Application of the “Codon-shuffling” Method
SYNTHESIS AND SELECTION OF DE NOVO PROTEINS AS ANTIBACTERIALS*[S]

Received for publication, March 21, 2005, and in revised form, April 15, 2005
Published, JBC Papers in Press, April 20, 2005, DOI 10.1074/jbc.M503056200

Alka Rao‡§, Sidharth Chopra‡§, Geeta Ram†, Ankit Gupta‡, and Anand Ranganathan‡§
From the ‡Recombinant Gene Products Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110067, India

Library-based methods of non-rational and part-rational designed de novo peptides are worthy beacons in the search for bioactive peptides and proteins of medicinal importance. In this report, we have used a recently developed directed evolution method called “codon shuffling” for the synthesis and selection of bioactive proteins. The selection of such proteins was based on the creation of an inducible library of “codon-shuffled” genes that are constructed from the ligation-based assembly of judiciously designed hexamer DNA duplexes called dicodons. Upon induction with isopropyl 1-thio-β-D-galactopyranoside, some library members were found to express dicodon-incorporated proteins. Because of this, the host cells, in our case Escherichia coli, were unable to grow any further. The bacteriostatic/lytic nature of the dicodon proteins was monitored by growth curves as well as by zone clearance studies. Transmission electron microscopy of the affected cells illustrated the extent of cell damage. The proteins themselves were overexpressed as fusion partners and subsequently purified to homogeneity. One such purified protein was found to strongly bind heparin, an indication that the interaction of the de novo proteins may be with the nucleic acids of the host cell, much like many of the naturally occurring antibacterial peptides, e.g. Buforin. Therefore, our approach may help in generating a multitude of finely tuned antibacterial proteins that can potentially be regarded as lead compounds once the method is extended to pathogenic hosts, such as Mycobacteria, for example.

The objective of recasting or constructing proteins from scratch to achieve altogether new entities that are then able to carry out a preordained catalytic or structural function is currently an area of intense scientific research. That small proteins or peptides can be constructed based on the first principles of secondary structure design has been known for some time, largely through some seminal contributions of DeGrado and co-workers (1) and others (2, 3). This so-called de novo protein design is the rational approach toward the creation of functional proteins, and it is perhaps only now, with the realization of the hitherto unreachable requirements of computational power, that one is witness to an increasing number of functional peptides and proteins being successfully generated de novo (4, 5). On the other hand, non-rational and part-rational approaches toward creating de novo proteins designed for specific purposes seem to be much more powerful methods at generating functional proteins, especially for in vivo biological requirements (6, 7). Methods like the “phage display” (8) and “mRNA display” (9) increasingly allow one to effectively generate and then subsequently pan non-rational peptide and protein libraries in order to select for the right protein. In situations where an effective assay-based selection system is available, part-rational and non-rational approaches may be more effective than a rational approach in the isolation of a functional protein, largely because the part-rational and non-rational approaches mimic evolutionary pathways of selection (of the fittest). In this context, two such studies are worthy of special attention.

A little more than a decade ago, Hecht and co-workers (10) put forth a most novel hypothesis, that of binary patterning of proteins. According to this simple concept, it is the binary patterning of polar (○) and non-polar (●) amino acids in the designed protein and not the exact identity of those amino acids that eventually dictates the secondary structure of the designed protein. The binary-patterning concept may be viewed as a part-rational approach, whereby one can predetermine the library folds but not the exact folds of its individual members. Indeed, de novo protein libraries based on binary patterning have furnished many functional as well as structural proteins (11, 12), although problems of a suitable assay system, as well as the limitation of a fixed length of library members, have yet to be addressed satisfactorily. The second seminal study by Mekalanos and co-workers (13) involves the generation of non-rational peptide libraries based on degenerate oligonucleotide design. Their work elegantly demonstrated that an inducible non-rational DNA library, once translated in vivo, could lead to the selection of a functional peptide. This so-called ABBIS approach was successfully used for isolation of a number of peptides that, upon overexpression in the host cell, led to severe growth attenuation of the host. The ABBIS method too has its limitations that are primarily because of its use of fixed-length degenerate oligonucleotides as a starting point for making a peptide library. Nonetheless, it remains a powerful example of a non-rational approach toward the de novo synthesis of useful proteins, in their case antimicrobial peptides (AMP).1

---

* This work was supported by internal grants of the International Centre for Genetic Engineering and Biotechnology and funding from the Council for Scientific and Industrial Research (to A. Rao and G. R.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains Supplemental data and Figs. S1–S8.

‡ Both authors contributed equally to this work.
§ Present address: School of Medicine, Stanford University, Palo Alto, CA 94305.
¶ To whom correspondence should be addressed. Tel.: 91-11-26195007; Fax: 91-11-26162316; E-mail: anand@igcgeb.res.in.

1 The abbreviations used are: AMP, antimicrobial peptide; trx, thioredoxin; DC, dicodon; HPDN, hairpin-encapsulated DC fragment; IPTG, isopropyl β-D-thiogalactoside; GST, glutathione S-transferase; Ni-NTA, nickel-nitrilotriacetic acid; TEM, transmission electron microscopy; ds, double-stranded.
Synthesis and Selection of Bioactive de Novo Proteins

Of late, there is an urgent need for generating newer more potent AMPs, given their newfound importance in the context of widespread antibiotic resistance among human pathogens (14–16). AMPs have been isolated from a variety of species, including around 40 odd AMPs from humans (15, 17). Thus far, around 700 AMPs have been isolated and characterized (18) and many among them are in advanced clinical trials as drug candidates against bacterial and fungal infections (15, 19). A general characteristic of AMPs is their net positive charge that is believed to be crucial for gaining entry into negatively charged bacterial outer membranes, although lately, some anionic AMPs have also been isolated (20, 21). Consequently, many models for bacterial entry have been postulated (17, 22). However, recent DNA-microarray studies on bacteria affected by AMPs have suggested that cell wall lysis may not be the sole mode of AMP action, given that the expression of a host of cytoplasmic genes is severely affected upon AMP cell entry (23, 24). Notwithstanding the uncertainty surrounding their action, it is clear that AMPs represent a class of molecules that may in the near future provide an effective alternative to the currently used antibiotics (14). Therefore, we believe that an approach that leads to the synthesis of a diverse pool of AMPs and one that addresses the earlier problems associated with AMP synthesis would be of considerable help.

In this report, we describe one such approach that is based on the application of a directed-evolution technique that we recently developed called codon shuffling (25). We also describe how, through codon shuffling, libraries are generated that: (a) can be both non-rational as well as part-rational; (b) have members large enough to be classified as proteins and not just peptides; (c) are not restricted by limits of length of their corresponding genes; and (d) can be preferentially skewed in amino acid attributes like charge and hydrophobicity, thereby narrowing down the search for a successful AMP.

EXPERIMENTAL PROCEDURES

Construction of Expression Plasmid pDNp28—An unrelated DNA fragment of 587 bp in length from the plasmid pET21c was PCR-amplified using pfu polymerase and the oligonucleotides 5′-aagcatg-ttcagttaccttggcgccgctctagaggcccgcgcgggcctctagagcggccg-3′ and 5′-aagcatgctctagaggcccgcgcgggcctctagagcggccg-3′ as forward and reverse primers, respectively. The PCR product was cloned in pBluescriptSRF vector, and the fragment was excised using NcoI and SnaBI restriction enzymes and cloned in pET28a vector previously cut with the same two enzymes. The resulting plasmid was amplified using NcoI and XhoI restriction enzymes and cloned in pBluescriptSRF vector, and the fragment was excised using NcoI and SnaBI restriction enzymes and cloned in pDNp28 vector previously cut with the same two enzymes. The resulting plasmid was designated pTEMDNp28. Plasmid pTEMDNp28 carries the SnaBI site into which DC fragment libraries can be inserted. The site is immediately downstream of the TEM-1 secretion signal sequence and just upstream of the His6 tag.

Western Blot Analysis—Whole-cell lysates of EQAMP1–3 cultures were run on 15% SDS-PAGE and transferred onto supported nitrocellulose membrane (Invironet). The membrane was blocked with 1% polyvinylpyrrolidone and then probed with anti-His monoclonal antibody (1:7500 dilution). The blot was developed by using 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (Promega).

Heparin Binding Studies—Procedures relating to the construction of GST-SKAMP1 gene and subsequent purification of the corresponding protein are described under supplemental data. 20 µg of pure GST-SKAMP1 protein was loaded onto a column containing heparin-Sepharose. The column was washed extensively with buffer C. Elution buffer (C with an NaCl gradient from 250 mM to 1 M) was then poured over the column, and fractions were collected.

Western Blot—The eluted fractions, as well as the heparin resin, were probed with anti-GST antibodies (1:25000 dilution) for 1 h. Nitrocellulose membrane was then washed four times with 1× phosphate-buffered saline plus Tween 20 (each wash of 5-min duration). Secondary probing was done with anti-rabbit IgG heavy and light chain AP-conjugated antibodies (1:5000 dilution) for 1 h. After washing the membrane with 1× phosphate-buffered saline plus Tween 20, the blot was developed with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium substrate.

RESULTS AND DISCUSSION

Exploring the Possibility of Using Codon Shuffling for Constructing de Novo Protein Libraries—The codon-shuffling method was initially conceptualized by us as a directed-evolution technique. However, because the method is based not on the selection of a progeny through inducement of random mutations in the parent gene as in gene shuffling (26) but rather on extending a truncated parent gene with new DNA fragments that are constructed de novo, codon shuffling can also be viewed as a method for creating stand-alone new DNA fragments. This is because codon shuffling fundamentally entails randomly assembling 6-bp DNA duplexes (called DCs) from a set containing 14 such dicodons (Table 1), each of which are judiciously chosen to encode amino acid pairs, such that in an equimolar dicodon pool, all 20 natural amino acids are represented and in proportions that mirror their usage in Escherichia coli. Furthermore, there is no restriction on the lengths of the resulting DNA fragments. Consequently, the amount of degeneracy achievable from ligation of the dicodons in making DNA fragments of ~250–300 bp in length is astronomically high and much beyond what may be required to exhaust all of the possible protein folds. Therefore, at first glance, codon shuffling represents a useful method for generating de novo protein libraries.

Here, it is pertinent to mention one disadvantageous feature of peptide libraries made from degenerate oligonucleotides. In addition to the problems of length restriction (the upper limit being 100 nt on average, resulting in 30–40-amino acid-long peptides) as well as random appearances of stop codons, such peptide libraries yield few folded peptides, thereby severely limiting their usefulness (27, 28). One reason why few peptides from such libraries are able to fold properly may be the completely non-rational nature of their parent DNA fragments that may result in a peptide that contains a string of polar (O...O...) or non-polar (●●●●●...) residues, rather than an arrangement of such residues that promotes a secondary structure feature (●●●●●...●●●●●...●●●●●...●●●●●...) or (●●●●●...●●●●●...●●●●●...●●●●●...●●●●●...). Indeed, by determining the degeneracy of parent oligonucleotides, such that the resulting peptides may possess polar/non-polar arrangements, Hecht and coworkers (10, 29) were able to generate a library of peptides wherein most members displayed a well-folded behavior. Consequently, the library size in such situations may not be nec-
Essentially large to begin with, much like in codon shuffling. One possible explanation in the codon-shuffling context may be that the 14 dicodons themselves possess a range of inherent secondary structure-forming attributes because of the manner in which they are paired, such as polar/non-polar, non-polar/polar, non-polar/turn, non-polar/non-polar, and other such combinations (Table I). For example, Glu-Leu or Asp-Ile repeats yield a binary pattern (●●●●●●●●●●) that represents β-sheet structure, whereas the string Met-His-Asp-Ile-Met-His-Glu-Leu-Ser-Thr yields a binary pattern (●●●●●●●●●●) that would preferentially form an α-helix. Additionally, the appearance of the dicodons Trp-Pro, Pro-Gly, or Gly-Ala may induce turns or breaks, whereas sputtering of Cys-Ala within sequences was established as a dicodon-dicodon pair (DC-DC) location of dicodon pairs within protein structures. The search of all of the proteins in the Protein Data Bank to find the combinations (13, 33, 34). However, whether the peptides are as effective once they are removed from their scaffolds or, more worryingly, whether the de novo peptide activity is partly due to the scaffold itself are some of the issues that remain largely unaddressed. However, one could be prudent and reduce the deleterious effects of the scaffold, simply by keeping its size to a minimum. Using a large protein scaffold may stimulate the de novo protein to take a shape that without the scaffold would be altogether different. Therefore, in designing the ds hairpin, we envisaged that the ds hairpin should code for a string of 6–7 nucleotides that are complementary to the hairpin sequence (30). Moreover, the ds hairpin, in addition to its primary role of amplifying DC fragments, could also be used as a scaffold for encapsulating DC proteins. Protein scaffolds have previously been employed to tighten randomly generated de novo peptides in hopes that the peptides may fold better as their N and C termini are no longer free (31, 32). One such scaffold, the thioredoxin (trx) protein, has been employed on numerous occasions (13, 33, 34). However, whether the peptides are as effective once they are removed from their scaffolds or, more worryingly, whether the de novo peptide activity is partly due to the scaffold itself are some of the issues that remain largely unaddressed. However, one could be prudent and reduce the deleterious effects of the scaffold, simply by keeping its size to a minimum. Using a large protein scaffold may stimulate the de novo protein to take a shape that without the scaffold would be altogether different. Therefore, in designing the ds hairpin, we focused on what was our initial goal: to isolate proteins that act as an antibacterial.

**Use of ds DNA Hairpins for Amplification of Codon-shuffled Fragments**—To address the above-mentioned problem, we envisaged the use of ds DNA hairpin that would first cap the two exposed ends of the DC-assembled DNA fragment and then be used as a template for fragment amplification using oligonucleotides that are complementary to the hairpin sequence (30). Moreover, the ds hairpin, in addition to its primary role of amplifying DC fragments, could also be used as a scaffold for encapsulating DC proteins. Protein scaffolds have previously been employed to tighten randomly generated de novo peptides in hopes that the peptides may fold better as their N and C termini are no longer free (31, 32). One such scaffold, the thioredoxin (trx) protein, has been employed on numerous occasions (13, 33, 34). However, whether the peptides are as effective once they are removed from their scaffolds or, more worryingly, whether the de novo peptide activity is partly due to the scaffold itself are some of the issues that remain largely unaddressed. However, one could be prudent and reduce the deleterious effects of the scaffold, simply by keeping its size to a minimum. Using a large protein scaffold may stimulate the de novo protein to take a shape that without the scaffold would be altogether different. Therefore, in designing the ds hairpin, we focused on what was our initial goal: to isolate proteins that act as an antibacterial.

**Table I**

| DC | BINARY PATTERN | SEARCH STRING (SS) | NUMBER OF SS | TOTAL EMBEDDED | FOUND EMBEDDED IN: | HELIX (%) | SHEET (%) | TURN (%) | RANDOM (%) |
|----|----------------|--------------------|--------------|----------------|----------------------|-----------|-----------|---------|-----------|
| EL | ●●             | XX-EL-DC-XX        | 14           | 1052           | 549 (52)             | 168 (16)  | 239 (23)  | 96 (9)  |
| DI | ●●             | XX-DI-DC-XX        | 14           | 778            | 193 (25)             | 333 (43)  | 141 (18)  | 111 (14)|
| KL | ●●             | XX-KL-DC-XX        | 14           | 971            | 270 (28)             | 260 (27)  | 303 (31)  | 138 (14)|
| NV | ●●             | XX-NV-DC-XX        | 14           | 387            | 92 (24)              | 160 (41)  | 84 (22)   | 51 (13) |
| GA | ●●             | XX-GA-DC-XX        |              | 1262           | 339 (27)             | 205 (16)  | 487 (39)  | 231 (18)|
| ST | ○○             | XX-ST-DC-XX        | 14           | 862            | 161 (18)             | 151 (17)  | 435 (49)  | 135 (16)|
| CA | ●●             | XX-CA-DC-XX        | 14           | 188            | 53 (28)              | 69 (37)   | 60 (32)   | 6 (3)   |
| FG | ●●             | XX-FG-DC-XX        | 14           | 727            | 119 (16)             | 19 (3)    | 533 (73)  | 56 (8)  |
| MH | ●●             | XX-MH-DC-XX        | 14           | 89             | 36 (40)              | 20 (23)   | 26 (29)   | 7 (8)   |
| WP | ●●             | XX-WP-DC-XX        | 14           | 136            | 11 (9)               | 0 (0)     | 99 (86)   | 6 (5)   |
| QL | ●●             | XX-QL-DC-XX        | 14           | 393            | 115 (29)             | 134 (34)  | 97 (25)   | 47 (12) |
| YV | ●●             | XX-YV-DC-XX        | 14           | 362            | 142 (39)             | 136 (38)  | 32 (9)    | 52 (14) |
| FE | ○○             | XX-FE-DC-XX        | 14           | 726            | 522 (72)             | 84 (12)   | 111 (15)  | 9 (1)   |
| RT | ○○             | XX-RT-DC-XX        | 14           | 423            | 73 (17)              | 65 (15)   | 231 (55)  | 54 (13) |

|                  |                  |                  | 196           | 8356           | 2675 (30) | 1804 (23) | 2878 (36) | 999 (11) |

* DCs that have high preferences within each fold element have their percentages depicted in bold.
binary patterns for the scaffold, we settled on the pattern shown in Fig. 1. The binary pattern, ○○○○○○○, along with its "antisense" pattern, ●●●●●●●, are both predicted to form an α-helix based on the Hecht hypothesis (10). We then searched the Protein Data Bank to document the presence of these two patterns in the various fold elements. Interestingly, both patterns show a generous preference for α-helices, thereby confirming the practicability of using Hecht binary patterning for de novo peptide design (Fig. 1). As a next step, we embedded some unique restriction enzyme (RE) sites (none of which appears in any given DC fragment) within the ds hairpin sequence, such that they would not disturb the binary pattern of the translated hairpin (Fig. 1B). The final sequence of the designed hairpin (shown in Fig. 1) coded for the following two sequences, "SGRHVFK" and "FKHVAAA," depending upon which hairpin strand was translated. Importantly, both these sequences conform to an α-helix-forming Hecht pattern.

Construction of de Novo DC Protein Libraries with the Hairpin Scaffold—Our next objective was to experimentally investigate the merit of using ds hairpin for amplifying DC fragments.
ments as well as for ably providing a scaffold for the DC proteins. As a starting point, we added a molar excess of ds hairpin to a mixture of equimolar DC pool (see "Experimental Procedures" for details). This was followed by amplification of encapsulated DC fragments with a 5'-phosphorylated oligonucleotide that was complimentary to the ds hairpin and therefore could act as both forward and reverse primers for the purposes of amplification. The PCR products were separated based on their length (50–400 bp) and ligated directly with SnaBI-cut expression vector pDNp28. Upon transformation of E. coli BL21(DE3) followed by selection on kanamycin, we obtained colonies roughly to the order of 10^4. We selected and sequenced some representative clones and found all of them to indeed contain DC sequences that were flanked on either end by the ds hairpin. The corresponding DC protein library showed little sequence similarity among its members and was vastly variant in terms of length and primary sequence attributes (Fig. 2A). To test the protein expression levels of our DC library, we induced growing cultures of a representative clone HPDN5 with IPTG. We observed overexpression of HPDN5 protein and were subsequently able to purify the protein using Ni-NTA chromatography (Fig. 2B). These results firmly indicated to us that the generation of DC proteins from ds hairpin-encapsulated DC fragment library represents a viable route toward isolation of de novo proteins that are expressed from their corresponding genes at good levels and therefore can be easily purified.

Finally, we wanted to explore a case where the DC protein is encapsulated by the traditional, often employed trx scaffold. Therefore, we ligated the hairpin-encapsulated DC fragment library with an expression vector (pTRXDNp21) that carried the modified trx gene wherein a unique SnaBI site had been positioned within the trx sequence that corresponded to the region between the two cystines (see supplemental data for details). The DNA of many of the obtained clones was sequenced and found to contain hairpin-encapsulated DC fragment sequences expectedly positioned within the SnaBI site (Supplemental Fig. S2A). One representative member of the trx-DC library was further analyzed to check for protein expression. Upon induction with IPTG, the protein trxHPDN3 was found to be generously overexpressed. However, most of the protein was found to be in the insoluble fractions, albeit the protein could be purified to homogeneity under denaturing conditions (8 M urea) using Ni-NTA chromatography (Supplemental Fig. S2B). Upon dialysis, we found the protein to be soluble in a minimum of 150 mM urea. Circular dichroism studies of THPDN3 along with native trx protein showed that the chimera displayed a predominant a-helical content (helix, 24.1%; sheet, 10.0%; turn, 34.4%; random, 34.4%), much like the parent trx protein (helix, 21.5%; sheet, 10.0%; turn, 31.3%; random, 32.8%; Supplemental Fig. S2C). Without the use more incisive techniques such as x-ray crystallography or NMR, it is not possible to speculate whether the DC protein has folded inside the trx scaffold or whether the obtained CD data are for a single protein entity. Further work in this direction is currently underway in our laboratory.

In summary, we have investigated the usefulness of DC proteins as candidates for de novo protein libraries. We have also found that it is worthwhile to encapsulate the DC proteins using scaffolds. Of the two scaffolds that we used for this purpose, we found the trx scaffold to yield chimeras that were insoluble and thus of little use in vivo.

**Design and Synthesis of de Novo Anti-bacterial Proteins**—Having put in place a method for the design, synthesis, and selection of de novo proteins with two different scaffolds, we now focused our efforts on the synthesis of de novo proteins that would act as an antibacterial. Our strategy was to select for AMPs in vivo by identifying those transformed bacterial colonies that were unable to grow in the presence of IPTG in contrast to robust growth in the absence of the inducer. Additionally, we decided to tether the DC proteins with an N-terminal signal sequence for the following reason. Because the

---

**A.**

| NAME | SEQUENCE | aa | pI | MW |
|------|----------|----|----|----|
| HPDN1 | MG**8**GRHVFKEGAKHLKDLIPGMDIKLMH | 136 | 5.7 | 15.7 |
| HPDN2 | MG**8**GRHVFKELESTRMHTRMIGAHMHI | 110 | 6.6 | 12.5 |
| HPDN3 | MG**8**GRHVFKEQJMKHELAMGTKMTRVNDI | 102 | 9.8 | 11.6 |
| HPDN4 | MG**8**GRHVFKEGSTIDIMNYVSTGAEALM | 114 | 6.9 | 13.0 |
| HPDN5 | MG**8**GRHVFKEQLQDMFPMGMDHIEEDINV | 162 | 6.5 | 18.7 |
| HPDN6 | MG**8**GRHVFKEVQCDIPMPSDIEDELMH | 98 | 6.2 | 11.3 |
| HPDN7 | MG**8**GRHVFKEKMLKLHLQIDLNIQMLGGA | 114 | 7.3 | 13.2 |
| HPDN8 | MG**8**GRHVFKEGSTDMPDMLQDIQRTDI | 88 | 6.2 | 7.8 |

---

**B.**

![Fig. 2. A, amino acid (aa) sequences of some ds hairpin-encapsulated DC protein library members showing variance in protein length, pl, and molecular mass (MW) (given in kDa). The hairpin sequences are displayed in boldface letters. Segments derived from the expression vector are underlined. B, 15% denaturing SDS-PAGE showing expression and purification of one library member, HPDN5. Lane 1, Amersham Biosciences low molecular mass markers (shown in kDa); lane 2, total protein obtained from an IPTG-induced culture of E. coli BL21(DE3)/HPDN5; lane 3, soluble protein loaded onto Ni-NTA column; lanes 4 and 5, eluted samples of HPDN5 (molecular mass of 19 kDa); lanes 6 and 7, concentrated HPDN5 protein.](image-url)
AMP site of action could be the cytoplasm, the periplasm, or indeed the extracellular environment of the host, in our case E. coli, we felt that it was important not to exclude, by a method of selection, DC proteins that were acting at any of the above-mentioned three sites. For example, by not choosing to tether the DC library with an N-terminal signal, all of the DC proteins would be restricted to the cytoplasm of the cell. On the other hand, choosing the ompA signal sequence as the primary tether for the AMPs, all DC proteins would be directed toward the extracellular environment of the host cell. As a result, AMPs that would have had a cytoplasmic or a periplasmic site of action would not be selected.

Keeping this viewpoint in mind, we decided to employ the N-terminal signal sequence of transmission electron microscopy-1 (TEM-1) β-lactamase as the primary AMP tether. This was because we and others (25, 36) have shown that, under induction conditions, the TEM signal sequence directs transport of the β-lactamase protein to all three regions of the bacterial cell (the cytoplasm, the periplasm, and the extracellular environment) in the approximate ratio of 70:25:5. Therefore, by choosing the TEM-1 signal, we would not be restricting the site of action of de novo AMPs to a particular cellular environment. The expression vector that would accept the in-frame translation products as well as physical properties of the three unique sequences are shown in Fig. 3. There is a wide variance seen in the protein length and pI among the three proteins named EQAMP1, EQAMP2, and EQAMP3. As a control, the DNA from a colony that grew well on both plates, I and II, was also sequenced (cEQ1). Consensus secondary structure predictions on all four proteins displayed a dramatic difference between the three proteins and the control protein (Fig. 3). Whereas EQAMP1–3 were predicted to be rich in α-helical content, the control protein, cEQ1, was predicted to be rich in random coils. This incongruity among the three EQAMP proteins and the control made us revisit the initial HPDN DC protein library (Fig. 2). Indeed, all HPDN1–8 proteins were predicted to have a random coil content >55%. Therefore, it is tempting to suggest that the EQAMP proteins, because of their in vivo AMP behavior, display pronounced helical fold characteristics that stand in contrast to the predicted fold characteristics of non-selected DC proteins.

Confirmation of AMP Behavior of EQAMP Proteins—To confirm beyond reasonable doubt that the in vivo AMP activity witnessed was due to DC proteins and nothing else, we carried out the following experimental checks. (a) Plasmid DNA from the strains exhibiting growth inhibition upon addition of IPTG were isolated, purified, and used to transform a fresh lot of E. coli competent cells. The addition of IPTG to growing cultures of the freshly transformed cells also displayed growth inhibition, thereby pointing to the fact that the growth inhibition was a property of the plasmid DNA used for transformation. (b) The DC fragment within each plasmid was amplified using universal primers that bound to only the vector regions flanking the DC fragments and nowhere else. The amplified products were digested with restriction enzymes that cut within the vector sequences and ligated to freshly cut pET28a vector. The resulting plasmids (that should be identical to the starting plasmids) were used to investigate growth inhibition and were all found to exhibit the same. This ruled out growth inhibition properties by any means (such as contamination and other factors) other than the DC fragments themselves. (c) The supernatant of IPTG-induced growth cultures of EQAMPs was isolated and spread on agarose plates containing E. coli DH5α lawns, and the plates were incubated for a period of 14–18 h at
37 °C. The absence of any plaque formation ruled out phage-mediated inhibition as the reason for growth inhibition of EQAMP cultures. (d) As a final check, we wanted to confirm that the growth inhibition was due to DC proteins and not because of their DC-mRNA. For this purpose, the unique NcoI site in each EQAMP plasmid (that also contained the ATG start codon) was cut and filled in using Klenow polymerase and the plasmids were self-ligated. This resulted in a frameshift starting from the NcoI ATG codon, which no longer yielded the DC proteins as the translated products. However, the mRNA in each case remained identical to the wild-type EQAMP mRNA. When the resulting plasmids EQAMPlnco1–3 were used, no growth inhibition was seen (Supplemental Fig. S4), thereby confirming that it was the DC protein and not the DC-mRNA that caused inhibition.

Monitoring of Growth Inhibition by EQAMPs—The extent of inhibition by EQAMP1–3 was directly followed by “growth-inhibition” curves. Again, the experiments were carried out in combination with extensive controls that were as described: pET28a vector; the scaffold vector pDNp28; E. coli BL21(DE3) strain; and the scaffold vector containing a dummy gene (in our case trx), the translated product of which does not cause growth inhibition. The control curves as well as those for EQAMP1–3 are shown in Supplemental Fig. S5 and Fig. 4, respectively.

![Fig. 4](image-url)

A–E, growth inhibition curves of isolated AMPs. Generally, growing E. coli cultures were induced with IPTG after 2–3 h of growth or after an A$_{600}$ of 0.4 had been reached. Curves for the induced cultures are shown in black. A, EQAMP1; B, EQAMP2; C, EQAMP3; D, SKAMP1; E, cEQ1.

F, plating of the culture containing the scaffold plasmid pDNp28 as well as plasmids containing the AMPs. The plate contained kanamycin and IPTG.
Also shown in Fig. 4 is the growth curve for the positive control cEQ1. As can be seen, there is a near complete growth inhibition upon IPTG addition in EQAMP samples, whereas there is little or no difference between the induced and uninduced curves for the controls, including cEQ1. Interestingly, the EQAMP samples seemed to display a bacteriostatic rather than a bacteriocytic behavior, because small aliquots of induced cultures taken at the endpoints when used as seed samples to inoculate IPTG-less media resulted in robust growth (results not shown). Identical growth inhibition results were also obtained when we employed “zone-clearance” studies. Small filter paper discs that were resting on growing lawns of EQAMP1–2 cultures were gently soaked with IPTG, and the plates were incubated for 14–18 h. Clearance zones appeared in EQAMP1–3 samples that were not present in the control plates (Supplemental Fig. S6), thereby confirming that the inhibition is inducer-driven and is the result of DC protein expression.

Identification of DC Proteins Responsible for Growth Inhibition— It is well recorded that gratuitous overexpression of a protein can also lead to severe growth inhibition or bacteriostasis of the host (37). To rule out the possibility that the growth inhibition we were witnessing was because the EQAMP proteins were being hugely overexpressed upon IPTG induction, we decided to look for the proteins in lysed cultures of EQAMPs as well as those of cEQ1. The collection of reasonable amounts of biomass for the above-mentioned purpose proved to be a tough challenge because of the bacteriocstatic nature of the expressed DC proteins. Nonetheless, pellets of 1-liter cultures could routinely be lysed and an adequate amount of total protein was extracted for SDS-PAGE analysis. We could not detect EQAMP1–3 proteins at their expected sizes through Coomassie Blue staining. On the other hand, the control cEQ1 protein was readily visible under the same conditions (with and without the TEM-1 signal; Fig. 5A). However, because all of the DC proteins are expressed as C-terminal His tag fusions, proteins EQAMP1–3 could be detected in Western blots using monoclonal anti-His antibodies (Fig. 5B). The presence of minute amounts of EQAMP proteins at their expected sizes and large amounts of the control cEQ1 protein provided strong evidence that the growth inhibition of the host was not because of gratuitous overexpression of DC proteins, but rather a consequence of the antibacterial nature of the latter. All of the efforts to isolate purified EQAMP proteins have as yet proven to be unsuccessful. In vitro transcription/translation of EQAMP genes may however yield sufficient amounts of the proteins for us to carry out further investigations on their antibacterial nature as well as their structure, and such efforts are currently underway in our laboratory.

As stated earlier, two major factors, namely protein charge and its tertiary structure, are responsible for AMP activity of peptide and proteins. Total charge profiles of EQAMP proteins provided an inconclusive picture. Whereas EQAMP1 was predicted to be positively charged at physiological pH, EQAMP2 and EQAMP3 possessed a pI of 6.6 and 6.5, respectively. However, an inspection of charge distribution of the latter two proteins provided for a possibility that these proteins may possess a domain structure wherein one domain could be positively charged (results not shown). Also, the charge state (whether negative or positive) at physiological pH of multiple histidine residues that are present in all EQAMPs could affect the overall charge of the proteins. In the matter of tertiary structure, consensus secondary structure predictions of the three EQAMP proteins label them as rich in α-helical elements (Fig. 3). Indeed, if one were to assume this prediction as correct, the helical-wheel projection drawing of EQAMP2, for example, nicely illustrates the demarcation of its residues, non-polar in the interior and polar at the exterior of the assumed all-helical bundle (Supplemental Fig. S7). However, presently, these are at best conjectures. Isolation of purified EQAMP proteins and their subsequent structural characterization would lead us to an accurate representation. As a next best alternative, we isolated representative AMPs (EQAMP2 and SKAMP1) as trx fusion proteins and obtained their CD spectra (Fig. 6). The AMP fusion proteins displayed a well folded structure as indicated by the CD data for EQAMP2 (helix, 31%; sheet, 0%; turn, 38%; random, 31%) and for SKAMP1 (helix, 29%; sheet, 4%; turn, 34%; random, 33%; Fig. 6B).

Direct Visualization of the in Vivo Effect of EQAMPs—Although growth inhibition experiments had earlier indicated to us that the EQAMP1–3 proteins were most probably bacteriostatic, rather than bacteriocytic in nature, we decided to investigate by TEM the state of the host cell post-IPTG induction with a non-induced cell culture as a reference control. TEM results (Supplemental Fig. S8, A–D) illustrated to us that induced samples displayed a morphology that was markedly different from the non-induced samples. In addition to the disintegration of cell wall, widespread cytoplasmic contraction was also clearly visible. The latter is generally caused when a membrane protein of a cell is affected or sequestered or when the nucleic acids of the cell are targeted (17). We decided to investigate this phenomenon further by studying the possible association of EQAMPs with heparin (heparan sulfate). It is well known that proteins that bind heparin also bind nucleic acids (38, 39). In addition, many AMPs have been shown to
bind heparin directly (40). Although obtaining pure EQAMPs had proved earlier to be unsuccessful, we were able to obtain good quantities of soluble EQAMPs as fusion partners of the GST protein. As a representative example of such fusion proteins, we decided to investigate the SKAMP1-GST fusion protein (see below for the synthesis and isolation of SKAMP1 protein). Purified SKAMP1-GST (molecular mass of 36.7 kDa) was able to bind to heparin at varying pH conditions, and the fusion protein could partly be eluted with buffer containing 1 M NaCl (Fig. 7). Some of the protein was seen bound to heparin resin, even after elution with the above-mentioned buffer, indicating very strong binding. It has previously been shown that GST on its own does not bind to heparin (41), a finding also confirmed by us using purified GST protein as a control (Fig. 7). Therefore, these results indicate association of AMPs with a negatively charged moiety (in this case heparin) that could be either a nucleic acid or indeed a negatively charged membrane protein. Further studies in this direction are ongoing.

Generating AMPs Using Skewed DC Libraries—As mentioned earlier, one exclusive advantage of codon shuffling over other methods of library generation is that skewed libraries can be generated wherein proteins can possess desired physical or chemical attributes on account of the excess presence or complete absence of one or many dicodons. We decided to take advantage of this facet of the codon-shuffling method by constructing an HPDN protein library from a DC mixture wherein the positively charged dicodons Lys-Leu and Arg-Thr were increased three times in proportion to the other 12 dicodons. The method that we have described results in de novo proteins and not just peptides. Because of this, we believe that the library is much less amenable to proteolysis by cellular proteases. The longer average length of the library members also presumably enhances protein folding. Secondly, we have also explored an exclusive property of the codon-shuffling method, which is the allowance for skewing library properties by inclusion, exclusion, or predominance of some DCs. We believe that the ability to skew a protein library further narrows the protein space that needs to be explored. Therefore, one can design skewed libraries to isolate proteins for specific needs. For example, a protein library of arginine-rich sequences may be generated by an overwhelming use of the Arg-Thr dicodon. Polyarginine peptides have lately gained much interest for their ability to act as cell-penetrating peptides that act by binding to the cell surface of a pathogen (42). Furthermore, the recent sequencing of organisms such as Mycobacterium tuberculosis and Plasmodium falciparum has unearthed many virulence-determining proteins that are exclusively composed of short sequence repeats (43). It is not known what selective advantage this may provide the pathogen. Proteins rich in DC-repeats constructed using our method could serve as a model for studying this hitherto unexplored natural phenomenon. Finally, our proof-of-concept experiments relating to the generation of AMPs against E. coli can be conveniently extended to pathogenic microorganisms like M. tuberculosis. The molecular targets of AMPs then can be identified using DNA microarrays. Such studies we believe would provide a novel direction toward discovering new molecules against common pathogens and, importantly, would further extend the scope of de novo protein design as an important chemical tool for tackling generic problems in medicine.

Acknowledgments—We thank Prof. Michael Hecht (Princeton University) for critically reading the manuscript. We also thank Recombinant Gene Products laboratory members for help and support and Malaria group members for their help in generating CD data. Expert analysis of TEM data by Dr. Nurapam Roy Chaudhary is gratefully acknowledged.

REFERENCES

1. Regan, L., and DeGrado, W. F. (1988) Science 241, 976–978
2. Cordes, M. H., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (1999) Science 284, 320–326
3. Hecht, M. H., Richardson, J. S., Richardson, D. C., and Ogden, R. C. (1990) Science 249, 884–889
4. Schneider, G., Schroll, W., Wallukat, G., Muller, J., Nissen, E., Ronspies, W., Wrede, P., and Kunze, R. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 12179–12184
5. Dahiyat, B. I., and Mayo, S. L. (1997) Science 278, 82–87
6. Taylor, S. W., Walter, K. U., Kast, P., and Hilvert, D. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 10596–10601
7. Roberts, R. W. (1999) Curr. Opin. Chem. Biol. 3, 268–273
8. Rodi, D. J., and Makowski, L. (1999) Curr. Opin. Biotechnol. 10, 87–93
9. Keele, A. D., and Szostak, J. W. (2001) Nature 410, 715–718
10. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Science 262, 1680–1685
11. Moflet, D. A., and Hecht, M. H. (2001) Chem. Rev. 101, 3191–3203
12. Roberts, R. W. (1999) Curr. Opin. Chem. Biol. 3, 268–273
13. Rodi, D. J., and Makowski, L. (1999) Curr. Opin. Biotechnol. 10, 87–93
14. Keele, A. D., and Szostak, J. W. (2001) Nature 410, 715–718
15. Katakas, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Science 262, 1680–1685
16. Moflet, D. A., and Hecht, M. H. (2001) Chem. Rev. 101, 3191–3203
17. Roberts, R. W. (1999) Curr. Opin. Chem. Biol. 3, 268–273
18. Rodi, D. J., and Makowski, L. (1999) Curr. Opin. Biotechnol. 10, 87–93
19. Keele, A. D., and Szostak, J. W. (2001) Nature 410, 715–718
20. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Science 262, 1680–1685
21. Moflet, D. A., and Hecht, M. H. (2001) Chem. Rev. 101, 3191–3203
22. Roberts, R. W. (1999) Curr. Opin. Chem. Biol. 3, 268–273
23. Rodi, D. J., and Makowski, L. (1999) Curr. Opin. Biotechnol. 10, 87–93
24. Keele, A. D., and Szostak, J. W. (2001) Nature 410, 715–718
25. Katakas, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Science 262, 1680–1685
26. Lehrer, R. I. (2004) in Advances in Molecular and Cellular Microbiology: Mammalian Host Defense Peptides (Devis, D. A., and Hancock, R. E. W., eds) pp. 5–8, Cambridge University Press, Cambridge, United Kingdom
Synthesis and Selection of Bioactive de Novo Proteins

21. Brogden, K. A., Ackermann, M., and Huttner, K. M. (1998) Infect. Immun. 66, 5948–5954
22. Powers, J. P., and Hancock, R. E. (2003) Peptides 24, 1681–1691
23. Tomasini, L., Scocchi, M., Mettuolo, R., and Zanetti, M. (2004) Antimicrob. Agents Chemother. 48, 3260–3267
24. Hong, R. W., Shchepetov, M., Weiser, J. N., and Axelsen, P. H. (2003) Antimicrob. Agents Chemother. 47, 1–6
25. Chopra, S., and Ranganathan, A. (2003) Chem. Biol. 10, 917–926
26. Stemmer, W. P. (1994) Nature 370, 389–391
27. Corey, M. J., and Corey, E. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 11428–11434
28. Prijambada, I. D., Yomo, T., Tanaka, F., Kawama, T., Yamamoto, K., Hasegawa, A., Shima, Y., Negoro, S., and Urabe, I. (1996) FEBS Lett. 382, 21–25
29. Wei, Y., and Hecht, M. H. (2004) Protein Eng. Des. Sel. 17, 67–75
30. Bittker, J. A., Le, B. V., and Liu, D. R. (2002) Nat. Biotechnol. 20, 1024–1129
31. Klevenza, B., Butz, K., and Hoppe-Seyler, F. (2002) Cell. Mol. Life Sci. 59, 1993–1998
32. Walker, J. R., Altman, R. K., Warren, J. W., and Altman, E. (2003) J. Pept. Res. 62, 214–226
33. Butz, K., Denk, C., Fitscher, B., Crnkovic-Mertens, I., Ullmann, A., Schroder, C. H., and Hoppe-Seyler, F. (2001) Oncogene 20, 6579–6586
34. Colas, P., Cohen, B., Jessen, T., Grishina, I., McCoy, J., and Brent, R. (1996) Nature 380, 548–550
35. Giangaspero, A., Sandri, L., and Tossi, A. (2001) Eur. J. Biochem. 268, 5589–5600
36. Bowden, G. A., and Georgiou, G. (1999) J. Biol. Chem. 265, 16760–16766
37. Deng, H., Nilsson, L., and Kurland, C. G. (1995) J. Bacteriol. 177, 1497–1504
38. Capila, I., and Linhardt, R. J. (2002) Angew. Chem. Int. Ed. Engl. 41, 391–412
39. Aoki, K., Matsumoto, S., Hirayama, Y., Wada, T., Ozeki, Y., Niki, M., Dome, K. Unemori, Y., Yamamoto, S., Mines, A., Matsumoto, M., and Kobayashi, K. (2004) J. Biol. Chem. 279, 39798–39806
40. Andersson, E., Rydengard, V., Senvasson, A., Mangelin, M., Bjorck, L., and Schmidtchen, A. (2004) Eur. J. Biochem. 271, 1219–1226
41. Rusnati, M., Coltrini, D., Oreste, P., Zoppetti, G., Albini, A., Noonan, D., d’Adda di Fagagna, F., Giacca, M., and Presta, M. (1997) J. Biol. Chem. 272, 11313–11320
42. Goncalves, E., Kitas, E., and Seelig, J. (2005) Biochemistry 44, 2692–2702
43. Ravi Chandra, B., Gowthaman, K., Raj Akhouri, R., Gupta, D., and Sharma, A. (2004) Protein Eng. Des. Sel. 17, 175–182