

A Search for Single Substitutions That Eliminate Enzymatic Function in a Bacterial Ribonuclease[†]

Douglas D. Axe,* Nicholas W. Foster, and Alan R. Fersht*

Cambridge University Chemical Laboratory and Cambridge Centre for Protein Engineering, MRC Centre, Hills Road, Cambridge CB2 2QH, U.K.

Received February 19, 1998

ABSTRACT: Exhaustive-substitution studies, where many amino acid replacements are individually tested at all positions in a natural protein, have proven to be very valuable in probing the relationship between sequence and function. The broad picture that has emerged from studies of this sort is one of functional tolerance of substitution. We have applied this approach to barnase, a 110-residue bacterial ribonuclease. Because the selection system used to score barnase mutants as active or inactive detects activity down to a level that can be approached by nonenzyme catalysts, mutants that test inactive are essentially devoid of enzymatic function. Of the 109 barnase positions subjected to substitution, only 15 (14%) are vulnerable to this extreme level of inactivation, and only 2 could not be substituted without such inactivation. A total of 33 substitutions (amounting to 5% of the explored substitutions) were found to render barnase wholly inactive. The profoundly disruptive effects of all of these inactivating substitutions appear to result from either (1) replacement of a side chain that is directly involved in substrate binding or catalysis, (2) replacement of a substantially buried side chain, (3) introduction of a proline residue, or (4) replacement of a glycine residue. Although substitutions of these types are functionally tolerated more often than not, the system used here indicates that only these sorts of substitution are capable of single-handedly reducing catalytic function to, or nearly to, levels that can be achieved by nonenzyme catalysts.

It is hoped that investigations of protein folding will ultimately yield a comprehensive understanding of the relationship between protein sequence and structure. Although this is an ambitious undertaking, it is only half of the larger effort aimed at elucidating the relationship between sequence and function. The other half, aimed at understanding the structure–function relationship, is no less ambitious.

While a general solution to the grand problem of relating protein sequence to function will clearly be some time in the making, we do have at our disposal, in the form of natural proteins, thousands of specific solutions to this problem. It therefore makes good sense for us to glean as much raw data as we possibly can from these natural solutions. Given the complexities and subtleties of the sequence–function relationship, it also makes sense for us to collect these data in a manner that is entirely unbiased by any a priori expectations we may hold regarding the nature of that relationship.

The simplest and most direct way to obtain such an unbiased data set is to produce and test a large collection of mutant proteins where all possible single-residue substitutions are represented. This might be termed the exhaustive-substitution approach. Since all single-replacement possibilities are examined by this approach, the data lack any imprint of the experimenter's expectations, and they are complete in the sense that the entire molecule is examined

(there is no question that more information would be gained by examining all possible double or triple substitutions, but such increases in the level of substitution lead very quickly to impractically large numbers of mutants).

A number of studies have captured the essence of this approach (1–8). One of the broad features to emerge from these studies (1, 4, 5, 7) and others (9, 10) is a positive correlation between the degree of solvent exposure at an amino acid position and the level of substitutional tolerance at that position. It has also become clear that natural proteins typically tolerate single substitutions, even nonconservative ones, at most positions without complete loss of function. A peculiar exception to this is the phage P22 Arc repressor, which was found to tolerate nonconservative substitutions at only 8 of its 53 positions (2). Perhaps the unusual behavior of this protein can be attributed in part to its unusually small size.

Another factor, one that clearly affects the results of any exhaustive-substitution study, is the activity threshold, the minimum level of activity necessary for a mutant to be scored active. Because of the large number of mutants involved, these studies typically rely upon rapid *in vivo* screens that produce a binary (i.e., “active” or “inactive”) indication of activity. It is generally possible to set the activity threshold at various levels within some range, the choice being more a matter of experimental convenience than necessity. For any experimental protein, a high threshold will lead to a more inclusive list of positions deemed to be functionally important than will a low threshold. In previous studies, thresholds

[†] This work was supported in part by a grant from the Office of Naval Research.

* Corresponding authors. Fax: 1223 402140.

have been in the range of 3–30% of wild-type (WT)¹ activity (1, 2, 4–6).

Barnase, a bacterial ribonuclease, provides an opportunity to apply the exhaustive-substitution approach using a very low activity threshold. The extreme autotoxicity of this enzyme (11) allows direct selection of mutants with very low (approximately 0.1% of wild-type and lower) activities (12), thereby enabling us to directly identify residues that profoundly influence enzyme function.² Here we report the results of an experiment designed to identify all single-base missense mutations that affect barnase function to this extent.

MATERIALS AND METHODS

Strains and Plasmids. The *Escherichia coli* strains and the plasmid used in this work have been described previously (12). Plasmid pSNBR carries a synthetic barnase gene, *synbar*, that is interrupted by two amber stop codons. Strain C-1a, a nonsuppressing strain, is used to prepare pSNBR DNA. Strain MX383, a suppressing strain, reads amber stop codons as serine codons, causing a full-length product to be produced.

Mutagenesis of *synbar*. The number of possible single-position substitutions for a protein the size of barnase is large enough to make individual preparation, sequencing, and testing of all single mutants impractical. Various approaches, all with strengths and weaknesses, have been employed in previous studies to overcome this difficulty. Three constraints were dominant in our choice of a method for mutagenesis. First, since this is a study of the effects of single substitutions, we required that the chosen method primarily produce singly substituted variants. Second, since the *synbar* selection system selects clones producing inactive barnase variants, we needed a method that would minimize accidental introduction of frame-shift mutations (virtually all of which would pass selection). And third, since a particular sort of amber suppression (encoded by the *supD* allele) is an integral part of the *synbar* selection system (12), methods involving a large set of strains with various amber-suppression phenotypes (1) could not be used.

These constraints can be adequately satisfied by a method that uses mutagenic oligonucleotides containing suitably small amounts of “contaminating” bases. By incorporating these oligonucleotides in a manner that avoids blunt-end ligations, we are able to achieve a low background frequency of frame-shift mutations. The primary limitation to this approach is the fact that it restricts the substitution set at a particular position to the amino acids that can be specified with a codon that differs only by one base from the wild-type codon. For *synbar*, this means that the number of accessible substitutions per position ranges from 5 to 7 (average = 6.2). Although most substitutions are inaccessible by this method, the number and variety of accessible substitutions ensure that multiple nonconservative substitu-

tions (in addition to more conservative ones) will be possible at all positions. This will enable us to obtain the desired information.

The coding region of *synbar* was conceptually divided into 8 contiguous regions covering from 12 to 14 codons each, starting from the second codon.³ With each region being treated separately, random base substitutions were introduced throughout the gene by using oligonucleotides prepared with mixed phosphoramidites. For each region, an oligonucleotide was synthesized that spanned the entire region and extended 10 bases beyond in both directions. The 10-base extensions were synthesized so as to perfectly complement the corresponding regions on the template plasmid, pSNBR. The central portion of each oligonucleotide, corresponding to one of the 8 *synbar* regions, was synthesized with small concentrations of contaminating phosphoramidites to introduce base substitutions at a low frequency. This was achieved by preparing the four standard phosphoramidite solutions at double concentrations and transferring 118 μ L from each of these bottles to a fifth bottle containing 20 mL of pure anhydrous acetonitrile. The synthesizer was programmed to draw only from the pure bottles for the first and last 10 base positions of each oligonucleotide but to draw equal volumes from the appropriate pure bottle and the mixed bottle at each of the central positions. The resulting product is a mixed population of oligonucleotides where each particular mutation occurs as frequently in isolation as it does in combination with other mutations (the latter being undesirable for this work).

In separate PCRs, each of the 8 mutagenic oligonucleotides was used with a single biotinylated oligonucleotide to amplify a large portion of plasmid pSNBR (Figure 1). After purification from agarose gels, the product DNA was methylated by incubation with *E. coli* Dam methylase. The nonbiotinylated strands were then separated from the biotinylated strands by using Dynabeads M-280 Streptavidin (Dyna) according to the manufacturer's protocol. The nonbiotinylated mutant strands were retained for producing mutant plasmid clones. Two additional oligonucleotide primers, one biotinylated, were used to amplify a portion of pSNBR such that the nonbiotinylated strand from this product can anneal to any of the nonbiotinylated mutant strands to form a gapped-duplex plasmid molecule (Figure 1). After isolation of the nonbiotinylated strand as before, the gapped-duplex product was prepared by combining this strand with each mutant strand (in equal proportions), heating to 75 °C for several minutes, and allowing the mixtures to cool slowly to room temperature.

The *synbar* Selection System. *E. coli* strain MX383 was transformed directly [by the previously described protocol (12)] with the mixtures of gapped-duplex DNA. Dam methylation of the mutant strand causes the cell to use this strand as the reference in correcting mismatches (14), thereby ensuring that *synbar* mutations are preserved. Because of the extreme autotoxicity of *synbar* expression (11, 12), only mutant genes producing largely inactive barnase variants

¹ Abbreviations: fMet, *N*-formylmethionine; FMOC, 9-fluorenylmethoxycarbonyl; HPLC, high-performance liquid chromatography; PCR, polymerase chain reaction; RNase, ribonuclease; WT, wild-type.

² To avoid confusion, it should be emphasized that throughout this work the terms inactive, inactivating, functional, sensitive, tolerant, and related words refer to properties of barnase or its variants (often with respect to particular amino acid positions), or to the effects of amino acid substitutions on barnase, but not to the bacterial host or to the effects of barnase variants on the host cell.

³ Codon and residue numbering correspond to the sequence of mature wild-type barnase, where the N-terminal residue is Ala. The N-terminal fMet resulting from *synbar* expression is expected to be removed within the cell (13), yielding the desired wild-type sequence. Since fMet removal depends on the identity of the adjacent residue, we have left the Ala codon (codon 1) undisturbed.

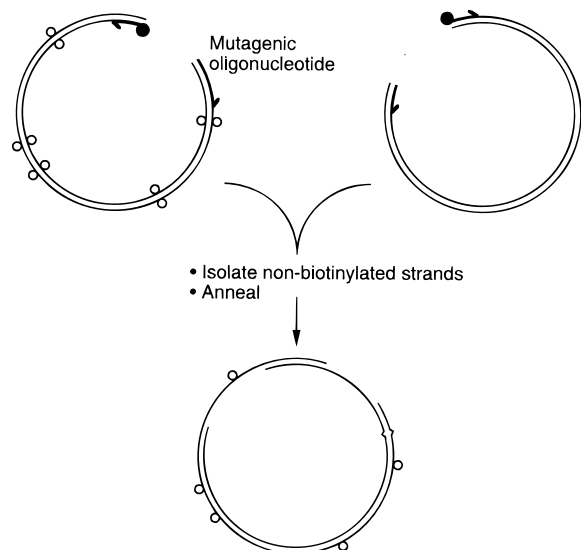


FIGURE 1: Introduction of random base substitutions into *synbar*. The upper illustrations depict the two PCRs used to produce the DNA strands that are combined to form the gapped-duplex product (lower illustration). Plasmid pSNBR serves as the template in both PCRs. Bold half-arrows indicate primers. Filled circles represent biotin groups that are covalently attached to primers. Open circles indicate DNA that has been methylated by Dam methylase. A base substitution in *synbar* results in local mispairing in the final product (as shown).

allow growth of the host cell. Consequently, inactivating mutations are directly selected by plating on Luria-Bertani agar with ampicillin (12).

Sequencing of *synbar* Mutants. Plasmid DNA was prepared from overnight cultures of clones passing selection. Sequences of mutant genes were determined by performing cycle-sequencing reactions with dye-labeled dideoxynucleotides (Perkin-Elmer) and analyzing products with an automated sequencer (Perkin-Elmer, model ABI 373). Clones with multiple substitutions or frame-shift mutations were excluded from our analysis.

Determination of OVA1 Activity. RNase activity has been reported for OVA1, a 15-residue peptide (NVMEERKIKVIL-PRM) corresponding to a portion of chicken ovalbumin (15). To perform a quantitative activity assay, the peptide was synthesized on a commercial synthesizer using Fmoc chemistry and purified by HPLC. RNA-hydrolysis activity was determined by measuring the absorbance (301 nm) of a solution consisting of 100 mM Tris buffer (pH 7.6 at 25 °C), 200 mM KCl, 5 mM MgCl₂, and 2.8 mg/mL torula yeast RNA (type VI, Sigma), then adding OVA1 (to 30 μM), and monitoring the decrease in absorbance as a function of time. As a reference, a parallel reaction was performed by adding wild-type barnase to the same assay buffer. The relative activity of OVA1 was calculated from the ratio of the absorbance slopes of OVA1 and barnase during the initial steady-state phase of hydrolysis.

RESULTS

Analysis of Completeness. Figure 2 summarizes our findings, indicating all single mutants found to be inactive and all single mutants inferred to be active. Because a limited number of trials were used to identify inactivating substitutions, some such substitutions may, by chance, have escaped detection. Consequently, before the implications of

these results are considered, it is important to consider how closely they represent the ideal data set that would result from an infinite number of trials.

One indicator of completeness is the proportion of identified inactivating substitutions for which only one example has been isolated. A very incomplete collection would be dominated by these unduplicated examples, whereas they would become rare as the collection approaches completeness. Our initial plan was therefore to continue the process of collecting and examining mutants until duplicates of all examples had been obtained. However, the highly nonuniform distribution of mutants following selection made it impractical to achieve this.

The problem we encountered is illustrated in Figure 3, which shows all mutants recovered from one of the 8 contiguous *synbar* regions. Codons 83 and 87 normally specify arginine residues. The base mixtures used to prepare the mutagenic oligonucleotide mixture (see Materials and Methods) are expected to give unaltered Arg codons in most plasmid molecules, with a small fraction of plasmids carrying a missense mutation at either of these codons. These mutant codons are expected to specify Cys, Ser, Gly, Leu, Pro, and His with equal frequency. Consequently, if all mutants passing selection were equally inactive, we would recover these inactive mutants at roughly equal frequencies. The fact that frequencies of recovery are highly nonuniform (Figure 3) suggests that significant variation in activity exists even among these inactive mutants.⁴ Variation of this sort was sufficiently common in a previous study (12) that a third classification was defined for mutants that impaired cell growth without completely preventing it.

Despite varying degrees of inactivity, however, the 64 inactive-mutant isolates represented in Figure 3 clearly demonstrate that positions 83 and 87, where all accessible substitutions lead to inactivation, are considerably more important for barnase function than any of the other positions in the region depicted. While we cannot be certain that an inactivating substitution at position 93, for example, would not be found if another 100 clones carrying mutations in this region were processed, we can be certain that most of these clones would carry substitutions at positions 83 or 87, and we can safely deem it highly unlikely that position 93 is actually as sensitive to substitution as positions 83 and 87 are. We can likewise be confident that positions showing limited sensitivity to substitution (positions 89–91) are less functionally critical than the highly sensitive positions. Moreover, because the identities of the inactivating substitutions found at these three positions are readily explicable in terms of the properties of the introduced side chains (at position 89, normally occupied by a well-buried leucine, Arg is the only accessible substitution that introduces a charged side chain; similarly, Asp is the only charged side chain of those accessible at position 90, normally occupied by a largely buried tyrosine; Pro introduces unusual backbone

⁴ The biological interpretation of the observed nonuniformity is that mutants having activities very close to the threshold level may or may not kill the original transformed host cell, depending upon whether the cell is able to make the necessary metabolic adjustments to compensate for the harmful RNase activity. Cells that do make the adjustments form colonies [often smaller than normal colonies (12)], but these colonies will be underrepresented to the extent that the survival rates of the initial transformants are reduced.

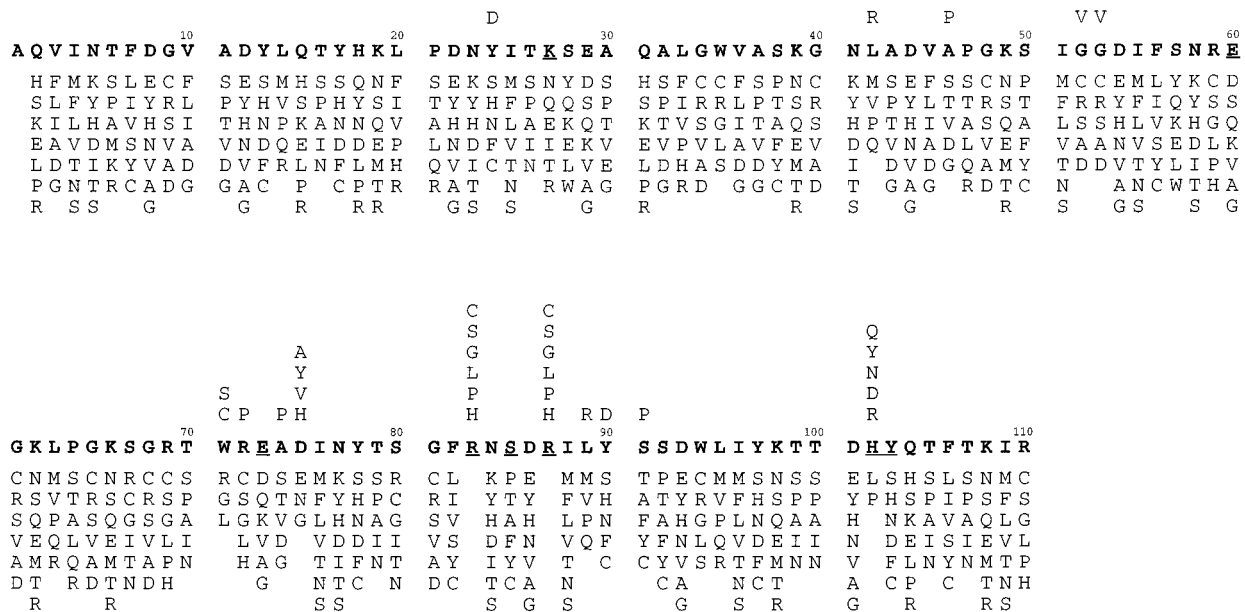


FIGURE 2: Summary of single-substitution results for barnase. The wild-type barnase amino acid sequence is shown in bold type, underlined letters indicating residues with side chains that interact directly with substrate (16). At each position, substitutions that were found to render the enzyme inactive are shown above the wild-type sequence. All other accessible substitutions (see Materials and Methods) are shown below the wild-type sequence.

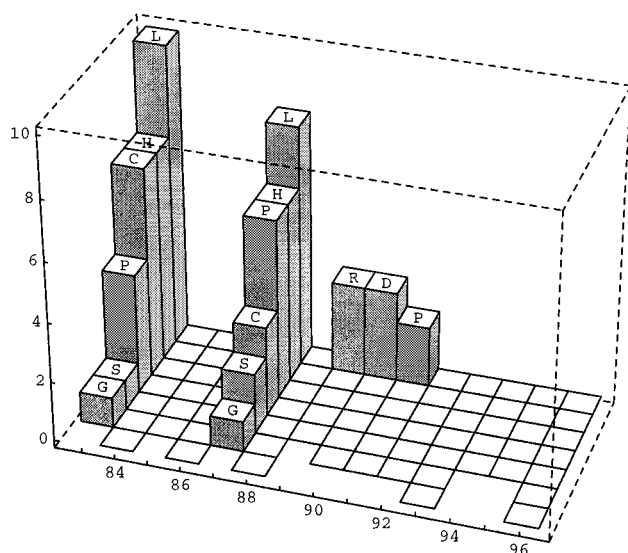


FIGURE 3: Number of independent isolates of inactive barnase variants carrying substitutions from position 83 to position 96. The residue position is indicated on the horizontal axis. Vertical stacks indicate the number (vertical scale) of examples recovered for each accessible substitution in this region. Stacks corresponding to substitutions for which at least one example was recovered are labeled to indicate the introduced amino acid (see Figure 2 for accessible substitutions that were not recovered).

constraints at position 91, normally occupied by a serine situated in the middle strand of a five-strand β -sheet), our approach appears to effectively identify the most severely disruptive substitutions at partially sensitive positions.

Since our primary aim is to identify the residues in wild-type barnase that most critically affect function, the overall substitutional sensitivity of each position is more important than the activity of any particular mutant. Some uncertainty at the level of individual substitutions can be accepted without losing the bigger picture at the level of amino acid positions because several substitutions are possible at each position. In light of this, a more appropriate measure of

completeness is the proportion of sensitive positions for which only a single instance of recovering an inactive variant occurred. By this measure, the data set is seen to be complete; multiple examples (averaging 8.6 per position) were isolated for all of the 15 positions found to be vulnerable to inactivating substitution.

Paucity of Inactivating Substitutions. The most striking aspect of the results presented in Figure 2 is the scarcity of inactivating substitutions. Of the 109 positions examined, 94 (86%) appear to tolerate all accessible substitutions, and only two (1.8%) are wholly intolerant of accessible substitutions. As discussed above, it is quite possible that some additional inactivating substitutions would be found if the process of mutant collection and examination were continued indefinitely. If so, this would decrease somewhat the fraction of positions showing complete tolerance. Furthermore, since our method of mutagenesis restricts the substitution set to about 6 substitutions per position, it is highly probable that some inactivating substitutions are inaccessible by this method. However, the overall picture of substitutional tolerance depicted in Figure 2 is not apt to be very different from the true picture (see above). In particular, the number of positions that are wholly intolerant of substitution is more likely to be lower (because of the restricted substitution set) than higher.

DISCUSSION

Extreme tolerance of substitution is not without precedent in studies of this kind. In a study of bacteriophage T4 lysozyme, Rennell et al. (4) found that more than half of the positions (55%) tolerate all of the 12 or 13 substitutions tested. More remarkably, only one position of the 163 examined in that study (0.6%) was found to be wholly intolerant of substitution. In another study, Wen et al. found that 75% of the 121 positions in a bacterial membrane protein tolerate nonconservative substitutions (6). Although functional tolerance of substitution is clearly an important theme

to emerge from exhaustive-substitution work, counterexamples do exist. The most striking of these, phage P22 Arc repressor, was noted above, as was the importance of the activity threshold in determining the outcome of an exhaustive-substitution study.

A more thorough consideration of the significance of the activity threshold will be instructive at this point. We noted previously that experimental systems used in exhaustive-substitution studies typically allow the experimenter to choose a threshold level from a wide range of feasible values. This raises the question as to whether the distinction between active and inactive is actually arbitrary or whether there is a natural fixed reference point by which these terms might be defined.

Significance of the Activity Threshold

Distinctive Mechanism of Enzymes. Although DNA-binding proteins have been the subject of a number of important exhaustive-substitution studies (1, 2, 7), we will narrow our focus here to enzymes. This class of molecules catalyzes chemical reactions by doing what all catalysts do, namely, lowering the free energy of the transition-state complex. What distinguishes enzymes from other catalysts is the way in which they accomplish this, and consequently the magnitude of their effect. Unlike simple catalysts, enzymes employ a spatially extensive structure to bind reactants, placing them in a geometrically precise and catalytically optimal orientation (17, 18) relative both to each other (where multiple reactants are involved) and to the catalytic group or groups, which are typically integral to the enzyme.

If separated from their protein scaffold, the catalytic groups alone may catalyze the same reaction in solution by, for example, simple acid or base catalysis. To demonstrate the purpose of the geometric scaffold, however, one need only compare rates of catalysis by small molecules under physiological conditions to rates of enzymatic catalysis for the same reactions. Rates of single-substrate reactions of biological relevance vary by many orders of magnitude when measured in neutral aqueous solutions in the absence of enzymes; first-order rate constants range from 10^{-1} to 10^{-16} s^{-1} for the set of reactions discussed by Radzicka and Wolfenden (19). In contrast, effective second-order rate constants (k_{cat}/K_m) for the corresponding enzymatic reactions appear to fall within a relatively narrow range, a typical value being 10^7 $s^{-1} M^{-1}$ (19). Using this value and a typical cytoplasmic concentration of 10^{-6} M for an enzyme, we estimate a typical pseudo-first-order rate constant for enzymatic reactions in vivo to be 10 s^{-1} . This can be compared directly to the range of nonenzymatic rate constants given above,⁵ indicating that the geometric role of the protein scaffold increases reaction rates by some 2–17 orders of magnitude, depending upon the reaction.

Basal Activity as a Natural Unit of Measure. So universal and so crucial is this geometric aspect of enzymes that it might be viewed as a defining property of this class of molecules. It follows that a natural basal limit to enzyme activity would be a catalytic rate just above that which can be obtained without the spatial positioning employed by enzymes (i.e., the maximal rate of catalysis by a nonenzymatic mechanism defines a limit that can be exceeded only

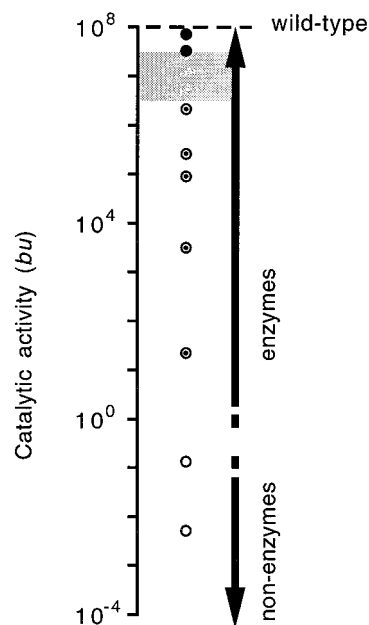


FIGURE 4: Natural scale of activity for a typical enzyme-catalyzed reaction. The scale indicates catalytic activity in terms of basal enzymatic activity units (*bu*), as discussed in the text. In this example, the basal level of activity is 8 orders of magnitude below the activity of the wild-type enzyme. Circles indicate activities of hypothetical mutant enzymes. Open circles correspond to mutants that fail to outperform nonenzyme catalysts and thus function as nonenzymes. All other mutants exceed the basal enzymatic activity level (1 *bu*) and thus function as enzymes. The shaded box indicates a range of activity thresholds from 3 to 30% of wild-type activity. Activity thresholds in the shaded region allow efficient enzymes (filled circles) to be distinguished from less efficient enzymes (circled dots) and nonenzymes, but they do not allow nonenzymes to be distinguished from enzymes.

by employing spatial positioning in the manner that is characteristic of enzymes). As discussed above, enzymes typically exceed this basal limit by many orders of magnitude, achieving catalytic perfection in some cases (19, 22, 23).

The scale of catalytic activity shown in Figure 4 uses basal activity as the unit of measure (1 basal enzymatic activity unit, *bu*, corresponds to the basal level of activity described above). The activity of the wild-type enzyme is taken to be 8 orders of magnitude higher than the basal level (i.e., 10^8 *bu*) so a typical enzyme might be represented. Activities of mutants resulting from single amino acid substitutions will span a wide range, from very close to the wild-type level to

⁵ Although Radzicka and Wolfenden (19) report intrinsic aqueous rate constants (where water is the only catalyst), these values are generally indicative of the level of catalysis that can be expected in micromolar aqueous solutions of nonenzyme solutes. The rate at which small-molecule solutes perform simple acid or base catalysis depends primarily upon their pK_a and concentration (see, for example, the discussion of the Brønsted equation in ref 20). In neutral aqueous solutions at room temperature (for the purpose of comparing them to enzymes), water is expected to have a more significant catalytic effect than any small solute present at micromolar concentrations, regardless of its pK_a . Like small molecules, macromolecules can act as nonenzyme catalysts, often with the added advantages of substrate binding and local solvent exclusion. For example, Hollfelder et al. have demonstrated that the Kemp elimination is fortuitously catalyzed by serum albumins (21). At an albumin concentration of 1 μ M, however, their data indicate that catalysis by water is comparable to catalysis by the albumin (pH 8.0 and 25 °C). At micromolar concentrations, then, even these more sophisticated catalysts typically fail to outperform water significantly.

the basal level and lower, depending upon the nature of the change they introduce. The natural significance of 1 *bu* activity, however, provides a meaningful division of this range into two regions; mutants with activities above 1 *bu* continue to exceed the performance of nonenzyme catalysts, whereas mutants with lower activities do not. Thus, even if these less active mutants continue to bind substrate molecules specifically, they fail to bind them in such a way as to enhance catalysis, and they consequently fail to exhibit the characteristic property of enzymes. Conversely, mutants having greater than 1 *bu* activity, however suboptimal they may be, continue to exhibit the characteristic property of enzymes.

Essential versus Refining Structural Features. In this sense, structural features removed upon substitution could be classed as refining features if the activity following substitution exceeds 1 *bu*. Essential features would then be those structural features that, upon removal, reduce activity to less than 1 *bu*.⁶ Note that the distinction has significance because of the qualitative difference between enzyme catalysis and nonenzyme catalysis, and not because of the quantitative difference per se. Indeed, for mutants in the vicinity of 1 *bu* activity, the quantitative differences in activities are relatively small. The qualitative distinction, however, remains important: where reversion of single mutants is concerned, the restoration of a refining feature can turn a poor enzyme into a good one, but it cannot turn a nonenzyme into an enzyme.

In exhaustive-substitution studies, the activity threshold effectively divides all single mutants into two classes according to activity (above threshold = active; below threshold = inactive). Although estimates of basal enzymatic activities are not generally made in the course of these studies, in most cases the relative proximity of the threshold to the activity of the wild-type enzyme implies that the threshold is several orders of magnitude above 1 *bu*. For example, in the study of T4 lysozyme by Rennell et al. (4), the threshold is placed at 3% of wild-type activity, while in the study of β -lactamase by Huang et al. (5), it is placed at about 30% of wild-type activity. This range of threshold levels is indicated in Figure 4. The wild-type activities of these two enzymes may be somewhat higher or lower than the value represented in Figure 4 (10^8 *bu*), but they are not apt to be many orders of magnitude lower. Consequently, we infer that the threshold levels in these studies are very much higher than the basal level of activity. These thresholds, then, would enable the researcher to distinguish refining features of relatively small effect from all other features (both refining and essential), but they would not be useful for distinguishing essential features from refining features.⁷ To

do this, one would need a system with an activity threshold in the vicinity of 10^0 *bu*.

Essential Features Are Particularly Relevant to Protein Design. In the pursuit of a complete understanding of the relationship between sequence and function, studies of refining features are no less relevant than studies of essential features. Essential features, however, may be of particular importance to efforts in protein design. A reasonable approach to the design of a novel enzyme would be to aim for a rudimentary enzyme as the initial target and then to apply iterative mutation–selection methods (25–28) to optimize the crude initial design. Since the sorts of features that are important for the function of natural enzymes will presumably have to be incorporated into any successful designed enzyme, one might hope to facilitate the first stage of the design process by using information from exhaustive-substitution studies of natural proteins to guide the initial design.

One might even be tempted to view an exhaustive-substitution study of a natural enzyme as an exercise aimed at determining the features that would need to be incorporated into a successful re-design of the same enzyme. However, an important limitation of exhaustive-substitution data in this regard is that they cannot be expected to reliably identify unimportant features (i.e., features that would not need to be included in a successful design). The reason for this is that the structural context in which single substitutions are made in an exhaustive-substitution study is that of the wild-type enzyme. Since the thermodynamic stability of natural proteins typically exceeds that which is necessary for function, we would expect there to be many destabilizing substitutions that are functionally tolerated when introduced in the context of the wild-type sequence. When the context is far less optimal, as would be the case for a crude initial design, the same sorts of substitutions might easily turn a weakly functional design into a nonfunctional design.

Can exhaustive-substitution studies be used to identify features that would need to be included in a successful design? In answering this, we will consider refining features and essential features separately. Upon removal of a refining feature, a wild-type enzyme is simply transformed into a suboptimal enzyme. Though suboptimal, this enzyme may still be considerably more active than a rudimentary enzyme of the sort we might hope to create from an initial design. That being the case, it is quite possible that the refining feature in question may only serve as a refinement in the context of other more basic refinements. In a crude enzyme lacking these basic refinements, we cannot assume that this feature would have any significant effect on function. Refining features therefore do not generally provide useful information for the design of rudimentary enzymes.

On the other hand, it seems inescapable that an essential feature of a wild-type enzyme will have a corresponding essential feature in any less optimal variant of that enzyme. That is, if some type of structural modification destroys enzyme function when it was initially optimal, it is difficult to imagine how the same type of modification would not have the same effect when applied to a suboptimal variant. An essential feature, then, points to a structural rule for the class of proteins that perform a particular function by means of a particular fold. Exhaustive-substitution data would therefore be of considerable value in obtaining design rules, provided

⁶ It is often appropriate to view amino acid substitutions as introducing features as well as (or instead of) removing them. For simplicity of expression, however, we are using the term “feature” in a broad sense to include both the presence and the absence of particular aspects of structure. The absence of a β -carbon at position 52, for example, is a feature of wild-type barnase that is “removed” upon substitution at that position.

⁷ A deletion study on the A chain of ricin (24) used a sensitive activity test capable of detecting activity down to 0.01% of the wild-type value. However, the extreme specificity of the reaction (cleavage of the N-glycosidic bond of a single adenosine base in the mammalian ribosome) suggests that the rate corresponding to 1 *bu* would be many orders of magnitude lower still.

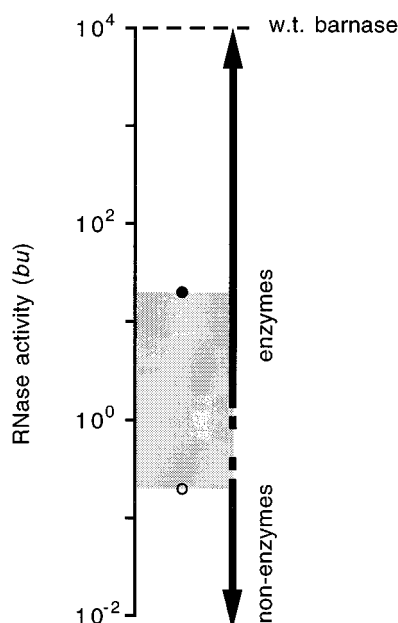


FIGURE 5: Estimation of the activity threshold for the *synbar* selection system. The open circle indicates the RNA-hydrolysis activity of OVA1, a 15-residue peptide corresponding to a portion of chicken ovalbumin (15). Because OVA1 has an unusually high activity for a nonenzyme (0.002% of that of wild-type barnase), we take it to represent an optimal or near-optimal nonenzyme catalyst. Taking basal enzymatic activity to be somewhat higher than the activity of OVA1, we define the basal enzymatic activity unit as $1 \text{ bu} \equiv 0.01\%$ of the wild-type activity. On this natural scale, the activity of OVA1 is $2.0 \times 10^{-1} \text{ bu}$, and the activity of barnase mutant E73A (filled circle) is $2.0 \times 10^1 \text{ bu}$ (29). As indicated by the shaded region, the activity threshold for the *synbar* selection system lies between these two values.

that a basal activity threshold is used to distinguish essential features from refining ones.

It should be noted, though, that essential features and design rules are different things, in that the former is equivalent to a sequence constraint that applies to a particular protein, whereas the latter is a more general constraint that applies to all possible sequences sharing the fold and function of that protein. While identification of essential features for a particular protein provides valuable information on the design rules for the whole class of proteins, these rules cannot necessarily be deduced from experiments on a single protein. Some amount of interpretation will therefore be necessary for tentative classwide design rules to be inferred from exhaustive substitution data.

The synbar System Approaches a Basal-Activity Selection System. To perform an exhaustive-substitution experiment with a basal-activity threshold, one would need a simple screening or selection procedure capable of detecting activity down to a level that approaches nonenzymatic activity. This presents considerable practical difficulties that may be insurmountable for many systems. The *synbar* selection system described here is one system where the necessary sensitivity appears to be attainable, or nearly so.

Figure 5 depicts the relationship between the *synbar* threshold and known enzymatic and nonenzymatic rates of RNA hydrolysis. The barnase mutant E73A is nearly 3 orders of magnitude less active than the wild-type enzyme because it lacks the side chain that normally acts as the catalytic general base in the first step of the hydrolysis reaction (30). Since this mutant tests active in the *synbar*

system (12), the activity threshold for this system must lie below the activity of the mutant, as indicated in Figure 5. As a lower bound to the selection threshold, we will consider a particular class of peptide catalysts.

Yanagawa and co-workers have demonstrated that some peptide fragments of barnase catalyze RNA hydrolysis (15). Although they drew the conclusion that this activity is relevant to the function of the whole enzyme, their demonstration that completely unrelated peptides show the same activity undermines that conclusion. Their further demonstration that for peptides to exhibit this activity they need only carry a net charge of +2 or more argues convincingly that the activity they have observed is not enzymatic in nature. However, with activities approximately 5 orders of magnitude below that of wild-type barnase, these are remarkably active catalysts for nonenzymes. Their ability to bind RNA (15) accounts, at least in part, for their catalytic performance.

We have synthesized one of these peptide catalysts (see Materials and Methods) and determined its activity to be 0.002% of that of wild-type barnase (mole-to-mole basis). Taking this to be an approximate upper-limit rate of nonenzymatic RNA hydrolysis under physiological conditions, we estimate basal enzymatic activity to be approximately 4 orders of magnitude below the activity of wild-type barnase. This defines the basal enzymatic activity unit used in the scale of Figure 5. Under the reasonable assumption that barnase mutants must outperform nonenzyme catalysts in order to exhibit lethality (evidence that lethality requires essentially barnase-like structure, and hence proper enzyme function, is discussed below), we conclude that the selection threshold for the *synbar* system must lie above 0.2 bu . Situated between 2.0×10^{-1} and $2.0 \times 10^1 \text{ bu}$, then, the threshold can reasonably be said to be in the vicinity of the basal level, $1 \times 10^0 \text{ bu}$. Consequently, substitutions that lead to activities below the threshold (i.e., to nonlethal barnase variants) must reduce activity to, or nearly to, nonenzymatic levels.

Inactivating Mutations in Perspective

Inspection of the collection of mutants having this dramatic effect suggests that they fall into three classes (Table 1). The first of these, class I, includes all substitutions that replace a side chain known to be directly involved in substrate binding or catalysis. The 17 substitutions falling into this class all involve replacement of Arg83, Arg87, or His102. As with all substitutions, there is a possibility of the local change at the site of substitution leading to a more extensive structural disturbance. However, because of the crucial and direct role of these three residues in function, such propagated disturbances would not need to be present to account for the effects of substitution. Therefore, class I substitutions will be excluded from subsequent classes, even if they might otherwise meet the criteria for inclusion.

Class II includes all substitutions (not in class I) that replace a side chain that is substantially buried (i.e., <10% solvent-exposed) in the wild-type structure. Although the number of mutants falling into this class is similar to the number in the previous class, the number of positions involved is significantly greater. Consequently, positions that contribute to this class tend to be considerably less vulnerable

Table 1: Classification of Inactivating Substitutions

WT residue	solvent exposure of the WT residue (%) ^a	class I	class II	class III
Tyr24	9 (1)	—	D	—
Leu42	0 (0)	—	R	—
Ala46	2 (0)	—	P	P
Gly52	4	—	V	V
Gly53	27	—	—	V
Trp71	2 (2)	—	C, S	—
Arg72	23 (28)	—	—	P
Ala74	0 (0)	—	P	P
Asp75	0 (0)	—	A, H, Y, V	—
Arg83		C, G, H, L, P, S	—	—
Arg87		C, G, H, L, P, S	—	—
Leu89	0 (0)	—	R	—
Tyr90	7 (7)	—	D	—
Ser91	1 (0)	—	P	P
His102		D, N, Q, R, Y	—	—

^a Solvent-accessible surface areas, calculated by the method of Lee and Richards (31), are given as percentages of the areas of each amino acid X in an extended Gly-X-Gly tripeptide (32). The first value for each position applies to the entire residue; values in parentheses apply to side chains alone (omitted for glycines). Exposure values are not given for residues that contact substrate because inactivating substitutions at these positions are exclusively assigned to class I.

to inactivating substitution (i.e., only a small fraction of the accessible substitutions destroy function) than the positions involved in class I substitutions. Position 75, where 4 of 7 possible substitutions are inactivating, provides a possible exception to this. In the wild-type enzyme, the aspartate side chain at this position forms a salt bridge with the arginine side chain at position 83, one of the 3 positions that account for all class I substitutions. The extreme sensitivity to modification at position 83 suggests that the sensitivity exhibited at position 75 might be due to the close interaction between these two residues.

Many of the substitutions in class II involve replacement of a hydrophobic side chain with a polar or charged one. The exceptions to this, however, are sufficiently numerous that they suggest another cause of inactivation. Thus, class III includes all substitutions (not in class I) that either introduce a proline residue or replace a glycine residue. Because the proline side chain places unusual restrictions on backbone conformation and the glycine side chain does just the opposite, both types of substitution have a strong tendency to introduce local backbone distortion. As shown in Table 1, 6 substitutions can be placed into class III, several of them also falling into class II. Since there are 66 positions where Pro is not an accessible substitution (see Figure 2), the true size of class III is probably somewhat larger.

It is noteworthy that these three classes give a complete description of the kinds of single substitutions that can destroy this enzyme. Considering the level of functional impairment required by *synbar* selection, we would expect the disruptive effects of these sorts of substitutions to be evident from previous work on other proteins. This is indeed the case. Earlier exhaustive-substitution studies, for example, have demonstrated the functional importance of particular active-site residues (4–6), the higher functional sensitivity at buried positions (1, 4, 5, 7), the unusually disruptive nature of proline (1), and the unusual sensitivity to substitution of particular glycine residues (3, 4). What the *synbar* system reveals is the extent to which the corresponding structural features (active-site groups, buried side chains, and local

backbone conformation) are essential to enzyme function at the most basic level.

Four primary conclusions are evident in this regard. First, it is a rare single substitution that is capable of destroying barnase function (i.e., reducing it to a level that can be approached by nonenzymes). Only about 5% of the accessible substitutions in this study were found to have an effect so severe. Second, in all cases where a substitution does have this effect, the cause (in broad terms) appears to be either (a) direct modification of the active site, (b) nonconservative replacement of a buried side chain, or (c) introduction of non-native local backbone constraints. Third, even among substitutions falling into these categories, elimination of enzyme function is atypical. For example, of the eight positions where side chains interact directly with substrate (indicated in Figure 2), only three are vulnerable to inactivating substitution, and only two of those appear to be wholly intolerant of substitution. Finally, all of the wild-type residues that exhibit extreme sensitivity to substitution (i.e., where a substantial majority of the accessible substitutions eliminate enzyme function) interact directly with substrate.

Having surveyed the set of inactivating substitutions, we can examine an important point that was presumed to be true in the previous section, namely, that barnase mutants must function as enzymes to exhibit lethality in the *synbar* selection system. The mere fact that some single substitutions can render barnase nonlethal in this system suggests that proper enzymatic activity [as opposed to the hydrolytic activity exhibited by peptides such as OVA1 (Figure 5)] is required for lethality. The fact that all instances of inactivation can be explained with reference to particular aspects of the structure and mechanism of wild-type barnase strengthens this conclusion because it indicates that these aspects are necessary for lethality. The conclusion becomes even more compelling when we consider what the inactivating substitutions tell us about the role of larger structural elements in forming a lethal protein.

Of the 15 positions that are sensitive to substitution, 12 are not involved in any direct interaction with substrate. Inactivation by substitution at these 12 positions must therefore result from propagated structural disturbances. As shown in Figure 6, these points of structural sensitivity are distributed among four elements of secondary structure (the third of three helices and the first three of five β -strands). Since the enzyme can be rendered nonlethal by propagated structural changes arising from substitutions in these structural elements, we can safely conclude that these elements must be at least partly formed for lethality to be possible. By preparing and testing truncated mutants lacking either the major α -helix (helix 1) or the final β -strand, we have further determined that these two elements must be present for lethality to be possible (D. D. Axe, unpublished result). This implies that all five strands of the sheet must be in place (the cooperative nature of β -sheets makes it implausible that a single internal strand could be unformed). Thus, of the eight elements of secondary structure in barnase, seven must be at least partly formed for a mutant to be lethal. Taken together with the three sensitive active-site positions, this means that the molecule must be largely intact for the host cell to be killed by it, and this confirms our earlier presumption that mutants scored as active are true enzymes.

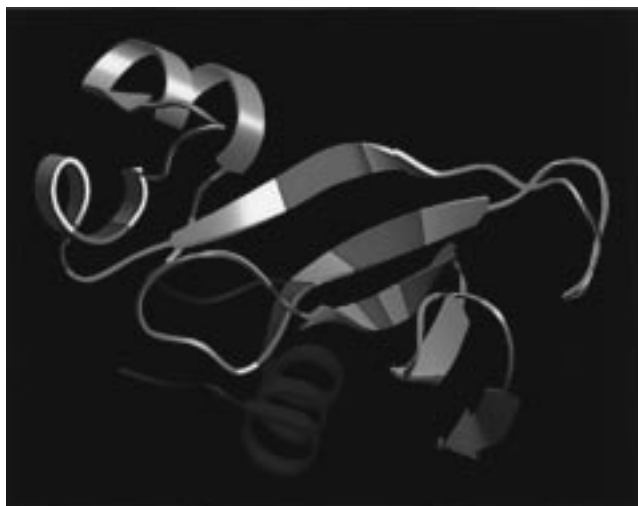


FIGURE 6: Location in the barnase structure of the 15 positions found to be vulnerable to inactivating substitution. Orange indicates the 3 positions where the side chain interacts directly with substrate in the wild-type enzyme (16). The other 12 sensitive positions are colored magenta. In addition, activity has been shown to be eliminated by both N-terminal and C-terminal truncations. The missing portions in these inactive constructs are shown in blue and green, respectively.

Issues Calling for Further Study

A number of interesting questions are raised by the results of this work. Class I is probably of more interest for what it does not contain than for what it does. That either of the active-site residues E73 or H102 [normally filling the roles of catalytic general acid and base in the two-step reaction (30)] can be replaced without destroying the enzyme raises the interesting question of how the enzyme compensates for their absence. Of particular interest among class II substitutions are those where the introduced side chain is not highly hydrophilic and the substitution does not fall into class III. W71C is the best example of such a substitution.

This tryptophan side chain normally packs against a cluster of other hydrophobic side chains to form a small hydrophobic core (33). Evidently, the structural shifts that result when a much smaller side chain is substituted can dramatically impair function. Although arginine is far more hydrophilic than cysteine, our inability to recover a W71R mutant suggests that the hydrophobic portion of the arginine side chain is a better tryptophan substitute in this case than the cysteine side chain is. Since W71C is inactive, one would have thought that W71G would also be inactive. As discussed above, we cannot conclusively declare a mutant to be active on the basis of our inability to recover it. It is possible, then, that W71G is inactive despite the fact that it was not recovered (Figure 2). The best way to conclusively verify that a particular mutant is active is to prepare the appropriate mutant plasmid and apply the *synbar* test as a screen (as in ref 12).

Among the interesting class III substitutions are the ones that replace either of the two glycines at positions 52 and 53. At both positions, valine is seen to be inactivating, but a number of nonconservative substitutions apparently do not completely eliminate activity. The fact that numerous independent examples of the valine substitutions were recovered (seven examples of G52V and four examples of G53V) suggests that most of the mutants that were not

recovered really are active. The presence of a β -bulge (a β -sheet distortion caused by a surplus residue in one strand) involving residues 53 and 54 (34) raises the interesting possibility that this small structural element may be important for barnase function. The results of a detailed investigation of the effects of substitutions in this region will be reported elsewhere.

Design Implications

In light of the above discussion on the significance of essential features, we should now consider what the results of this study imply for the design of a barnase-like enzyme. For the reasons indicated previously, we must here focus on those features that appear to be essential, recognizing that the list of these will probably be incomplete. The first implication to consider is that any protein design aspiring to emulate the fold and mechanism of barnase will need two arginines to fill the roles of R83 and R87, and probably a histidine to fill the role of H102 as well (H102 does not appear to be truly essential, but it is sufficiently sensitive to substitution to suggest that it may be essential in anything but a highly optimal context). As discussed above, though, essential features of wild-type barnase cannot generally be construed as design rules for barnase-like enzymes. Since the experiment described here looks at single mutants only, it does not rule out the possibility that one or more of these active-site residues might be replaceable in the context of appropriate compensating substitutions. If we view this work more broadly, though, as giving us a picture of what kinds of single-residue features are indispensable in the context of a natural enzyme, we conclude that a barnase-like enzyme with barnase-like activity will have a few indispensable active-site residues, their exact identities possibly varying for various designs. A less optimal design, striving only for basal enzymatic activity, would at the very least need to have these few residues in their proper spatial orientations.

Conceivably, this is essentially all that is needed for an enzyme to have basal activity, the trick being to design a scaffold that holds the few key residues in their proper orientations. That is, it is reasonable to view direct interaction with substrate as a prerequisite for a residue to be considered to have a direct role in function. The role of the remaining residues, the scaffold residues, is then to impart the necessary orientations, structural dynamics, and chemical properties to the residues on the "front line". Again, though, experiments with single substitutions cannot be expected to give a full picture of the complexity of this front line. It may well be that some of the barnase residues known to interact directly with substrate but found here not to be vulnerable to inactivating substitutions (e.g., E73) would be essential in a less optimal context. What we conclude from this study, then, is that none of the scaffold residues are irreplaceable in the context of an otherwise wild-type sequence (even D75, the scaffold residue found to be most sensitive to substitution, can be replaced without eliminating activity). Studies involving combined substitutions (as in ref 12) will provide a clearer picture of the sequence requirements for basal barnase function. This work constitutes an essential first step, as it provides information that is needed for the design of multiple-substitution experiments.

ACKNOWLEDGMENT

We thank Dr. Wai Chen for his assistance in the preparation of Figure 6 and Dr. Brian DeDecker for his assistance in the calculation of solvent-accessible surface areas.

REFERENCES

1. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994) *J. Mol. Biol.* **240**, 421–433.
2. Bowie, J. U., and Sauer, R. T. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2152–2156.
3. Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., and Hutchison, C. A. (1989) *Nature* **340**, 397–400.
4. Rennell, D., Bouvier, S. E., Hardy, L. W., and Poteete, A. R. (1991) *J. Mol. Biol.* **222**, 67–87.
5. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S., and Palzkill, T. (1996) *J. Mol. Biol.* **258**, 688–703.
6. Wen, J., Chen, X., and Bowie, J. U. (1996) *Nat. Struct. Biol.* **3**, 141–148.
7. Terwilliger, T. C., Zabin, H. B., Horvath, M. B., Sandberg, W. S., and Schlunk, P. M. (1994) *J. Mol. Biol.* **236**, 556–571.
8. Huang, X., and Boxer, S. G. (1994) *Nat. Struct. Biol.* **1**, 226–229.
9. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) *Science* **247**, 1306–1310.
10. Suckow, J., Markiewicz, P., Kleina, L. G., Miller, J., Kisters-Woike, B., and Muller-Hill, B. (1996) *J. Mol. Biol.* **261**, 509–523.
11. Hartley, R. W. (1988) *J. Mol. Biol.* **202**, 913–915.
12. Axe, D. D., Foster, N. W., and Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5590–5594.
13. Hirel, P. H., Schmitter, J. M., Dessen, P., Fayat, G., and Blanquet, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8247–8251.
14. Modrich, P. (1991) *Annu. Rev. Genet.* **25**, 229–253.
15. Yanagawa, H., Yoshida, K., Torigoe, C., Park, J. S., Sato, K., Shirai, T., and Go, M. (1993) *J. Biol. Chem.* **268**, 5861–5865.
16. Buckle, A. M., and Fersht, A. R. (1994) *Biochemistry* **33**, 1644–1653.
17. Knowles, J. R. (1991) *Nature* **350**, 121–124.
18. Mesecar, A. D., Stoddard, B. L., and Koshland, D. E. (1997) *Science* **277**, 202–206.
19. Radzicka, A., and Wolfenden, R. (1995) *Science* **267**, 90–93.
20. Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman, New York.
21. Hollfelder, F., Kirby, A. J., and Tawfik, D. S. (1996) *Nature* **383**, 60–63.
22. Blacklow, S. C., Raines, R. T., Lim, W. A., Zamore, P. D., and Knowles, J. R. (1988) *Biochemistry* **27**, 1158–1167.
23. Ellerby, L. M., Cabelli, D. E., Graden, J. A., and Valentine, J. S. (1996) *J. Am. Chem. Soc.* **118**, 6556–6561.
24. Morris, K. N., and Wool, I. G. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4869–4873.
25. Stemmer, W. P. C. (1994) *Nature* **370**, 389–391.
26. Wright, M. C., and Joyce, G. F. (1997) *Science* **276**, 614–617.
27. Tarasow, T. M., Tarasow, S. L., and Eaton, B. E. (1997) *Nature* **389**, 54–57.
28. Zhang, B., and Chech, T. R. (1997) *Nature* **390**, 96–100.
29. Mossakowska, D. E., Nyberg, K., and Fersht, A. R. (1989) *Biochemistry* **28**, 3843–3850.
30. Day, A. G., Parsonage, D., Ebel, S., Brown, T., and Fersht, A. R. (1992) *Biochemistry* **31**, 6390–6395.
31. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400.
32. Miller, S., Janin, J., Lesk, A. M., and Chothia, C. (1987) *J. Mol. Biol.* **196**, 641–656.
33. Serrano, L., Kellis, J. T., Cann, P., Matouschek, A., and Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 783–804.
34. Chan, A. W. E., Hutchinson, E. G., Harris, D., and Thornton, J. M. (1993) *Protein Sci.* **2**, 1574–1590.

BI9804028