Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 5590–5594, May 1996
Biochemistry

Active barnase variants with completely random hydrophobic cores

(protein design/protein evolution/molecular evolution/combinatorial mutagenesis)

DOUGLAS D. AXE, NICHOLAS W. FOSTER, AND ALAN R. FERSHT

Medical Research Council Unit for Protein Function and Design, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Contributed by Alan R. Fersht, January 30, 1996

ABSTRACT **The central structural feature of natural proteins is a tightly packed and highly ordered hydrophobic core. If some measure of exquisite, native-like core packing is necessary for enzymatic function, this would constitute a significant obstacle to the development of novel enzymes, either by design or by natural or experimental evolution. To test the minimum requirements for a core to provide sufficient structural integrity for enzymatic activity, we have produced mutants of the ribonuclease barnase in which 12 of the 13 core residues have together been randomly replaced by hydrophobic alternatives. Using a sensitive biological screen, we find that a strikingly high proportion of these mutants (23%) retain enzymatic activity *in vivo*. Further substitution at the 13th core position shows that a similar proportion of completely random hydrophobic cores supports enzyme function. Of the active mutants produced, several have no wild-type core residues. These results imply that hydrophobicity is nearly a sufficient criterion for the construction of a functional core and, in conjunction with previous studies, that refinement of a crudely functional core entails more stringent sequence constraints than does the initial attainment of crude core function. Since attainment of crude function is the critical initial step in evolutionary innovation, the relatively scant requirements contributed by the hydrophobic core would greatly reduce the initial hurdle on the evolutionary pathway to novel enzymes. Similarly, experimental development of novel functional proteins might be simplified by limiting core design to mere specification of hydrophobicity and using iterative mutation–selection to optimize core structure.**

Tight, well-ordered packing of interior hydrophobic side chains plays an important role in stabilizing the unique active conformations of natural proteins (1). Introduction of even small changes to a hydrophobic core typically diminishes the packing quality (2, 3, 33), the stability (2–6), and (often) the activity of the protein (7, 8). Reproduction of such exquisite core packing in proteins designed *de novo* has proven to be a considerable challenge (9, 10). Recent progress has been made on a four-helix bundle design by incorporating carefully engineered complementary side-chain packing and strategically placed polar residues (11). Because more complex folds involving β-sheets will pose an even greater challenge (12), it is worth considering whether the ultimate design objective, high activity, might be achieved without having to confront the packing problem.

If a designed fold lacking an engineered core exhibited even slight activity, this might serve as a starting point for experimental evolution (13–15) of a highly active novel structure. Highly functional natural proteins provide an opportunity to test the feasibility of this approach. Elegant studies using random core substitutions in a repressor domain (7, 16) have established, on the one hand, that some degree of function can be retained after several hydrophobic substitutions [up to four substitutions (16) in a core of ≈10 residues (7)] but, on the other, that the extent of functional impairment correlates with the extent of sequence modification (7). Is it possible, then, for a protein completely devoid of native-like core packing to exhibit phenotypically detectable activity? To tackle this question, we have saturated the core of barnase with random hydrophobic substitutions and applied a very sensitive biological screen to detect activity in the resulting mutants.

Experimental Approach. Codons for the 13 residues composing the main hydrophobic core of barnase (Fig. 1) were randomized in three stages, in which the extent of core randomization is progressively increased at each stage. Mutagenic primers were synthesized with thymine at the middle nucleotide of each core codon and a mixture of bases for the first and third nucleotides. The outcome at each core position is thereby restricted to a set of codons specifying five hydrophobic amino acids (7): Val, Leu, Ile, Met, and Phe. As the aim is to eliminate all traces of native core packing after the final mutagenesis stage, base mixtures were chosen to minimize the probability of retaining wild-type residues without reducing the number of possible substitutions.

Activity was detected in the core mutants by taking advantage of the extreme autotoxicity of barnase (18) when produced in *Escherichia coli* in the absence of its natural inhibitor, barstar. The principal component of the screening system is *synbar*, a synthetic barnase gene designed to produce mature barnase in the cytoplasm of *E. coli* (Fig. 2). Amber stop codons, replacing the serine codons normally at positions 28 and 57, prevent lethal production of barnase in a nonsuppressing (sup^-) *E. coli* host strain. When plasmid DNA prepared from the sup^- strain is used to transform a suppressing (*supD*) strain, the intervening stop codons are read as serine codons, causing a full-length product to be produced. If the *synbar* gene has not been modified, the product will be wild-type barnase, and no transformants will survive. Only *synbar* mutations yielding barnase variants with dramatically reduced activity allow the *supD* strain to grow. These mutants are scored as inactive, whereas those preventing growth of the *supD* strain are scored as active. Mutants having only 0.2% of wild-type activity test active in this system (see note † in Table 1).

MATERIALS AND METHODS

Bacterial Strains. *E. coli* strain MX383 is an amber-suppressing (*supD*), *Eco*K-restricting strain used as the experimental strain in the activity screen. During translation of mRNA, this strain reads a substantial fraction of amber (UAG) stop codons as Ser codons. Strain C-1a is a nonsuppressing (sup^-), nonrestricting strain used for preparation of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: amp^r, ampicillin resistance or ampicillin-resistant; $chlor^r$, chloramphenicol resistance or chloramphenicol-resistant.
Data Deposition: The sequence reported in this paper has been deposited in the GenBank data base (accession no. U46664).

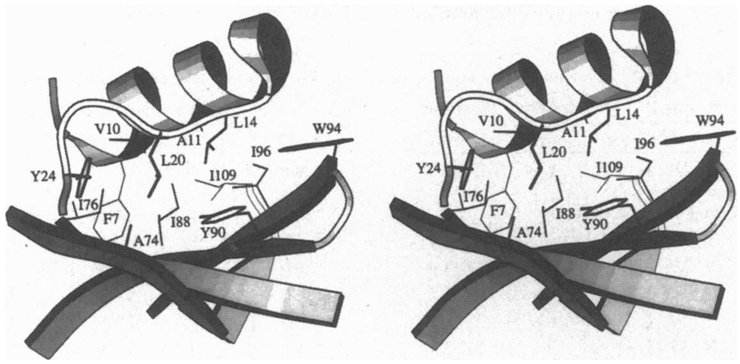

FIG. 1. Stereoview of the main hydrophobic core of barnase, formed by the packing of a 13-residue α-helix against a five-strand antiparallel β-sheet [rendered with MOLSCRIPT (17)].

mutant plasmid DNA and as a control strain in the activity screen. Both strains were obtained from the *E. coli* Genetic Stock Center (Yale University).

Plasmid Construction. The parent plasmid, pUC18, was modified by insertion (at the *Xba*I site) of a Genblock DNA fragment (Pharmacia) carrying the chloramphenicol-resistance gene, *cat*. Two alterations to the ampicillin-resistance gene, *bla*, were then introduced. First, codon 40 was changed from GCA to GCG (both coding for Ala) to disrupt the only *Eco*K restriction site on the plasmid. Then, codon 68, which specifies a catalytically essential Ser residue (23), was changed to an amber stop codon. The resulting plasmid confers chloramphenicol resistance ($chlor^r$) to all *E. coli* host strains but ampicillin resistance (amp^r) only to *supD* strains. Next, the *synbar* gene (Fig. 2) was inserted into the polylinker region of the plasmid such that transcription from the lac promoter directs expression of *synbar* followed by expression of *cat*. The final construct, designated pSNBR, causes sufficient *cat* expression for growth of the *supD* and sup^- strains in the presence of chloramphenicol (>25 mg/liter) without induction of the lac promoter.

Randomization of Hydrophobic-Core Codons in *synbar*. The 13 *synbar* codons specifying hydrophobic-core residues were randomized in three stages, as shown in Fig. 3. At stage 1, codons corresponding to the six residues on the helix side of the core (positions 7, 10, 11, 14, 20, and 24; Fig. 1) were simultaneously randomized by PCR. The product of this stage, a large mixed plasmid population, was used for two purposes. First, by using it to transform the sup^- strain, a collection of plasmid clones with six randomized core codons was obtained. These plasmid clones were used for sequence analysis and activity screening. Second, the same pool of DNA was carried over to the next mutagenesis stage (stage 2), which simultaneously randomized six of the sheet-side positions (positions 74, 76, 88, 90, 94, and 96; inclusion of codon 109 at stage 2 would have required undesirably long primers). Again, the resulting plasmid mixture was used to obtain a collection of plasmid clones for sequence analysis and activity screening. Clones at this stage are randomized at all core positions except position 109. Finally, at stage 3, position 109 was randomized in eight individual clones from the second stage, and 14 of the resulting clones were isolated for analysis.

```
<-P1------------------------------VTS-------------HTS-NTS-------------DTS-|-----
ATG GCT CAG GTT ATC AAC ACG TTT GAT GGT GTT GCT GAT TAC CTG CAG
M   A   Q   V   I   N   T   F   D   G   V10 A   D   Y   L   Q
-------------------DTK-----------NTK--------------------------------
ACT TAC CAT AAG CTC CCG GAT AAT TAC ATT ACT AAA TAG GAG GCG CAG
T   Y   H   K   L20 P   D   N   Y   I   T   K   *Am E   A30 Q
------->P2
GCT CTC GGT TGG GTT GCC TCT AAG GGT AAT CTG GCC GAC GTT GCT CCG
A   L   G   W   V   A   S   K   G40 N   L   A   D   V   A   P

GGC AAG TCT ATT GGT GGC GAT ATC TTT TAG AAC CGT GAG GGT AAG CTG
G   K   S50 I   G   G   D   I   F   *Am N   R   E60 G   K   L
        P3<-------------------------------NTS-------NTS-------------
CCG GGT AAG AGC GGT CGT ACT TGG CGT GAA GCT GAT ATT AAC TAC ACT
P   G   K   S   G   R   T70 W   R   E   A   D   I   N   Y   T
-----------------------------|---NTS-----NTK-------------NTS-----
AGC GGT TTC CGT AAC TCT GAC CGC ATT CTG TAC TCT TCT GAC TGG CTG
S80 G   F   R   N   S   D   R   I   L   Y90 S   S   D   W   L
NTS------------------------------------->P4
ATT TAC AAG ACT ACC GAC CAC TAC CAG ACC TTT ACT AAG ATC CGT TGA
I   Y   K   T   T100 D  H   Y   Q   T   F   T   K   I   R110 *
```

FIG. 2. Nucleotide sequence of the synthetic barnase gene, *synbar*. Residue numbering corresponds to the mature barnase sequence, where the N-terminal residue is Ala (18). The N-terminal Met resulting from expression of *synbar* in a suppressing host strain is expected to be removed inside the host bacterium (19), yielding the mature wild-type enzyme. Annealing positions of mutagenic primers *P1-P4* are indicated with arrows showing the direction of extension. Base mixtures introduced at the first 12 core codons are shown (D = A, G, T; H = A, C, T; K = G, T; N = A, C, G, T; S = C, G; and V = A, C, G). Primers *P5* and *P6* (not shown) introduce an NTK codon mixture at position 109.

Primers *P1–P6* were synthesized with the base mixtures indicated in Fig. 2 and purified by PAGE. PCR mixtures (50 μl) contained the appropriate primers (0.4 μM each) and 7.5 units (stages 1 and 2) or 3.8 units (stage 3) of recombinant *Pfu* DNA polymerase (Stratagene) in the supplied buffer. Amplification was achieved by incubation at 94°C for 4 min, followed by 15 cycles of denaturation (1 min at 94°C), annealing (1 min at 65°C for stages 1 and 2; 1 min at 58°C for stage 3), and extension (5.5 min at 72°C). After amplification, the principal PCR product was purified by agarose gel electrophoresis. Approximately 100 ng of purified DNA was circularized with T4 DNA ligase (New England Biolabs), and a portion of the ligated DNA was used to transform the sup^- strain, plating on Luria–Bertani agar with chloramphenicol (25 mg/liter). $Chlor^r$ colonies were cultured in liquid medium (2× TY medium) with chloramphenicol (25 mg/liter). Plasmid DNA was prepared from these overnight cultures using Wizard miniprep kits (Promega) according to the supplied protocol.

Sequencing of *synbar* Mutants. Plasmid clones prepared from the sup^- strain were used as templates in single cycle-sequencing reactions using [α-^{33}P]dATP (Amersham) and one dideoxynucleotide termination mix from the Amplicycle sequencing kit (Perkin–Elmer). After PAGE and autoradiography, lanes were inspected for base deletions within *synbar*. Of the clones examined after stage 1, approximately one-fourth carried deletions. Further accumulation of deletions at the second stage resulted in two-thirds of stage-2 clones having frameshift mutations. Only plasmid clones lacking deletions were examined further. These were used for PCR amplification of the mutant *synbar* genes along with flanking regions. The product DNA was purified using Wizard PCR prep kits (Promega) according to the supplied protocol, and complete sequencing of mutant *synbar* genes was accomplished by performing cycle-sequencing reactions with dye-labeled dideoxynucleotides (Perkin–Elmer) and analyzing products with an automated sequencer (Perkin–Elmer, ABI 373).

Barnase Activity Screen. Plasmid clones found to lack *synbar* deletions were used to transform the *supD* strain and the sup^- strain in parallel using the method of Chung and coworkers (24) except that dimethyl sulfoxide was omitted from the

Table 1. Sequence and activity data for hydrophobic-core mutants

ID	Amino acid residues by core position* (non-wild-type residues underlined) 7	10	11	14	20	24	74	76	88	90	94	96	109	Mutant activity score†	Total residue vol.‡ (Å^3)
wt:	F	V	A	L	L	Y	A	I	I	Y	W	I	I	++	2183.9
	Stage-1 mutants (ascending volume order)														
1:	L	L	L	V	V	V	A	I	I	Y	W	I	I	-	2136.7
2:	I	M	V	V	V	M	A	I	I	Y	W	I	I	-	2143.4
3:	I	I	I	V	I	V	A	I	I	Y	W	I	I	-	2166.5
4:	L	I	I	M	V	V	A	I	I	Y	W	I	I	++	2167.6
5:	L	F	L	V	V	V	A	I	I	Y	W	I	I	+	2172.2
6:	I	L	I	V	L	L	A	I	I	Y	W	I	I	-	2190.9
7:	L	L	M	I	L	V	A	I	I	Y	W	I	I	++	2192.9
8:	L	M	L	I	V	L	A	I	I	Y	W	I	I	+	2192.9
9:	L	M	L	V	M	M	A	I	I	Y	W	I	I	-	2197.8
10:	V	I	L	I	F	V	A	I	I	Y	W	I	I	-	2200.2
11:	I	I	L	V	F	V	A	I	I	Y	W	I	I	-	2200.2
12:	V	M	L	V	L	F	A	I	I	Y	W	I	I	++	2201.3
13:	M	M	L	I	L	L	A	I	I	Y	W	I	I	+	2222.0
14:	L	L	L	V	I	F	A	I	I	Y	W	I	I	++	2225.5
15:	L	L	L	I	F	V	A	I	I	Y	W	I	I	-	2225.5
16:	I	F	I	V	L	L	A	I	I	Y	W	I	I	+	2226.4
17:	L	M	L	L	F	V	A	I	I	Y	W	I	I	++	2227.5
18:	M	L	L	V	F	L	A	I	I	Y	W	I	I	-	2227.5
19:	M	L	L	I	F	V	A	I	I	Y	W	I	I	-	2228.4
20:	M	I	V	F	V	F	A	I	I	Y	W	I	I	++	2237.7
21:	M	L	I	M	I	F	A	I	I	Y	W	I	I	++	2258.4
22:	M	I	L	F	L	F	A	I	I	Y	W	I	I	++	2290.1
23:	L	M	I	I	F	F	A	I	I	Y	W	I	I	++	2291.0
	Stage-2 mutants (ascending volume order)														
24:	L	L	L	V	L	V	I	V	L	V	V	M	I	-	2056.4
25:	I	M	L	V	V	L	V	L	L	V	V	L	I	-	2056.4
26:	L	L	L	I	V	M	I	I	V	V	V	V	I	-	2058.2
27:	L	I	L	V	I	L	I	I	M	V	V	V	I	-	2085.3
28:	I	M	I	M	V	V	L	I	V	L	L	V	I	-	2087.3
29:	L	I	M	V	V	V	I	M	L	L	V	I	I	-	2087.3
30:	L	M	M	I	V	V	L	L	V	M	V	L	I	+	2088.4
31:	L	L	V	F	V	V	V	I	L	V	L	L	I	-	2089.0
32:	V	L	V	M	V	F	I	L	L	V	L	V	I	-	2091.9
33:	I	L	L	I	F	V	V	M	L	V	V	V	I	-	2092.8
34:	L	L	L	I	I	V	L	I	L	V	L	V	I	++	2107.7
35:	L	L	L	M	V	V	L	L	L	L	L	V	I	-	2107.9
36:	L	I	L	V	L	L	M	I	V	L	L	V	I	-	2109.7
37:	I	I	L	V	V	L	L	V	L	L	F	V	I	-	2116.1
38:	L	F	V	I	L	V	I	I	V	L	V	L	I	+	2117.0
39:	M	L	F	V	V	L	L	I	V	L	L	V	I	-	2118.1
40:	M	I	L	I	V	V	V	I	V	L	L	F	I	-	2119.9
41:	L	I	L	V	F	V	L	V	V	F	V	L	I	-	2124.5
42:	I	I	L	V	L	L	L	L	V	L	L	L	I	-	2133.0
43:	L	I	L	V	V	F	I	I	L	L	L	V	I	++	2143.2
44:	L	I	L	I	F	L	I	L	V	L	V	V	I	-	2143.2
45:	L	L	I	I	F	L	L	V	V	I	V	L	I	-	2143.2
46:	L	L	L	V	F	V	M	L	V	L	L	L	I	-	2143.4
47:	I	L	I	I	L	L	V	I	L	V	F	V	I	++	2144.1
48:	L	I	L	I	F	V	I	V	L	L	V	I	I	-	2144.1
49:	L	I	L	M	F	V	L	V	V	L	L	L	I	-	2144.3
50:	L	L	L	V	V	L	I	M	L	V	I	F	I	-	2145.2
51:	L	L	M	V	I	L	I	L	L	F	V	V	I	-	2145.2
52:	I	I	L	I	V	F	L	M	V	V	L	L	I	+	2146.1
53:	M	I	L	V	F	V	I	V	M	M	L	L	I	-	2151.0
54:	L	L	F	I	V	V	L	M	V	F	V	L	I	-	2153.6
55:	L	F	V	V	F	V	I	V	M	F	V	L	I	+	2162.9
56:	L	L	M	I	L	I	I	L	F	V	V	L	I	-	2172.3
57:	L	I	M	I	L	V	L	I	I	V	F	L	I	-	2173.2
58:	I	I	M	I	F	I	I	L	L	V	L	V	I	++	2174.1
59:	L	M	L	I	M	L	I	L	F	L	V	V	I	-	2174.3
60:	L	L	L	V	L	V	I	V	L	F	L	F	I	-	2176.9
61:	I	M	L	I	F	V	I	L	V	F	L	V	I	-	2181.6
62:	I	M	M	I	V	V	I	L	F	L	V	F	I	-	2184.5
63:	I	M	L	I	F	V	M	L	V	V	M	F	I	+	2186.5
64:	L	L	I	V	V	V	F	I	V	F	F	M	I	-	2190.0
65:	I	I	I	M	L	L	M	L	L	L	L	L	I	-	2192.1
66:	L	F	L	I	I	V	I	L	L	L	L	L	I	-	2195.6
67:	L	L	L	M	V	L	I	M	M	L	L	F	I	+	2202.5
68:	L	I	I	I	V	V	I	I	L	F	F	L	I	++	2206.7
69:	I	I	M	L	V	L	F	L	V	F	L	L	I	-	2206.9
70:	L	I	M	F	F	L	I	I	V	L	L	V	I	-	2207.8
71:	I	I	L	M	F	F	I	I	L	V	V	L	I	-	2208.7
72:	L	I	M	F	F	F	V	F	L	V	V	V	I	+	2224.6
73:	L	L	I	M	L	V	L	L	F	L	M	F	I	-	2235.1
74:	L	L	I	F	L	F	I	I	F	L	V	L	I	-	2266.6
75:	V	L	L	L	F	L	M	M	F	L	L	F	I	-	2269.7
	Stage-3 mutants (ordered by parentage)														
34.1:	L	L	L	I	I	V	L	I	L	V	L	V	L	++	2106.8
34.2:	L	L	L	I	I	V	L	I	L	V	L	V	F	-	2142.3
40.1:	M	I	L	I	V	V	V	I	V	L	L	F	L	-	2119.0
40.2:	M	I	L	I	V	V	V	I	V	L	L	F	V	-	2092.8
43.1:	L	I	L	V	V	F	I	I	L	L	L	V	L	++	2142.3
43.2:	L	I	L	V	V	F	I	I	L	L	L	V	F	-	2177.8
47.1:	I	L	I	I	L	L	V	I	L	V	F	V	L	++	2143.2
48.1:	L	I	L	I	F	V	I	V	L	L	V	I	L	-	2143.2
48.2:	L	I	L	I	F	V	I	V	L	L	V	I	V	-	2117.0
58.1:	I	I	M	I	F	I	I	L	L	V	L	V	V	++	2147.0
58.2:	I	I	M	I	F	I	I	L	L	V	L	V	L	++	2173.2
63.1:	I	M	L	I	F	V	M	L	V	V	M	F	L	++	2185.6
63.2:	I	M	L	I	F	V	M	L	V	V	M	F	F	-	2221.1
68.1:	L	I	I	I	V	V	I	I	L	F	F	L	L	+	2205.8

Several specific packing constraints are suggested by these data. Among the active first- and second-stage mutants, Phe is significantly overrepresented at positions 10 and 24 (see also text regarding the effect of Phe at position 109). The distribution of codons at some positions deviates significantly from the presumed distribution in the primer population, probably due to sequence-dependent differences in primer annealing during the PCR. The effect is not so pronounced as to prevent the isolated mutants from displaying extreme sequence diversity, and because the base mixtures used preclude any significant bias toward wild-type residues (Fig. 2), the effect is completely unrelated to mutant activity. Consequently, it is accurate to describe the first- and second-stage collections as being random with respect to barnase activity.

*Deduced from complete nucleotide sequences of mutant *synbar* genes obtained as described in *Materials and Methods*.

†Obtained from biological activity screen as described in *Materials and Methods*. Stable active-site mutants were used to gauge the sensitivity of the activity screen. Mutants Glu-73 → Ala and Glu-73 → Tyr test active (++) while lacking the glutamate residue that normally acts as the catalytic general base (20) in the first step of the hydrolysis reaction. Measurement of the RNA hydrolysis rate *in vitro* for Glu-73 → Ala determines it to be 500-fold lower than wild-type barnase (21). A similar activity has been found for Glu-73 → Tyr (unpublished result).

‡Individual residue volumes (22) were summed for each core sequence. No relationship between activity and total residue volume is evident among second-stage mutants, although there appears to be a preference for larger volumes among first-stage mutants.

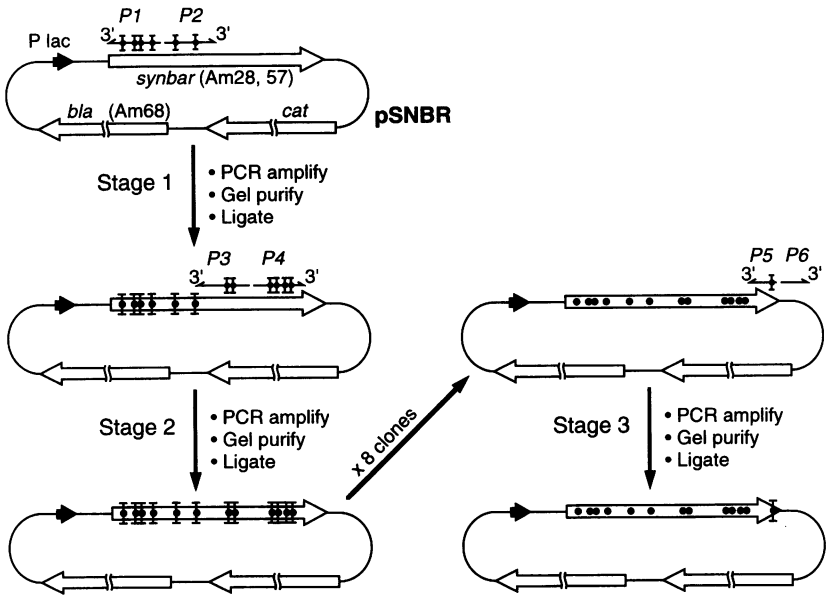

FIG. 3. Three-stage randomization of hydrophobic-core codons in *synbar*. Dots with range bars indicate positions where six or eight codon sequences exist within a mixed DNA population. Dots alone indicate positions where all DNA molecules in a population have the same codon sequence and that sequence was acquired as a result of prior randomization. See *Materials and Methods* for experimental details. Details of primer design are given in Fig. 2.

transformation mixtures. Control transformations were also performed on both strains using plasmid pSNBR and a derivative of this plasmid, pSNBRΔ, that carries a frameshift mutation in *synbar*. After incubation of transformation mixtures at 37°C to allow expression of *cat* and *bla*, 40 μl of each mixture was removed and spread onto a pre-warmed Luria–Bertani agar plate containing antibiotic (chloramphenicol at 25 mg/liter for the sup$^-$ strain and ampicillin at 50 mg/liter for the *supD* strain; for control transformations, both plate types were used for both strains). Plates were inspected after incubation at 37°C for 20 hr. The control transformations confirmed the competence and suppression phenotypes of the two strains: in the sup$^-$ strain, no amp^r colonies but hundreds of chlor^r colonies were present; in the *supD* strain, hundreds of amp^r and chlor^r colonies from pSNBRΔ but none from pSNBR were present. All chloramphenicol plates from core-mutant transformations showed hundreds of colonies. Most ampicillin plates showed either hundreds of colonies or no colonies, with a few showing small numbers of colonies. Clones giving no amp^r colonies were given the activity score ++, indicating extreme autotoxicity. Amp^r colonies were picked (where present) to inoculate 2-ml cultures of 2× TY medium with ampicillin (50 mg/liter) for 24-hr incubation at 37°C in a rotary shaker. Clones showing dense or moderate growth were scored −, indicating little or no autotoxicity. Clones showing very little or no growth (usually corresponding to plates with few or small colonies) were scored +, indicating moderate autotoxicity. The term active in this work refers to clones scored either + or ++. Transformation tests were repeated for verification. To confirm that lack of growth is caused by mutant toxicity*, mutant *synbar* genes were excised by restriction and ligation. Ligation products were then used to transform the *supD* strain. In all cases, this produced numerous amp^r colonies capable of producing dense amp^r liquid cultures.

*Two alternative explanations for a lack of amp^r colonies might be (*i*) accidental inactivation of the *bla* gene by mutation during PCR or (*ii*) coincidental production of an *Eco*K restriction site by a mutation at a core codon in *synbar*. A search for potential *Eco*K sites in a nucleotide sequence including mixed bases demonstrates that the second explanation is not possible. The first explanation is refuted by the described experiment.

RESULTS AND DISCUSSION

Results of sequence analysis and activity tests following all three mutagenesis stages are given in Table 1. Of the 23 mutants analyzed after the first stage, 13 (57%) show unambiguous autotoxicity in the *supD* strain. Because all six core residues projecting from the helix side (Fig. 1) have been randomized, the entire packing interface between the helix and sheet sides is affected. Consequently, any unusual complementarity between the hydrophobic surfaces on the two sides of the native core has been destroyed at this stage. The prevalence of active mutants demonstrates that enzymatic function does not require such complementarity.

Randomization of residues projecting from the sheet side of the core is not expected to diminish further the quality of packing at the interface. The lower proportion of active mutants following the second mutagenesis stage (23%; Table 1) must, therefore, result from deterioration of intrasheet contacts relative to those present in the native sheet. An activity rate of nearly 1 in 4 among second-stage mutants is nonetheless remarkable. Most of the active mutants (7 of 12) have non-wild-type residues at all 12 randomized positions, leaving Ile-109 as the only unchanged core residue. Exquisite native-like packing must, therefore, be unnecessary for rudimentary enzymatic function.

The fact that Ile-109 finds itself in a wide variety of packing contexts in active second-stage mutants makes it implausible that Ile is functionally essential at this position. Screening of third-stage mutants confirms the dispensability of this residue by showing that most 13-position mutants derived from active 12-position mutants retain activity. However, there does appear to be a negative sequence constraint at position 109 that is largely context independent. The three mutants that lost activity at the third stage (clones 34.2, 43.2, and 63.2; Table 1) all have Phe at position 109, suggesting that this bulky residue is unacceptable at this position.

These results demonstrate that barnase function shows profound tolerance toward changes in the hydrophobic core. Before considering the implications this carries for protein evolution and design, it is important to consider the possibility that the activity reported here might be achieved by misfolded structures that bear little resemblance to barnase. Given the sensitivity of the activity screen, some of the mutants scored as active are expected to be

very much less active than wild-type barnase. Might these levels of activity be produced by structures that would be unsuitable as starting points for evolution (natural or experimental) of enzymes that approach the performance of barnase?

Some illumination of this point can be obtained from a collection of single-site barnase mutants (to be described in detail elsewhere) that test inactive by the same screen. One of these is a truncation mutant comprising the first 104 residues of the 110-residue enzyme. A very similar barnase variant, produced by *in vitro* cleavage of a Thr-105 → Met mutant, exhibits native-like structure and function *in vitro* but is considerably less stable than the complete protein (25, 26). The reduced stability evidently renders the truncated protein inactive *in vivo*, presumably due to rapid digestion of loosely folded proteins in the cytoplasm (27). This implies that mutants testing active must have well-formed and reasonably stable structures. Other inactive single-site mutants indicate that the major structural features of barnase, including the active site, the main hydrophobic core, and both smaller cores, must be formed for a mutant to test active.

Thus, the results presented here demonstrate that mere hydrophobicity is nearly sufficient as a criterion for the construction of a barnase core that provides adequate structural integrity for enzymatic function *in vivo*. Of all possible amino acid combinations at the 13 core positions, approximately 1 in 3×10^4 would fully satisfy the hydrophobicity criterion [taking 9 of the 20 amino acids to be hydrophobic, $(9/20)^{13} \approx 1/(3 \times 10^4)$]. Here we find that approximately 1 in 5 core combinations meeting this criterion (or approximately 1 in 2×10^5 overall) supports enzymatic function within the cell. Hence, whatever further criteria may be needed for a sufficient specification of functional barnase cores, they accomplish only a 5-fold reduction in the number of core sequences.

Although complete core replacement will have to be performed in other proteins before these findings can be deemed typical, the fact that barnase responds in a typical way to single core substitutions (1) justifies tentative consideration of this possibility. Viewed in the context of previous hydrophobic-core studies, then, this work suggests that refinement of a crudely functional core entails a substantially greater restriction on the core sequence than does the initial attainment of crude function. For example, suppose there is one hydrophobic alternative to each native barnase core residue such that any or all native residues could be replaced by the alternatives without affecting function. It follows that approximately 1 in 3×10^8 barnase core sequences that satisfy the hydrophobicity criterion would be functionally equivalent to the wild-type core $[(2/9)^{13} \approx 1/(3 \times 10^8)]$. This would mean that barnase cores providing wild-type activity are >1000-fold less prevalent within a pool containing all fully hydrophobic core sequences than cores providing rudimentary activity are within a pool containing all possible core sequences $(3 \times 10^8/2 \times 10^5 > 10^3)$. The number of core sequences that are functionally equivalent to the wild type may well be overestimated in this example [experiments detecting relatively subtle effects on repressor function (7, 8)] suggest that very few core substitutions have no adverse effect), in which case the disparity between the sequence constraints accompanying attainment and refinement of enzymatic activity would be even greater than this calculation suggests.

That the requirements for attainment of a functional core are relatively lax has important implications for protein evolution. The considerable optimization power of biological selection (13–15) requires a pre-existing structure exhibiting a phenotypically significant level of activity. The presence or absence of such starting structures is, therefore, critical in determining the evolutionary course of any system undergoing replication with random mutation. If protein folds required unique core sequences for function, starting structures for evolutionary optimization of noncore residues would be profoundly rare: for a 13-residue core, only 1 in 20^{13} ($\approx 10^{17}$) random sequences would qualify. Relative to this, the core permissiveness demonstrated here amounts to an increase by 12 orders of magnitude in the prevalence of starting structures.

The implication of this permissiveness for protein design is that it should indeed be possible to achieve low-level enzymatic activity from *de novo* designs that disregard the packing problem. The "binary-code" design approach (28), in which the final design consists simply of a sequence of unspecified hydrophobic and polar residues, finds support in an analysis of the globin protein family (29) and in lattice-model studies (30–32), and has been applied in the design of a four-helix bundle (28). The behavior of barnase suggests that, for core design at least, this approach may prove useful in the development of sophisticated functional folds.

We thank G. Schreiber for the plasmid used to produce the Glu-73 → Tyr mutant, K. Wong and V. Arcus for assistance with MOLSCRIPT, and the Office of Naval Research and Medical Research Council for financial support.

1. Richards, F. M. & Lim, W. A. (1994) *Q. Rev. Biophys.* **26,** 423–498.
2. Karpusas, M., Baase, W. A., Matsumura, M. & Matthews, B. W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8237–8241.
3. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992) *Science* **255,** 178–183.
4. Kellis, J. T., Nyberg, K., Sali, D. & Fersht, A. R. (1988) *Nature (London)* **333,** 784–786.
5. Sandberg, W. S. & Terwilliger, T. C. (1989) *Science* **245,** 54–57.
6. Shortle, D., Stites, W. E. & Meeker, A. K. (1990) *Biochemistry* **29,** 8033–8041.
7. Lim, W. A. & Sauer, R. T. (1991) *J. Mol. Biol.* **219,** 359–376.
8. Milla, M. E. & Sauer, R. T. (1995) *Biochemistry* **34,** 3344–3351.
9. Betz, S. F., Raleigh, D. P. & DeGrado, W. F. (1993) *Curr. Opin. Struct. Biol.* **3,** 601–610.
10. Kamtekar, S. & Hecht, M. H. (1995) *FASEB J.* **9,** 1013–1022.
11. Raleigh, D. P., Betz, S. F. & DeGrado, W. F. (1995) *J. Am. Chem. Soc.* **117,** 7558–7559.
12. Hecht, M. H. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8729–8730.
13. Chen, K. Q. & Arnold, F. H. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5618–5622.
14. Bartel, D. P. & Szostak, J. W. (1993) *Science* **261,** 1411–1418.
15. Wilson, C. & Szostak, J. W. (1995) *Nature (London)* **374,** 777–782.
16. Lim, W. A. & Sauer, R. T. (1989) *Nature (London)* **339,** 31–36.
17. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
18. Hartley, R. W. (1988) *J. Mol. Biol.* **202,** 913–915.
19. Hirel, P.-H., Schmitter, J.-M., Dessen, P., Fayat, G., & Blanquet, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8247–8251.
20. Day, A. G., Parsonage, D., Eble, S., Brown, T. & Fersht, A. R. (1992) *Biochemistry* **31,** 6390–6395.
21. Mossakowska, D. E., Nyberg, K. & Fersht, A. R. (1989) *Biochemistry* **28,** 3843–3850.
22. Chothia, C. (1975) *Nature (London)* **254,** 304–308.
23. Joris, B., Ghuysen, J.-M., Dive, G., Renard, A., Dideberg, O., Charlier, P., Frère, J.-M., Kelly, J. A., Boyington, J. C. & Moews, P. C. (1988) *Biochem. J.* **250,** 313–324.
24. Chung, C. T., Niemela, S. L. & Miller, R. H. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2172–2175.
25. Vuilleumier, S., Sancho, J., Lowenthal, R. & Fersht, A. R. (1993) *Biochemistry* **32,** 10303–10313.
26. Kippen, A. D. (1994) Ph.D. thesis (Univ. of Cambridge, U.K.).
27. Parsell, D. A. & Sauer, R. T. (1989) *J. Biol. Chem.* **264,** 7590–7595.
28. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993) *Science* **262,** 1680–1685.
29. Bashford, D., Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196,** 199–216.
30. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1229–1233.
31. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22,** 3986–3997.
32. Dill, K. A., Bromberg, S., Yue, K. Z., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995) *Protein Sci.* **4,** 561–602.
33. Buckle, A., Henrick, K. & Fersht, A. R. (1993) *J. Mol. Biol.* **234,** 847–860.