate the utility of this simple pairing strategy, we constructed a combinatorial library of scFvs comprising the four most abundant V_L and V_H genes from each of the two mice immunized with C1s. scFv antibodies were expressed in *E. coli*. Binding analysis by ELISA revealed that all of the highest antigen-binding clones possessed the same V_L-V_H gene combinations predicted by our pairing strategy (**Supplementary Table 10**).

As mouse 2 immunized with C1s (C1s-2) displayed the highest serum titers (**Supplementary Table 11**), we selected antibodies from this animal for biophysical characterization of antigen binding affinity by surface plasmon resonance. Antibodies were expressed from synthetic genes and purified as monomeric scFv fragments in

© 2010 Nature America, Inc. All rights reserved.

Table 1 Antigen binding of antibody single-chain variable fragments (scFvs) from high frequency V_L and V_H genes

V_L-V_H pair	% V_L	CDRL3	% V_H	CDRH3	scFv binding
α–OVA					
1.1L-1.1H	11.70	WQGTHFPLT	7.11	GSSYYAMDY	+
1.2L-1.2H	4.40	QQYNSYPLT	1.10	LLWLYAMDY	+
1.3L-1.3H	3.38	QQSNSWYT	0.57	DVYDGYAMDY	+
1.4L-1.4H	2.20	QHHYGTPPWT	0.54	NPYAMDY	–
2.1L-2.1H	5.32	WQGTHFPLT	7.61	RTTVSRDWYFDV	+
2.2L-2.2H	4.05	QQYNSYPLT	3.23	YYYGSSAMDY	+
2.3L-2.3H	3.46	QQYSSYPLT	2.22	DGWYYFDY	+
2.4L-2.4H	2.01	QQHYSTPWT	2.10	EDDYDLFAY	+
α–C1s					
1.1L-1.1H	12.95	WQGTHFPQT	7.93	GNYYYAMDY	+
1.2L-1.1H	6.94	QQWSSYPQLT	7.93	GNYYYAMDY	+
1.3L-1.2H	3.81	QNDHSYPLT	2.64	DMISYWYFDV	+
1.4L-1.3H	3.16	QQGQSYPFT	1.67	EDYGNYWYFDV	+
1.4L-1.4H	3.16	QQGQSYPFT	1.67	EGYYYGSSYFDY	–
2.1L-2.1HA	17.10	FQGSHVPLT	10.99	SDRYDGYFDY	+
2.1L-2.1HB	17.10	FQGSHVPLT	9.93	SDRFDGYFDY	+
2.2L-2.2H	2.62	QQSNEDPWT	3.30	WLLLAY	+
2.3L-2.2H	2.20	WQGTHFPH	3.30	WLLLAY	+
2.3L-2.3H	2.20	WQGTHFPH	1.65	SDGYYYFDY	+
2.4L-2.4H	1.64	QQHYSTPFT	1.15	YYDYDKAYYFDY	–
α–Br					
1.1L-1.1H	6.64	LQYASSPFT	7.20	HDYGNYVDY	+
1.2L-1.2H	4.73	WQGTHFPRT	5.62	DGNYQEDYFDY	–
1.3L-1.3H	4.51	QQNNEDPRT	1.91	EGYAYDVDY	+
1.4L-1.4H	3.59	QQRSSYPLT	1.20	YDYGKDFDY	+
2.1L-2.1H	7.24	WQGTHFPQT	2.57	RGDGNYFFDY	+
2.2L-2.2H	4.50	QQGQSYPWT	2.27	GDEAWFAY	–
2.3L-2.3H	3.12	LQYASSPYT	2.03	EGDFDY	–
2.4L-2.4H	2.58	FQGSHVPWT	1.63	GGNYDYAMDY	+

E. coli whole-cell lysates expressing antibody scFvs that were constructed by pairing the most abundant V genes (as shown above). V_L and V_H gene pairing was determined by relative frequency (%) of the respective V genes in the BMPC repertoires. ELISA analysis was performed to determine antigen binding (Online Methods). +, more than threefold stronger ELISA signal on antigen-coated wells relative to wells coated with unrelated antigen (BSA and/or gelatin). OVA, ovalbumin; BR, bright.

© 2010 Nature America, Inc. All rights reserved.

E. coli and as full-length IgG antibodies in HEK 293F cells. Pairing of the most abundant light (2.1L) and heavy (2.1H-B) V genes (frequencies, 17.10% CDRL3 and 9.93% CDRH3) from mouse C1s-2 yielded an antibody with a K_D of 20 nM as a scFv ($k_{on} = 2.3 \times 10^4\ M^{-1}\ sec^{-1}$; $k_{off} = 5.0 \times 10^{-4}\ sec^{-1}$) and unexpectedly, a slightly lower monovalent K_D of 50 nM ($k_{on} = 2.4 \times 10^4\ M^{-1}\ sec^{-1}$; $k_{off} = 1.2 \times 10^{-3}\ sec^{-1}$) as an IgG. From the same mouse, pairing of C1s genes 2.2L with 2.2H (frequencies, 2.62% CDRL3 and 3.30% CDRH3) resulted in an IgG that displayed low binding affinity (K_D of ~500 nM, data not shown). However, the pairing of C1s genes 2.3L with 2.2H (frequencies, 2.20% CDRL3 and 3.30% CDRH3) generated an IgG with subnanomolar binding affinity ($K_D = 0.43$ nM; $k_{on} = 4.5 \times 10^5\ M^{-1}\ sec^{-1}$; $k_{off} = 1.9 \times 10^{-4}\ sec^{-1}$; **Supplementary Fig. 10** and **Supplementary Table 12**), indicating that the natural pairing is likely 2.3L–2.2H. Furthermore, the antibodies were suitable for functional assays, such as sandwich ELISA and immunoprecipitation of C1s from human serum (**Supplementary Figs. 11** and **12**).

Our approach capitalizes on mining the antibody repertoire of BMPCs, a population of B cells that is responsible for the synthesis of the large majority of circulating immunoglobulins in animals[16]. Although we have validated this methodology in mice, there is no reason to believe that the same approach cannot be readily extended to primates, including humans. Furthermore, it is possible that this technology could be extended for antibody discovery with more complex antigens such as viral and bacterial pathogens, as in these situations BMPCs may still develop polarity. We note, however, that the polarization of the BMPC repertoire in instances when the antigen may contain multiple highly immunogenic epitopes requires further evaluation. The mechanisms that dictate the selection of B-cell differentiation into plasma cells and homing into the bone marrow are complex and appear to partially relate to high antigen affinity[23,24]. As the highly abundant BMPCs correspond to abundant circulating antibodies, it seems plausible to hypothesize that these antibodies have been selected by the immune system (at least partly) because they display more potent pathogen neutralization. Therefore, antibodies generated by the mining of the BMPC repertoire may prove particularly useful for therapeutic purposes. The hybridoma technology and other B-cell immortalization methods interrogate the antibody-producing cells in pre–plasma cell B-cell populations, specifically in memory B cells, or in circulating short-lived plasmablasts[9]. Fully differentiated plasma cells are not amenable to most of these analyses, as they do not survive outside their biological niches. Very recently, microwell arrays and single-cell cloning were used to isolate antibodies from spleen plasma cells[12]. Nonetheless, despite the use of a sophisticated screening technology, only small numbers of antigen-specific clones could be isolated. Consequently, information on the repertoire and relative abundance of V genes could not be obtained by this method.

Our use of high-throughput DNA sequencing, bioinformatic analysis and automated gene synthesis can lead to the isolation and expression of mAbs with minimal effort. In our hands, it takes ~10 person hours for sample preparation for DNA sequencing. With automated bioinformatic processing of the 454 sequencing data, no extra effort is required to identify highly abundant V_L and V_H genes for DNA synthesis. Synthetic genes can be constructed either by an automated facility (as described herein) or through commercial gene-synthesis vendors. Furthermore, antibody genes can be codon optimized as desired for either bacterial or mammalian expression and subsequent characterization studies. Thus, in terms of effort by dedicated personnel (not including DNA sequencing and synthesis, which are carried out by multi-user services) and time line required for antibody discovery, our method compares very favorably to methods involving hybridomas, B-cell immortalization, and B-cell screening and/or single-cell cloning. Currently, the most expensive part of our antibody discovery process is DNA sequencing followed by gene synthesis. However, the cost for these technologies is declining at a rapid and exponential pace, resembling Moore's law for microelectronics[25,26]. Taken within this context, we envisage that the expense for our approach to antibody discovery will eventually not be a limitation.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Accession codes. GenBank: HQ15046,47,49,51-91.

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

ACKNOWLEDGMENTS

We thank C. Das for her invaluable technical assistance on mammalian cell culture and transfection and M. Byrom for assistance with mouse experiments. This work was funded by grants from the Clayton Foundation, the Cockrell Chair in Engineering (to G.G.) and by a fellowship from Natural Sciences and Engineering Research Council of Canada (to X.G.).

AUTHOR CONTRIBUTIONS

S.T.R. and G.G. developed the methodology, designed experiments, analyzed the data and wrote the manuscript; X.G., C.C., and K.H.H. carried out the

bioinformatic analysis; S.T.R., A.E.M., R.A.H., S.H.K. and K.H.H. performed the experiments; S.P.H.-S. performed 454 DNA sequencing; B.L.I., P.W.T. and A.D.E. helped analyze the data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).
2. Kwakkenbos, M.J. *et al.* Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming. *Nat. Med.* **16**, 123–128 (2010).
3. Clackson, T. *et al.* Making antibody fragments using phage display libraries. *Nature* **352**, 624–628 (1991).
4. Feldhaus, M.J. *et al.* Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library. *Nat. Biotechnol.* **21**, 163–170 (2003).
5. Harvey, B.R. *et al.* Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from *Escherichia coli*-expressed libraries. *Proc. Natl. Acad. Sci. USA* **101**, 9193–9198 (2004).
6. Schaffitzel, C., Hanes, J., Jermutus, L. & Plückthun, A. Ribosome display: an in vitro method for selection and evolution of antibodies from libraries. *J. Immunol. Methods* **231**, 119–135 (1999).
7. Hoogenboom, H.R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23**, 1105–1116 (2005).
8. Traggiai, E. *et al.* An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. *Nat. Med.* **10**, 871–875 (2004).
9. Wrammert, J. *et al.* Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453**, 667–671 (2008).
10. Meijer, P.-J. *et al.* Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. *J. Mol. Biol.* **358**, 764–772 (2006).
11. Mazor, Y., Blarcom, T.V., Mabry, R., Iverson, B.L. & Georgiou, G. Isolation of engineered, full-length antibodies from libraries expressed in *Escherichia coli*. *Nat. Biotechnol.* **25**, 563–565 (2007).
12. Jin, A. *et al.* A rapid and efficient single-cell manipulation method for screening antigen-specific antibody-secreting cells from human peripheral blood. *Nat. Med.* **15**, 1088–1092 (2009).
13. Love, J.C., Ronan, J.L., Grotenbreg, G.M., van der Veen, A.G. & Ploegh, H.L. A microengraving method for rapid selection of single cells producing antigen-specific antibodies. *Nat. Biotechnol.* **24**, 703–707 (2006).
14. Cobaugh, C.W., Almagro, J.C., Pogson, M., Iverson, B. & Georgiou, G. Synthetic antibody libraries focused towards peptide ligands. *J. Mol. Biol.* **378**, 622–633 (2008).
15. Persson, H., Lantto, J. & Ohlin, M. A focused antibody library for improved hapten recognition. *J. Mol. Biol.* **357**, 607–620 (2006).
16. Manz, R.A., Hauser, A.E., Hiepe, F. & Radbruch, A. Maintenance of serum antibody levels. *Annu. Rev. Immunol.* **23**, 367–386 (2005).
17. Shapiro-Shelef, M. & Calame, K. Regulation of plasma-cell development. *Nat. Rev. Immunol.* **5**, 230–242 (2005).
18. Manz, R.A., Thiel, A. & Radbruch, A. Lifetime of plasma cells in the bone marrow. *Nature* **388**, 133–134 (1997).
19. Weinstein, J.A., Jiang, N., White, R.A., Fisher, D.S. & Quake, S.R. High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324**, 807–810 (2009).
20. Boyd, S.D. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* **1**, 12ra23 (2009).
21. Glanville, J. *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc. Natl. Acad. Sci. USA* **106**, 20216–20221 (2009).
22. Ge, X., Mazor, Y., Hunicke-Smith, S.P., Ellington, A.D. & Georgiou, G. Rapid construction and characterization of synthetic antibody libraries without DNA amplification. *Biotechnol. Bioeng.* **106**, 347–357 (2010).
23. Radbruch, A. *et al.* Competence and competition: the challenge of becoming a long-lived plasma cell. *Nat. Rev. Immunol.* **6**, 741–750 (2006).
24. Phan, T.G. *et al.* High affinity germinal center B cells are actively selected into the plasma cell compartment. *J. Exp. Med.* **203**, 2419–2424 (2006).
25. Carlson, R. The changing economics of DNA synthesis. *Nat. Biotechnol.* **27**, 1091–1094 (2009).
26. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008).

© 2010 Nature America, Inc. All rights reserved.

ONLINE METHODS

Immunization. Purified C1s (CalBiochem), purified chicken egg ovalbumin (Sigma), or recombinant bacterially expressed human B-cell regulator of IgH transcription (Bright) were resuspended in sterile-filtered PBS at 1.0 mg/ml. On the day of primary immunization, 25 μl of antigen solution was thoroughly mixed with 25 μl of complete Freund's adjuvant (CFA; Pierce Biotechnology) and 50 μl of sterile PBS and stored on ice. Female BALB/c mice (Charles Rivers Laboratories) 6–8 weeks old were housed in conventional barrier space and were maintained on a normal chow diet. Before injections, mice were bled from the tail vein and ~25 μl of blood was collected and stored at −20 °C for later analysis. Day 1 was designated as the day primary immunizations were performed. 100 μl of the antigen-CFA mixture per mouse was injected with a 26-gauge needle subcutaneously into the backpad. Mice were monitored daily by animal housing staff and cages were changed twice per week.

For secondary immunization, 25 μl of antigen solution was thoroughly mixed with 25 μl of incomplete Freund's adjuvant (IFA; Pierce Biotechnology) and 50 μl of sterile PBS and stored on ice. On day 21, mice were given the secondary immunization intraperitoneally at 100 μl of antigen-IFA mixture per mouse. On day 26, mice were euthanized by CO_2 asphyxiation and blood, femurs and tibia were collected.

Isolation of BMPCs. Muscle and fat tissue were removed from the harvested tibias and femurs. The ends of both tibia and femurs were clipped with surgical scissors and bone marrow was flushed out with a 26-gauge insulin syringe (Becton Dickinson, BD). Bone marrow tissue was collected in sterile-filtered buffer no. 1 (PBS with 0.1% BSA/2 mM EDTA). Bone marrow cells were collected by filtration through a 70-μm cell strainer (BD) with mechanical disruption and washed with 20 ml of PBS and collected in a 50 ml tube (Falcon, BD). Bone marrow cells were then centrifuged at 335*g* for 10 min at 4 °C. Supernatant was decanted and the cell pellet was resuspended with 3.0 ml of red blood cell lysis buffer (eBioscience) and shaken gently at 25 °C for 5 min. Cell suspension was then diluted with 20 ml of PBS and centrifuged at 335*g* for 10 min at 4 °C. Supernatant was decanted and cell pellet was resuspended in 1.0 ml of buffer no. 1.

Each isolated bone marrow cell suspension was incubated with 2.5 μg and 1.5 μg of biotinylated rat anti-mouse CD45R(B220) and biotinylated rat anti-mouse CD49b (eBioscience), respectively. Cell suspension was rotated at 4 °C for 20 min. Cell suspensions were then centrifuged at 930*g* for 6 min at 4 °C, supernatant was removed and the cell pellet was resuspended in 1.5 ml of buffer no. 1. Streptavidin conjugated M280 magnetic beads (Invitrogen) were washed and resuspended according to manufacturer's protocol. 50 μl of magnetic beads were added to each cell suspension and the mixture was rotated at 4 °C for 20 min. Cell suspensions were then placed on Dynabead magnet (Invitrogen) and supernatants (negative fraction, cells unconjugated to beads) were collected and cells bound to beads were discarded.

Prewashed streptavidin M280 magnetic beads were incubated for 30 min at 4 °C with biotinylated rat anti-mouse CD138 (BD Pharmingen) with 0.75 μg antibody per 25 μl of magnetic beads. Beads were then washed according to manufacturer's protocol and resuspended in buffer no. 1. The negative cell fraction (depleted of $CD45R^+$ and $CD49b^+$ cells) collected as above was incubated with 50 μl of CD138-conjugated magnetic beads and the suspension rotated at 4 °C for 30 min. Beads with $CD138^+$ bound cells were isolated by the magnet, washed 3 times with buffer no. 1, the negative ($CD138^-$) cells unbound to beads were discarded (or saved only for analysis). The positive $CD138^+$ bead-bound cells were collected and stored at 4 °C until further processed.

Preparation of V_L and V_H genes. $CD45R^-$ $CD138^+$ BMPCs isolated as described herein were centrifuged at 930*g* at 4 °C for 5 min. Cells were then lysed with TRI reagent and total RNA was isolated according to the manufacturer's protocol in the Ribopure RNA isolation kit (Ambion). mRNA was isolated from total RNA with oligodT resin and the Poly(A) purist kit (Ambion) according to the manufacturer's protocol. mRNA concentration was measured with an ND-1000 spectrophotometer (Nanodrop).

The isolated mRNA was used for first-strand cDNA synthesis by reverse transcription with the Maloney murine leukemia virus reverse transcriptase (MMLV-RT, Ambion). cDNA synthesis was performed by RT-PCR using 50 ng of mRNA template and oligo(dT) primers according to manufacturer's protocol of Retroscript kit (Ambion). After cDNA construction, PCR amplification was performed to amplify the V_L and V_H genes with a standard mix of degenerate primers[27]. A complete list of primers can be found in **Supplementary Table 13**. A 50 μl PCR reaction consisted of 0.2 mM of forward and reverse primer mixes, 5 μl of Thermopol buffer (NEB), 2 μl of unpurified cDNA, 1 μl of Taq DNA polymerase (NEB) and 39 μl of double-distilled H_2O. The PCR thermocycle program was: 92 °C for 3 min; 4 cycles (92 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min); 4 cycles (92 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min); 20 cycles (92 °C for 1 min, 63 °C for 1 min, 72 °C for 1 min); 72 °C for 7 min; 4 °C storage. PCR gene products were gel purified and submitted to SeqWright and Genomic Sequencing and Analysis Center at the University of Texas Austin for Roche GS-FLX 454 DNA sequencing.

High-throughput sequencing of V_L and V_H repertoires. V-gene repertoires isolated from BMPC of eight mice were sequenced using high-throughput 454 GS-FLX sequencing (University of Texas, Austin, TX; SeqWright). In total, 415,018 sequences were generated, and 454 data quality-control filtered and grouped >97% of the sequences into data sets for each mouse according to their Multiplex Identifiers usages.

Bioinformatic analysis. (1) CDR3 identification. A search method was developed based on conserved flanking sequence motifs found upstream and downstream of CDR3. Searching motifs for CDRH3 and CDRL3 were determined based on amino acids that occur with an average frequency of 99% at specific positions in V genes from the Kabat database. (**Supplementary Table 1**). V_H sequences were searched for the motif DXXX(Y/F)(Y/F)C (Kabat # 86-92) and WGXG(T/S) (Kabat # 103-107) at N- and C- termini of CDRH3, respectively. Analogously, V_L genes were found by searching for the motifs DXXXY[F/Y]C (Kabat # 82-88) and FGXGT (Kabat # 98-102). This approach correctly identifies >94% of V_H and 92% of V_L full-length sequences in the Kabat database. Any sequences or reverse complements containing these motifs were extracted as either V_H or V_L genes, respectively. Only the sequences with in-frame CDR3 and without stop codons were further analyzed. For each sample, the most highly represented CDR3 sequences (typically represented at frequencies >1%) were discovered, and their relative abundances in all the other seven samples were calculated. To find a consensus full-length V_H/V_L gene sequence, sequences containing high-frequency CDR3s of interest were analyzed for pairwise homology by BLAST, and the sequence with the highest score was chosen. **Supplementary Figure 3** summarizes the bioinformatics analysis of the V-gene sequences. Analysis was performed using Perl scripts in a Unix environment, which were converted into a graphical user interface using the Matlab 7.1 GUI builder for enhanced visualization of results (**Supplementary Data**).

(2) Analysis of CDR3 expression across samples from different mice. CDR3 sequences found in multiple samples were extracted and analyzed for their prevalence in all mice. First, principle component analysis was performed using Matlab to analyze the variance of CDR3 expression in different mice (**Supplementary Fig. 7**). The majority of the variance between mouse samples was categorized into seven principle components. Second, the percent of CDRH3 sequences found in multiple samples was calculated. Because it could not be determined whether replicate sequences were due to contamination or a true biological effect, a permutation test was performed to determine whether the percentage of sequences shared across four samples was biased by samples analyzed on a specific day. The percent of shared sequences was calculated for all 70 possible combinations of the eight samples selected four times, subsequently ranked by percentage overlap. The top three ranked combinations were considered significant and not attributed to random combinations.

(3) Frequency distribution of abundant CDRH3. A heat map was generated to illustrate the prevalence of highly abundant CDRH3s from each sample in mice receiving different antigens. Only CDRH3 sequences with statistically significant frequencies in the top 5% of the distribution (frequency cutoff ~0.03%) were represented (**Fig. 2**).

(4) Homology analysis of full-length V genes. Full-length V genes were found for sequences containing identical CDR3s. First, sequences were placed

© 2010 Nature America, Inc. All rights reserved.

doi:10.1038/nbt.1673

in frame by docking CDR3 motifs. Second, full-length V-gene sequences were accepted if they did not contain stop codons and covered all three CDR regions. Nonidentical, full-length V genes (containing at least one amino acid difference) were aligned to determine pairwise homology using the multiple sequence alignment tool in Geneious Software (Biomatters Ltd.; **Supplementary Fig. 6** and **Supplementary Table 7**).

(5) Germline analysis. The top four full-length consensus V_L and V_H genes were analyzed by the IMGT/V-Quest Tool[28]. Additionally, the top 30 ranked CDRH3 sequences of four mice (adjuvant-1, adjuvant-2, C1s-1 and C1s-2) were further analyzed for V(D)J recombination using the IMGT/V-QUEST tool. The V segment germline usage and V_H gene somatic mutations were identified after the IMGT/V-QUEST analysis. These data are reported in **Supplementary Figure 5** and **Supplementary Table 6**.

Construction of synthetic antibody genes. The coding sequences for the selected V_L and V_H genes were designed using the GeneFab software component of our in-house protein fabrication automation (PFA) platform[29]. After reverse translation of the primary amino acid sequences for each V_L and V_H using an *E. coli* class II codon table, the coding sequences for each V_L and V_H were paired based upon their relative frequency from the sequencing data (most abundant V_L with the most abundant V_H, and so forth). The antibody V_L and V_H sequences were built into scFvs with a polyglycine-serine linker $(GGGGS)_4$ between the V_L and V_H sequences. The scFv genes were aligned using the sequence encoding the common $(GGGGS)_4$ linker sequence and a universal randomly generated stuffer sequence was applied to the ends of the scFv sequences to ensure that all of the constructs were of the same length (808 bp). This design format reduced the number of oligonucleotides needed for gene synthesis as oligonucleotides with identical sequences between the different scFv constructs could be reused. SfiI restriction endonuclease sites were added, flanking each gene sequence to facilitate cloning of the synthetic gene constructs into compatible pMoPac16 vectors[30].

The scFv genes were synthesized from overlapping oligonucleotides using a modified thermodynamically balanced inside-out nucleation PCR[31]. The 80 mer oligonucleotides necessary for the construction of the various scFv genes were designed using the GeneFab software with a minimal overlap of 30 nucleotides between oligonucleotide fragments. The oligonucleotides were synthesized using standard phosphoramidite chemistry at a 50 nmol scale using a Mermade 192 oligonucleotide synthesizer (Bioautomation) using synthesis reagents from EMD Chemical and phosphoramidites from Glen Research. All of the oligonucleotide liquid-handling operations necessary for assembling the various genes were done on a Tecan Evo 200 workstation (Tecan) with reagent management and instrument control done through the FabMgr software component of the PFA platform[29]. The gene assembly PCRs were performed using KOD-Hotstart polymerase using buffers and reagents supplied with the enzyme (Novagen). To facilitate cloning of the V_L and V_H genes separately into vectors for IgG expression, the genes for the various V_L and V_H pairs were either built as gene fusions similar to the scFvs except without the $(GGGGS)_4$ linker or as separate genes. These constructs contained sites for the restriction enzymes BssHII and BsiWI flanking the V_L gene and the BssHII and NheI sites flanking the V_H gene.

Antibody expression and antigen binding analysis. Antibody fragments were expressed as scFv fusions to the human light chain constant region Cκ (scAbs), followed by a C-terminal polyhistidine (polyHis) tag. Cloning was accomplished by SfiI digestion of antibody genes and ligation into the expression vector pMoPac16 followed by electroporation transformation into *E. coli Jude 1* cells, which were then plated on Luria Broth (LB, Miller) agar plates supplemented with 100 μg/ml ampicillin. Single colonies were used to inoculate cultures in microtiter 96-well plates with 200 μl/well of Terrific Broth (TB, Miller) supplemented with 2% glucose and 100 μg/ml ampicillin; plates were shaken for 16 h at 30 °C. 10 μl of each well was used to inoculate 200 μl/well of fresh 96-well plates containing TB media supplemented with 100 μg/ml ampicillin and 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG, Calbiochem).

After a 4 h IPTG induction at 25 °C with shaking, plates were centrifuged at 3,600*g* for 10 min at 4 °C, the supernatant was decanted and cell pellets were resuspended in 20% BugBuster HT (Novagen) in PBS at 150 μl/well. Plates were then shaken at 25 °C for 30 min, and then centrifuged at 3,600*g* for 15 min at 4 °C. 50 μl/well of cell lysates were then added to an ELISA 96-well plate that was precoated with antigen (e.g., ovalbumin, C1s, Bright) at 2 μg/ml in PBS and preblocked with 0.5% BSA or 1% gelatin. A standard indirect ELISA protocol was followed with the detection anti-polyHis antibody (Sigma) conjugated to horseradish peroxidase (HRP) and developed with TMB substrate (Dako) for 15–45 min and stopped with 2N H_2SO_4. The absorbance was measured at 450 nm with a 96-well spectrophotometer (BioTek). Positive wells were identified when the absorbance value was at least threefold above background binding to BSA.

For IgG expression, synthetic V_L and V_H genes were digested with BssHII/BsiWI and BssHII/NheI, respectively, and then ligated into the vectors pMAZ-IgL and pMAZ-IgH, respectively[32]. pMAZ-IgL carries the constant human kappa light chain antibody region and pMAZ-IgH carries the constant human heavy chain antibody region of IgG1. Vectors were transformed into *E. coli Jude 1* cells and plated on LB agar plates supplemented with 100 μg/ml ampicillin. Single colonies were selected and verified for correct V gene sequence. *E. coli* cells carrying pMAZ-IgL and pMAZ-IgH vectors were then grown in 2 ml TB supplemented with 100 μg/ml ampicillin; after overnight growth, plasmid DNA was isolated and purified. 20 μg each of purified pMAZ-IgL and pMAZ-IgH were used for cotransfection and transient expression from HEK 293F cells following the Freestyle MAX expression system (Invitrogen). HEK 293F cells were grown for 96 h after transfection and medium was harvested and IgG was purified by a protein-A agarose chromatography column.

Surface plasmon resonance. C1s was covalently immobilized on a CM5 chip (GE Healthcare) at a level of ~200 response units via standard amine coupling chemistry as described in the manufacturer's protocol. BSA was similarly coupled for baseline correction. All kinetic analyses were performed at 25 °C in HBS-EP (10 mM HEPES, 150 mM NaCl, 50 μM EDTA, 0.005% P-20, pH 7.4) on a BIAcore 3000 (GE Healthcare). Antibodies were injected over immobilized antigen at a flow rate of 50 μl/min or 100 μl/min and the chip was regenerated with a single 10s injection of 20 mM NaOH. Each sensogram was run in duplicate. Kinetic and equilibrium constants were determined by global fitting to a bivalent model using BIAevaluation software (GE Healthcare).

Software. Software is available upon request and on our website: (http://www.che.utexas.edu/georgiou/home.htm).

27. Krebber, A. *et al.* Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J. Immunol. Methods* **201**, 35–55 (1997).
28. Brochet, X., Lefranc, M.P. & Giudicelli, V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* **36**, W503–W508 (2008).
29. Cox, J.C., Lape, J., Sayed, M.A. & Hellinga, H.W. Protein fabrication automation. *Protein Sci.* **16**, 379–390 (2007).
30. Hayhurst, A. *et al.* Isolation and expression of recombinant antibody fragments to the biological warfare pathogen Brucella melitensis. *J. Immunol. Methods* **276**, 185–196 (2003).
31. Gao, X., Yo, P., Keith, A., Ragan, T.J. & Harris, T.K. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **31**, e143 (2003).
32. Mazor, Y., Barnea, I., Keydar, I. & Benhar, I. Antibody internalization studied using a novel IgG binding toxin fusion. *J. Immunol. Methods* **321**, 41–59 (2007).

© 2010 Nature America, Inc. All rights reserved.