J|A|C|S COMMUNICATIONS

Published on Web 06/25/2009

Evolution of Proteins with Genetically Encoded "Chemical Warheads"

Chang C. Liu, Antha V. Mack, Eric M. Brustad, Jeremy H. Mills, Dan Groff, Vaughn V. Smider,* and Peter G. Schultz*

Department of Chemistry and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received April 14, 2009; E-mail: vvsmider@scripps.edu; schultz@scripps.edu

The genetic code specifies the 20 amino acid building blocks from which all proteins are translated. Nevertheless, it is clear from nature that there exists an evolutionary impetus for additional chemical functionalities beyond those found in the canonical 20 amino acids; examples include cofactor-dependent catalysis, post-translational modifications, and pyrrolysine incorporation in methanogenic archea.[1] We recently demonstrated that a 21 amino acid code (containing a genetically encoded sulfotyrosine) can confer a selective advantage in a phage-based system for the evolution of gp120-binding antibodies.[2] Here we ask whether unnatural amino acids containing "chemical warheads" (like those found in many serine and metalloprotease inhibitors) could be advantageous to the evolution of proteins that target specific functional groups. We report the generation of antibodies in an *Escherichia coli* strain (Boro-X-*E. coli*)[2,3] that encodes *p*-boronophenylalanine (BF, **1**) in response to the amber nonsense codon TAG and show that from a library of mutants, specific clones containing BF emerge from selection for binding to an acyclic glycan. These findings further support the notion that additional genetically encoded amino acids can confer an evolutionary advantage and represent a significant advance toward the generation of novel protein therapeutics that selectively target glycans, nucleophilic serine hydroxyl groups, and the like using proteins with uniquely reactive unnatural amino acid side chains.

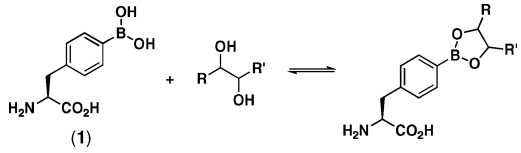

Figure 1. *p*-Boronophenylalanine can form boronate esters with diols.

We hypothesized that glycan-binding proteins could be evolved in Boro-X-*E. coli* because the boronic acid group has an intrinsic affinity for diols[4,5] (Figure 1). This effectively converts the challenge of optimizing multiple weak interactions to one of optimizing a smaller number of strong interactions. The latter is a more tractable problem for experimental evolution, which is limited in depth and route of sequence search. Starting with a human germline antibody containing V_H 3−23, we completely randomized six residues in the V_H CDR3 loop by site-saturation mutagenesis with the codon NNK (N = A, T, G, C; K = G, T) and also partially randomized two flanking residues. This fragment was combined with the A27 light chain and cloned into the pSEX phagemid backbone to yield pSEX-GermNNKLib, a library (maximal experimental diversity = 5×10^8) encoding antibodies in the single-chain variable fragment (scFv) format for phage display as multivalent pIII fusions. This library was then transformed into Boro-X-*E. coli*, infected with hyperphage,[6] and used to produce phage in a medium containing BF. Sequencing showed that 17% of the resulting phage contained a TAG codon in the randomized region (the number of clones sequenced was $n = 100$), confirming that a subset of the population had BF in their displayed proteins. This percentage is consistent with the codon frequency of TAG in NNK randomization (1 in 32) and the expected expression bias against sequences containing BF.[2] As a negative control, phage were similarly produced from Boro-X-*E. coli* but without BF in the medium. As expected, almost no clones sequenced contained the TAG codon; the 2% of clones that did ($n = 50$) were likely the result of low-level background amino acid incorporation by the BF-specific aminoacyl-tRNA synthetase, $B(OH)_2$PheRS, which occurs in the absence of its preferred substrate, BF.[3]

The scFv phage library was then used to select for antibodies that bind glucamine, an amino sugar of sorbitol. Phage (4×10^8) were incubated with the resin XUS43594.00 (a polystyrene resin with *N*-methylglucamine conjugated as a tertiary amine) for 4 h and washed thoroughly. The remaining phage particles were eluted with 1.75 μg/mL trypsin (a trypsin site joins the phage coat protein pIII with the displayed antibody) and used to infect Boro-X-*E. coli* for two additional rounds of selection. Enrichment was assayed by phage titers, and the percentage of phage clones containing the TAG codon was determined by DNA sequencing before and after each round of selection. As shown in Table 1, selection resulted in an increase in the proportion of eluted phage and a corresponding rise in the frequency of clones containing a TAG. This enrichment for TAG occurred during selection for glucamine binding and not during phage expression: sequencing showed that for each selection round *N*, the phage population eluted in round $N - 1$ had a higher number of TAG-containing clones than the phage population obtained after amplification and expression as input for round *N* (this was expected from the expression bias against unnatural-containing sequences). Therefore, the emergence of clones containing TAG resulted from the added functional advantage of BF in binding the glucamine resin. The emergence of TAG-containing clones was quite dramatic: ~50% of the clones contained TAG after only one round of selection, and >80% of the clones contained a TAG after three rounds (Table 1).

Table 1. Enrichment Statistics

round	fraction eluted	clones with TAG
preselection	N/A	17 of 100 (17%)
1	9×10^{-5}	15 of 28 (54%)
2	2×10^{-3}	14 of 31 (45%)
3	4.8×10^{-2}	39 of 47 (83%)

Although the population converged toward TAG-containing clones after three rounds of selection, convergence toward any particular TAG-containing clone was weak (see the Supporting Information for a list of selected sequences). This observation raised the question of whether the selected clones bind glucamine using

10.1021/ja902985e CCC: $40.75

BF alone or require additional residues around BF to achieve maximal affinity. Analysis of the selected clones suggested that BF is not the sole basis for their selection, since specific sequence patterns clearly emerged from selection (e.g., the location of BF in the randomized region converged to NNK-randomized positions 5 and 6, the sequence Asp-Val-Asp was overrepresented, and the frequency of hydrophilic residues around BF was high). To test this experimentally, we randomly chose a number of library phage that contained a TAG but had not undergone selection and compared their activities to the selected TAG-containing clones from round 3. As shown in Figure 2a, the round-3 clones (post) bind glucamine more effectively than do the randomly chosen BF-containing clones from before selection (pre). This is even more apparent when the two most frequent clones from round 3, 172-6 and 172-7, are compared to the preselection clones. BF alone is therefore not enough to explain the fitness of the selected clones; the sequence around BF also contributes to binding and selectivity (see below).

To characterize the binding of our selected phage to a number of model glycans, sugars conjugated to bovine serum albumin (BSA) were then immobilized on a Maxisorp plate and incubated with phage clones (see the Supporting Information). As Figure 2b shows, selected clones 172-6 and 172-7 bound the glucamine−BSA conjugate (G-BSA) but not BSA alone, indicating specificity for glucamine. Clones 172-3 and 172-4, which are phage that contain BF but did not undergo selection, only weakly bound G-BSA over BSA, confirming that the presence of BF alone cannot account for the fitness of 172-6 and 172-7. In addition, none of the phage clones bound sialyl-LewisX− and LewisX−BSA conjugates (sL-BSA and L-BSA), demonstrating that 172-6 and 172-7 are selective for glucamine over these other glycans (this selectivity may result from the natural preference of aryl boronates for acyclic glycans, which is exaggerated in the evolved sequences).

We next converted the best-performing sequence, 172-6, into the antigen binding fragment format (Fab) by cloning its light and heavy chains into a Fab expression vector (see the Supporting Information). Boro-X-*E. coli* was used to produce soluble Fab (Fab-172-6), and the resulting protein was purified by Protein G with a yield of 0.5 mg/L. In parallel, we expressed and purified Fab-172-6-Y, which contains tyrosine instead of BF at the position corresponding to the TAG codon, with a yield of 1.0 mg/L. To determine relative binding affinities to the glucamine resin used in the selection, we incubated Fab-172-6, Fab-172-6-Y, and a control Fab (that neither contained BF nor had been selected for glucamine binding) with 25 mg of the glucamine resin at a Fab concentration of 1 μM in 50 μL of phosphate buffered saline (pH 7.4). The resin was washed thoroughly and eluted with a 1 M sorbitol solution. The relative amounts of eluted protein were determined by an enzyme-linked immunosorbent assay (ELISA). As Figure 2c shows, Fab-172-6 bound the glucamine resin much more effectively than Fab-172-6-Y, confirming that the BF is required for activity. Still, Fab-172-6-Y retained some function, as it bound the glucamine resin more effectively than did the control antibody, demonstrating that the sequence content around BF in Fab-172-6 can contribute to binding independently of BF. There may also be a BF-dependent contribution of these surrounding residues, as they contain side chains that can coordinate BF and thus increase its exchange rate.

These studies come with two caveats. First, glucamine is an acyclic sugar, and thus, the generality of this approach for targeting biologically important cyclic glycans remains to be demonstrated. Second, the glucamine resin contains a high density of possible binding sites and a tertiary amine; thus, the value of BF during selection for binding sparsely glycosylated proteins is unknown. We are currently exploring these questions with libraries based on known antiglycan antibodies, complex germline collections, or lectins. In addition, the boronate group may also be advantageous in the evolution of antibody or peptide inhibitors of serine proteases and other proteins with nucleophilic hydroxyl groups.

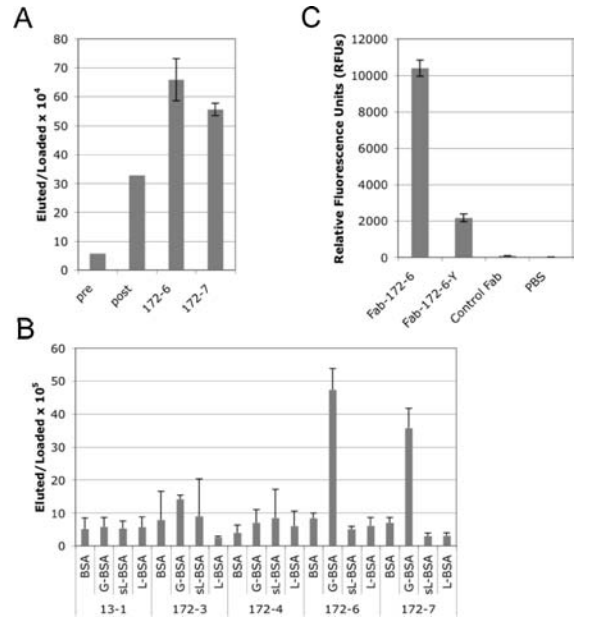

Figure 2. (A) Binding of glucamine resin by phage clones. Mean binding of three unique clones containing BF from before selection (pre) is compared to mean binding of five unique clones containing BF from after round 3 of selection (post) and binding of the two most frequent clones obtained from selection (172-6 and 172-7). Binding is reported as the fraction eluted, as determined by infection titers, multiplied by 10^4. Error bars are $\pm\sigma$ (σ = standard deviation) for triplicate experiments. (B) Binding of BSA and BSA-conjugated glycans by phage clones. The samples compared are a clone from before selection that does not contain a BF (13-1), two BF-containing clones from before selection (172-3 and 172-4), and the most frequently selected phage clones 172-6 and 172-7. Binding is reported as the fraction eluted, as determined by infection titers, multiplied by 10^5. Error bars are $+\sigma$ for triplicate experiments. (C) Binding of glucamine resin by expressed soluble Fab's. Eluted protein was coated onto wells of an ELISA plate, bound by an HRP-conjugated antibody against human κ light chain, and detected using a fluorogenic substrate (QuantaBlu). Error bars represent $\pm\sigma$ for triplicate experiments.

Acknowledgment. This research was supported by the DOE, NIH, Hertz Foundation (C.C.L.), and NSF (C.C.L.).

Supporting Information Available: Experimental details, selected sequences, and protein characterization. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Krzycki, J. A. *Curr. Opin. Microbiol.* **2005**, *8*, 706.
(2) Liu, C. C.; Mack, A. V.; Tsao, M. L.; Mills, J. H.; Lee, H. S.; Choe, H.; Farzan, M.; Schultz, P. G.; Smider, V. V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17688.
(3) Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2008**, *47*, 8220.
(4) Manimala, J. C.; Wiskur, S. L.; Ellington, A. D.; Anslyn, E. V. *J. Am. Chem. Soc.* **2004**, *126*, 16515.
(5) Li, M.; Lin, N.; Huang, Z.; Du, L.; Altier, C.; Fang, H.; Wang, B. *J. Am. Chem. Soc.* **2008**, *130*, 12636.
(6) Broders, O.; Breitling, F.; Dubel, S. *Methods Mol. Biol.* **2003**, *205*, 295.

JA902985E