Synthesis and Selection of De Novo Proteins That Bind and Impede Cellular Functions of an Essential Mycobacterial Protein[▽]

Alka Rao,^{1†} Geeta Ram,^{1†} Adesh Kumar Saini,² Reena Vohra,² Krishan Kumar,¹
Yogendra Singh,² and Anand Ranganathan^{1*}

Recombinant Gene Products Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi-110067, India,¹ and Institute of Genomics and Integrative Biology, Mall Road, Delhi-110007, India²

Received 20 October 2006/Accepted 4 December 2006

Recent advances in nonrational and part-rational approaches to de novo peptide/protein design have shown increasing potential for development of novel peptides and proteins of therapeutic use. We demonstrated earlier the usefulness of one such approach recently developed by us, called “codon shuffling,” in creating stand-alone de novo protein libraries from which bioactive proteins could be isolated. Here, we report the synthesis and selection of codon-shuffled de novo proteins that bind to a selected *Mycobacterium tuberculosis* protein target, the histone-like protein HupB, believed to be essential for mycobacterial growth. Using a versatile bacterial two-hybrid system that entailed utilization of HupB and various codon-shuffled protein libraries as bait and prey, respectively, we were able to identify proteins that bound strongly to HupB. The observed interaction was also confirmed using an in vitro assay. One of the protein binders was expressed in *Mycobacterium smegmatis* and was shown to appreciably affect growth in the exponential phase, a period wherein HupB is selectively expressed. Furthermore, the transcription profile of *hupB* gene showed a significant reduction in the transcript quantity in mycobacterial strains expressing the protein binder. Electron microscopy of the affected mycobacteria elaborated on the extent of cell damage and hinted towards a cell division malfunction. It is our belief that a closer inspection of the obtained de novo proteins may bring about the generation of small-molecule analogs, peptidomimetics, or indeed the proteins themselves as realistic leads for drug candidates. Furthermore, our strategy is adaptable for large-scale targeting of the essential protein pool of *Mycobacterium tuberculosis* and other pathogens.

Tuberculosis, the disease caused by *Mycobacterium tuberculosis*, continues to be a major cause of human suffering and mortality in the developing world (15, 31, 52; http://www.who.int/tb/publications/tb_global_facts_sep05_en.pdf). Further compounding this worrying scenario is the fact that the last drug approved for treatment of tuberculosis was approved almost 40 years ago (16). Consequently, there is an urgent need to support the conventional drug discovery processes by exploring potential nonchemical routes and methodologies for developing new molecules against this pathogen. Herein, we describe how de novo protein and peptide inhibitors offer a viable alternative to small-molecule inhibitors that are traditionally selected from combinatorial chemistry- or natural-products-based libraries. To begin with, the diversity that is achievable from a de novo protein library is immense and dependent largely on the library design. Moreover, the area of de novo protein design has increasingly kept pace with advances in molecular biology and computational methods from its initial foundations in peptide chemistry and protein secondary structure knowledge-based approaches (10, 37, 49). Recent additions to this research area are the generation of novel proteins/peptide binders through the use of nonrational and part-rational approaches that do not require any detailed knowledge

of the enzyme target (22, 39, 40). In the context of *M. tuberculosis*, in which roughly one-third of the total number of proteins is completely unknown or made up of conserved hypotheticals (i.e., showing homology only to other mycobacterial proteins), these methods may be particularly useful, more so as protein targets are increasingly being selected from lists that are themselves the outcome of global, whole-genome searches. Accordingly, the identity, structural features, and other biochemical attributes of a target protein, all information that is required when employing a rational approach, become less relevant than just the knowledge that the target protein is essential for the survival of the pathogen.

On the other hand, criticism is sometimes levied on nonrational approaches because they furnish few well-folded and therefore useful proteins. Helpfully, this has fundamentally been addressed by Hecht and coworkers (21, 29). It follows from their binary-patterning concept that de novo proteins that lack substantial binary patterning of polar (○) and nonpolar (●) amino acids (○●○●○●○●○●○ or ○●○●○●○●○●○ for beta-sheet and alpha-helix, respectively) would fail to form adequate secondary structure and therefore not be able to fold properly (35). A nonfolded protein is more amenable to in vivo proteolytic degradation and is therefore less useful (9, 34). Thus, de novo protein libraries wherein members display considerable binary patterning would be a good starting point in the search for drug-like candidates. Even so, there are instances where peptides produced from degenerate oligonucleotide-based libraries have been isolated and show strong binding to the target protein (4, 5, 7). However, the obtained entities in most cases are small (approximately 20 amino acids

* Corresponding author. Mailing address: Recombinant Gene Products Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi-110067, India. Phone: 91-11-26195007. Fax: 91-11-26162316. E-mail: anand@icgeb.res.in.

† Both authors contributed equally to this work.

▽ Published ahead of print on 22 December 2006.

5'-tggtcca-3'	5'-cccgagg-3'	5'-agtact-3'	5'-cagctg-3'	5'-atgcat-3'	5'-tgcgca-3'	5'-tacgta-3'
3'-accggt-5'	3'-ggggccc-5'	3'-tcatga-5'	3'-gtcgac-5'	3'-tacgta-5'	3'-accggt-5'	3'-atgcat-5'
W P	P G	S T	Q L	M H	C A	Y V
5'-ggcgcc-3'	5'-ttcgaa-3'	5'-aacgtt-3'	5'-aagctt-3'	5'-cgtagc-3'	5'-gatatc-3'	5'-gagctc-3'
3'-ccggcg-5'	3'-aagctt-5'	3'-ttgcaa-5'	3'-ttcgaa-5'	3'-gcatgc-5'	3'-ctatag-5'	3'-ctcgag-5'
G A	F E	N V	K L	R T	D I	E L

FIG. 1. The dicodon set used for codon-shuffling experiments.

in length) and are tethered to a comparatively large scaffold. It has been hypothesized that these peptides may not display their original structural fold once removed from the scaffold and, as a result, become prone to proteolytic degradation. Furthermore, the binding activity may, in many instances, be partly due to the bulky scaffold itself (3, 53). In comparison, we have earlier shown that "codon shuffling" is a versatile method for generating de novo protein libraries that possess a considerable diversity of amino acid compositions as well as protein lengths (6, 35). We had also postulated that the constituent library members would theoretically possess adequate secondary structure elements, as they are formed from dicodons (DC) that have defined binary patterning (Fig. 1) (35). Protein libraries that are programmable, display good binary patterning, and show diversity in length and composition represent, we believe, a good pool from which potent candidates can be isolated. We now extend this approach towards the construction of programmable codon-shuffled libraries aimed at selecting de novo proteins that strongly bind and inhibit in vivo the functioning of HupB, an essential *M. tuberculosis* protein.

MATERIALS AND METHODS

Reagents, bacterial strains, plasmids, and growth conditions. Bacteriomatch two-hybrid kit was purchased from Stratagene. X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and the beta-galactosidase inhibitor 2-phenylethyl beta-D-thiogalactoside were purchased from Sigma. *Escherichia coli* strains BL21(DE3) and TOP10 and the two-hybrid XL-1 Blue MRF⁺::Kan^r-derived reporter strain were obtained from Novagen, Invitrogen, and Stratagene, respectively. Plasmids pGEX4T3 and pThiohisA were from Amersham Biosciences and Invitrogen, respectively. Routine cloning and transformation procedures for *E. coli* were as described earlier (41). All new plasmids were sequenced using the dideoxy method in order to validate their authenticity.

Construction of pBT- and pTRG-derived plasmids pBTnn and pTRGnn. A 587-bp-long DNA fragment from plasmid pBT was PCR amplified using the oligonucleotides 5'-AAGCGGCCGCGGATCCTACGTATGAGAATTCTA GACCTCGAGTTAATTAATTAATTAAGATCT-3' and 5'-CACCATGGGG AAATATTATACGCAAGGCGAC-3' as the forward and reverse primers, respectively. The PCR product was cloned in the pBluescript SRF vector, the fragment excised using NcoI and NotI restriction enzymes and cloned in a pBT vector previously cut with the same two enzymes. The resulting plasmid was designated pBTnn. Plasmid pBTnn carries the SnaBI site into which DC fragment libraries can be inserted. The site immediately follows the end of the λ cI gene and just precedes a stop codon. Expression of this plasmid would result in libraries of codon-shuffled proteins, all fused with λ cI. For the creation of plasmid pTRGnn, a 580-bp BamHI fragment from the pBluescript SRF plasmid that contained the PCR product mentioned above was excised and cloned in the BamHI-cut pTRG. A clone that contained the BamHI fragment in the wrong orientation was specifically selected and digested with XhoI, and the vector was self-ligated to finally yield pTRGnn. Plasmid pTRGnn carries the SnaBI site into which the codon-shuffled libraries can be inserted. The site immediately follows the end of the gene encoding the α subunit of RNA polymerase (RNAP) and just precedes a stop codon. Expression of this plasmid would result in libraries of codon-shuffled proteins, all fused with the α subunit of RNA polymerase.

Construction of plasmid pGEX4T3nn. The 580-bp BamHI fragment that was used for the construction of pTRGnn was cloned in the BamHI-cut pGEX4T3. A clone that contained the BamHI fragment in the wrong orientation was specifically selected and digested with XhoI, and the vector was self-ligated to

finally yield pGEX4T3nn. Plasmid pGEX4T3nn carries the SnaBI site into which DC fragment libraries can be inserted. The site immediately follows the end of the glutathione S-transferase (GST) gene and just precedes a stop codon. Expression of this plasmid would result in libraries of codon-shuffled proteins, all fused with the GST protein.

Construction of plasmid pThiohisann. A 945-bp XbaI fragment was excised from pBTnn and inserted in the XbaI-cut pThiohisA. A clone that contained the XbaI fragment in the right orientation was specifically selected and digested with BamHI, and the vector was self-ligated to finally yield pThiohisann. Plasmid pThiohisann carries the SnaBI site into which DC fragment libraries can be inserted. The site immediately follows the end of the *trx* gene and just precedes a stop codon. Expression of this plasmid would result in libraries of codon-shuffled proteins, all fused with the TRX protein.

Construction of plasmid series containing *M. tuberculosis* HupB and its derivatives. (i) **pBTnnHupB and pTRGnnHupB.** The full-length *hupB* gene (Rv2986c) was PCR amplified from *M. tuberculosis* H37Rv genomic DNA using the oligonucleotides 5'-CCTACGTAATGAACAAAGCAGAGCTCATTGAC GTG-3' and 5'-CCTACGTATTGCGACCCCGCCGAGCGTTGCC-3' as the forward and reverse primers, respectively. The 658-bp PCR product was isolated, purified, and subsequently cloned in pBluescript SRF. The resulting plasmid was digested with SnaBI, and the 648-bp-long *M. tuberculosis* *hupB* gene was inserted in frame in plasmids pBTnn and pTRGnn, which were earlier cut with SnaBI. Expression of the resulting plasmids, pBTnnHupB and pTRGnnHupB, in a suitable two-hybrid host results in λ cI-HupB and RNAP α subunit-HupB fusion proteins.

(ii) **HupBpGEX4T3nn.** The SnaBI-cut full-length *hupB* gene was inserted in plasmid pGEX4T3nn, which was earlier cut with the same restriction enzyme, to yield plasmid hupBpGEX4T3nn. Expression of this plasmid in *E. coli* resulted in a GST-HupB fusion protein (molecular mass [MW], 49.0 kDa; the MW of GST alone is 26.57 kDa).

(iii) **pBTnnHup2 and pTRGnnHup2.** The DNA fragment corresponding to the C-terminal half of the *M. tuberculosis* HupB protein was PCR amplified using the oligonucleotides 5'-GGATACGTAGCTGTTAAGCGTGGTGTGGG-3' and 5'-CCTACGTATTGCGACCCCGCCGAGCGTTGCC-3' as the forward and reverse primers, respectively. The 358-bp-long PCR product was isolated, purified, and subsequently cloned in pBluescript SRF. Using procedures similar to those described for *hupB*, the SnaBI-flanked *hup2* gene was inserted in plasmids pBTnn, pTRGnn, pGEX4T3nn, pThiohisann, and pMALc2x to generate *hup2* gene fusions that would result in Hup2 fusion proteins with λ cI, the RNAP α subunit, GST, thioredoxin (TRX), and maltose-binding protein (MBP), respectively.

Construction of codon-shuffled libraries. The protocol for library generation has been described earlier (35) and was faithfully followed, except that in the present sets of experiments, the proportions of dicodons were skewed and not kept equimolar. The degrees of skewing were as follows.

Library-versus-library (LvL) experiments. For library 1, KL and RT dicodons were added one, three, and five times in molar excess and DI and EL dicodons were removed from the initial mix. For library 2, DI and EL dicodons were added one, three, and five times in molar excess; KL and RT dicodons were removed from the initial mix.

HupB/Hup2 experiments. DI and EL dicodons were added one and three times in molar excess; KL and RT dicodons were removed from the initial mix.

All libraries preferentially contained DC-fragment sizes larger than 250 bp (this corresponds to DC proteins at least 80 amino acids in length). Purified libraries were stored at -20°C until further use.

Two-hybrid assays (in vivo). All in vivo protein-protein interaction studies were carried out using the Bacteriomatch two-hybrid system (13). The Bacteriomatch reporter strain, bearing the ampicillin resistance gene and the *lacZ* reporter gene, was cotransformed with equal amounts of bait (pBTnn-derived plasmids or libraries) and prey (pTRGnn-derived plasmids or libraries) DNA. Protein-protein interaction was detected by observing the level of transcriptional activation of *lacZ* reporter gene in various cotransformants of the reporter strain, when the latter was plated on X-Gal plates that were supplemented with appro-

priate antibiotics (chloramphenicol and tetracycline), 25 μ M IPTG (isopropyl- β -D-thiogalactoside), and β -galactosidase inhibitor. Plasmids pBT-LGF2 (λ cI-LGF2 fusion) and pTRG-Gal11P (RNAP α subunit-Gal11P fusion) provided the positive controls, whereas empty pBTnn and pTRGnn plasmids served as negative controls for the interaction studies. Routine protocols performed during the assay were according to the manufacturer's instructions.

Liquid β -galactosidase assays. Confirmation of the in vivo interactions detected on X-Gal plates was obtained by the measurement of beta-galactosidase (LacZ) activity in a given cotransformant strain. Colonies were grown at 30°C to mid-log phase in the presence of 25 μ M IPTG. Cells were permeabilized with sodium dodecyl sulfate (SDS)-CHCl₃ and assayed for β -galactosidase activity essentially as described previously (27). Assays were performed twice, in duplicate, with different cotransformants on separate occasions, and the values reported are the means of these measurements.

Construction of plasmids LvL7+pthiohisann and LvL7-pthiohisann. The full-length clone 7 gene of the positively charged library (LvL7+) (318 bp) and the full-length clone 7 gene of the negatively charged library (LvL7-) (360 bp) were PCR amplified from their pBTnn and pTRGnn clones, respectively, using the 5'-phosphorylated oligonucleotide 5'-AGCGGCCGCGGTTTAA-3', which served as both a forward and reverse primer. The PCR products were eluted using DEAE membrane and inserted into the SnaBI-cut pThiohisann to yield plasmids LvL7+pthiohisann and LvL7-pthiohisann. Expression of these plasmids resulted in the production of TRX-LvL7 fusion proteins.

Expression and purification of TRX-LvL7+ and TRX-LvL7- proteins. Liquid cultures of *E. coli* DH5 α strains containing plasmids LvL7+pthiohisann and LvL7-pthiohisann were independently grown to mid-log phase (optical density at 600 nm [OD₆₀₀], 0.4 to 0.6) and then induced with 0.5 mM IPTG. The induction was continued for a further 4 h at 37°C, after which the cells were pelleted, washed, and resuspended in buffer A (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.8% NaCl). Postlysis and removal of cell debris, the supernatant was loaded onto a Ni-nitrilotriacetic acid column that had previously been equilibrated with buffer A. As both proteins are TRX fusions, they are expected to bind the nickel column by virtue of the thioredoxin His patch. Subsequently, the proteins were eluted with buffer A containing 250 mM imidazole. Protein fractions were visualized by 15% SDS-polyacrylamide gel electrophoresis (PAGE) (the MW of TRX-LvL7+ and TRX-LvL7- are 24.33 kDa and 26.03 kDa), and those containing purified proteins were pooled and dialyzed against 10 mM Tris-HCl, pH 8.0. The proteins were concentrated using YM-5 concentrators and utilized for circular dichroism (CD), matrix-assisted laser desorption/ionization (MALDI), and in vitro analysis experiments.

In vitro ion-exchange pull-down assay. The pull-down assay was conducted using two different approaches. In the first approach, the purified proteins TRX-LvL7+ and TRX-LvL7- were taken in equal amounts (10 μ g; phosphate buffer, pH 7.0) and mixed for 1 h at 4°C. The protein mixture was added to 100 μ l of SP-Sepharose beads and the mix further incubated for 1 h. The beads were washed extensively with 50 to 150 mM NaCl (phosphate buffer, pH 7.0) and the proteins eluted with 500 mM NaCl. In the second approach, 10 μ g of TRX-LvL7+ was bound onto 100 μ l of SP beads and the mix incubated for 1 h at 4°C. To this was added 10 μ g of TRX-LvL7- protein, and the mixture was further incubated with mild shaking for 1 h at 4°C. The protein recovery procedure was identical to the one used in the first approach. Purified TRX protein and unbound SP beads acted as controls in the experiments mentioned above. The proteins were visualized using Coomassie blue on 15% SDS-PAGE gels.

Expression and purification of GST-C4 protein. The protein GST-C4, as well as GST alone, was found to be preferentially expressed in the soluble form. GST-C4 and GST were routinely purified using glutathione-Sepharose 4B beads. The beads were washed extensively with buffer B (10 mM Tris-HCl, pH 8.0, 14 mM β -mercaptoethanol, 0.1% Tween 20, 20% [vol/vol] ethanol, 0.5 M NaCl), and the proteins eluted with buffer C (10 mM reduced glutathione, 50 mM Tris-HCl, pH 7.5). Purified fractions of eluted proteins were analyzed by 15% SDS-PAGE and subsequently pooled and dialyzed.

In vitro protein interaction assay. The in vitro protocol was essentially as described earlier (36), with some variations. Briefly, the purified GST-C4 protein (10 μ g) or the GST protein alone (5 μ g) was incubated with 45 μ g of glutathione-Sepharose 4B beads at 4°C for 1 h, followed by repeated washing and blocking with 1% (wt/vol) polyvinylpyrrolidone at 4°C for 1 h. Subsequently, the beads were incubated with purified recombinant Hup2 protein (10 μ g) at 4°C for 4 h. Finally, after extensive washings, all bound protein was eluted by boiling the beads in 50 μ l of 2 $\times$ SDS loading buffer (20% [vol/vol] glycerol, 4% SDS, 0.125 M Tris-HCl, pH 6.8, 4% [vol/vol] β -mercaptoethanol) and resolved by 15% SDS-PAGE for subsequent visualization by Coomassie staining. Various controls (e.g., beads alone or GST plus Hup2) for the pull-down assay were also subjected to the same treatment.

Construction of plasmid pSD5GFP. The full-length green fluorescent protein (GFP) gene was PCR amplified from the commercially available vector pEGFPn1 (Amersham) using the oligonucleotides 5'-TCCACCCTGCAGTAC TATGGTGAGCAAG-3' and 5'-AGAGTCGCGACGCGTTACTGTACA G-3' as the forward and reverse primers, respectively. The 750-bp-long PCR product was isolated, purified and subsequently cloned in pBluescript SRF. From this plasmid, the GFP gene was excised using PstI and MluI and cloned in plasmid pSD5, which was earlier cut with the same two restriction enzymes. The resulting plasmid carries a unique ScaI site just upstream of the GFP initiation codon. The insertion of genes into this site, followed by expression of the resulting plasmids, yields GFP fusion proteins.

Construction of pSD5-C and pSD5GFP-C series plasmids. Full-length genes corresponding to the "C series" were PCR amplified from their pBTnn or pTRGnn clones, using the 5'-phosphorylated oligonucleotide 5'-AGCGGCCG CCACGTGTTTAAA-3', which served as both a forward and reverse primer. The PCR products were eluted using the DEAE membrane and inserted in ScaI-cut pSD5, as well as ScaI-cut pSD5GFP, to yield plasmids ready for expression in *Mycobacterium smegmatis*. All inserts reside upstream of the GFP gene in pSD5GFP.

Growth analysis and colony count of mycobacteria. *Mycobacterium smegmatis* strain LR222 (28) was used throughout the course of this study. Mycobacterial transformation procedures were as described earlier (19). Five-milliliter starter cultures of *M. smegmatis* strains with pSD5 as well as pSD5C4 were prepared from frozen stock in Middlebrook 7H9 medium (Difco Laboratories) supplemented with 1% (vol/vol) ADC medium (5% albumin, 2% dextrose) and 0.05% Tween 80. Cultures were incubated at 37°C with aeration for 3 days, or until the stationary phase of growth was achieved. These were used as seed cultures for subsequent investigations. For growth curve analysis, 100-ml secondary cultures were seeded with 1% of primary cultures and incubated at 37°C. Periodically, aliquots were removed over a period of 50 h and their OD₆₀₀ measured. For CFU analysis, the 100-ml secondary cultures were grown at 37°C until mid-log phase (OD₆₀₀ of 0.9) was reached. At this point, the growing cultures were shifted to a 10°C water bath shaker (180 rpm). Subsequently, 10- μ l aliquots were removed and made up to 100 μ l with liquid media. Fifty microliters of this culture was plated onto agar plates containing kanamycin and the plates incubated for 3 days at 37°C. Finally, the number of colonies that appeared for each set (the total number of days was 10) was plotted as a percent increase over the number present on day 0.

RNA extraction. Fifty-milliliter cultures of wild-type and recombinant strains of *M. smegmatis* were grown at 37°C. Cell pellets were rinsed with ice-cold diethyl pyrocarbonate-treated sterile double-distilled water and resuspended in 400 μ l of TES [N-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid] buffer. To the tubes was added 400 μ l water saturated with phenol-chloroform/isoamylalcohol and the contents mixed vigorously. Glass beads were added for efficient cell disruption. The tubes were incubated at 65°C for 1 h with intermittent mixing, following which the tubes were incubated on ice for 2 min and then centrifuged. Aqueous phase was transferred to fresh tubes and the RNA was precipitated with 3 M sodium acetate (pH 5.2, 0.1 volumes) and chilled ethanol (2 volumes). After incubation at -20°C for 30 min, RNA samples were pelleted (10,000 rpm, 4°C, 15 min), washed with 70% ethanol, dried, and resuspended in 20 μ l of diethyl pyrocarbonate-treated sterile water. RNA was checked for quality and quantity spectrophotometrically as well as by denaturing agarose gel electrophoresis.

Blotting. Total RNA (20 μ g) was electrophoresed on 1% denaturing agarose-formaldehyde gel, photographed to check for comparable levels of both samples, and then transferred to nylon membrane (Hybond N+; Amersham), following which the membrane was exposed to UV light for cross-linking. Transfer was checked with methylene blue. Linear DNA probe was generated using *M. smegmatis* hupB-specific oligonucleotides (5'-CCCTACGTAATGA ACAAGCGGAGCTCATCGAC-3' and 5'-CCCTACGTACCTGCGGCC TTCTTGCCGG-3' as forward and reverse primers, respectively) and labeled using a nick translation system (Invitrogen) according to the manufacturer's specifications. After denaturation, this probe was added to the prehybridization buffer A (6 $\times$ SSC [1 $\times$ SSC is 0.15 M sodium chloride plus 0.015 M sodium citrate], 50% formamide, 0.5% SDS, 100 μ g/ml salmon sperm DNA, 5 $\times$ Denhardt's reagent) and incubated at 65°C for 3 to 6 h with the membrane. The membrane was subsequently washed first with buffer B (2 $\times$ SSC, 2% SDS) at room temperature for 15 min and then with buffer C (0.5 $\times$ SSC, 2% SDS) at 65°C for 45 min. The membrane was finally washed with buffer D (0.2 $\times$ SSC) to remove excess SDS, salts, and free radioactivity. Intensity was measured by a PhosphorImager (Amersham Biosciences). The profile seen was reproducible.

A.

MTU	HupB	MNKAELIDVLTQKLGSDRRQATAAENVVDTIVRAVHKGDSVTITGFGVFQRRRAARVARNPRTGETVVKVKPTSVPAFRPGAQFKAUVSGAQRLPAAEGPAVKRGVGAS	
MSM	HupB	MNKAELIDVLTQKMGTDRRQATAAENVVDTIVRAVHKGDSVTITGFGVFQRRRAARVARNPRTGETVVKVKPTSVPAFRPGAQFKAUVSGAQRLPAAEGPAVKRGVGTAG	

MTU	HupB	AAKQVAKKAPAKKAT---KAAKKAATKAPARKAATKAPAKKAATKAPAKKAVKATKSPAKVTVTKVKTAVKASVRKAATKAPAKKAAAKRPATKAPAKKATARRGRK	214
MSM	HupB	PAKKAAPAKKAPAKKAAAKKATKAAAKKAPAKKAATKAPAKKAATKAPAKKA---ATKAPAK-----KAATKAPAKKAAAKAPAKKAATKAPAKKAAAKKAPAKKGRR	208

B.

HpbNT	MNKAELIDVLTQKLGSDRRQTAAVENVDITIVRAVHKGDSVTITGFGVFQRRAARVARNPRTGETVKVKPTSVPAFRPQAQFKAVVSQAQLPAEGP	100
HU DBP	MNKTELINAVAETSGLSKKDATKAVDAVFDISITALRKGDQKVLIIGFNGFEVRERAARKGRNPOTGEEMIIPASKVPAPFKPGKALKDAVK-----	90
HU A2	MNKTLQDLDIAEKAEBSKTQAKAALSTLAAITESLKEGDVAQLVGFFGTFTKNVHAERTGNRPOTGKEIKIAANVPAFVSGKALKDAVK----- ***:*::*:~::~**:**:~::::*****~:::****~:::~:::	90

C.

[illegible]

D.

FIG. 2. (A) Amino acid sequence alignment of full-length *M. tuberculosis* (MTU) and *M. smegmatis* (MSM) HupB proteins. For panels A through C, asterisks indicate identity while colons and periods indicate high and low similarities, respectively. (B) Amino acid sequence alignment of the N-terminal half of *M. tuberculosis* HupB with its closest homologues (revealed from BLAST search). HpBNt, N-terminal half of *M. tuberculosis* HupB; HU DBP, DNA-binding protein Hu from *Bacillus stearothermophilus*; HU A2, $\alpha 2$ protein HU from *E. coli*. (C) Amino acid sequence alignment of the C-terminal half of *M. tuberculosis* HupB with its closest homologues (revealed from BLAST search). HpBCt, C-terminal half of *M. tuberculosis* HupB; MBO, *Mycobacterium bovis*; MLE, *M. leprae*; MSM, *M. smegmatis*; MAV, *M. avium*. (D) Models depicting ribbon diagrams and surface electrostatic potentials of the N-terminal (left panel) and C-terminal (right panel) halves of the *M. tuberculosis* HupB protein. The models were generated using threading algorithms and drawn using the molecular graphics programs RIBBONS and MOLMOL.

Circular dichroism studies. CD spectra were recorded on a JASCO model J-810 spectropolarimeter. Data were collected at room temperature (25°C) with a pathlength of 1 mm. The concentrations of proteins (suspended in 1× phosphate-buffered saline, pH 8.0) were 12.5 μM for TRX-LvL7+ and 10.0 μM for TRX-LvL7−.

TEM studies. Fresh colonies of *M. smegmatis* LR222 transformed with expression plasmids pSD5 and pSD5C4 were cultured until mid-log phase. Five milliliters of each culture was centrifuged and the pellet washed once with buffer A (10 mM phosphate-buffered saline, pH 7.4). For cell fixing, the cell pellet was fixed using 2.5% glutaraldehyde and 2% paraformaldehyde on ice for 45 min and subsequently washed with buffer A. Pellets were postfixed with 1% buffered osmium tetroxide (1 h, room temperature), followed by dehydration against a graded series of acetone (30 to 100%). The postfixed samples were infiltrated with resin and toluene and were finally embedded in LR white resin (TAAB, United Kingdom). Polymerization of resin was done at 60°C for 12 h. For transmission electron microscopic (TEM) analysis, ultrathin sections (~70 nm) of each embedded sample were cut and lifted on copper grids followed by staining with 4% uranyl acetate and 0.4% lead acetate for 5 min. The sections were examined and photographed using a FEI-PHILIPS MORGAGMI.268D digital transmission electron microscope.

Protein analysis. SDS-PAGE analysis was carried out on 15% gels by using a method described previously (24). All protein samples were treated with mercaptoethanol before they were loaded onto SDS-PAGE gels. Multiple amino acid sequence alignment was carried out using the CLUSTALW program (available at <http://www.ebi.ac.uk>). Theoretical modeling of HupB and C4 proteins was carried out using the 3DPSSM and PHYRE algorithms (available at <http://www.sbg.bio.ic.ac.uk/>, courtesy of Structural Bioinformatics Group, Imperial College, London, United Kingdom).

RESULTS AND DISCUSSION

Suitability of the codon-shuffling method and choice of target *M. tuberculosis* protein. We decided to employ the bacterial two-hybrid system for selection of de novo protein binders against the chosen *M. tuberculosis* target. Additionally, we reasoned that instigation of a charge-based interaction between the bait and prey could be effectively employed for selecting protein binders. It has previously been reported that charge-based forces that lead to protein-protein interactions constitute a strong primary force that includes effects such as charge-based induced fit, electrostatic steering, hot-spot navigation, and electrostatic tethering (23, 45, 47, 51). Here it is noteworthy to emphasize an exclusive advantage of codon shuffling over other methods, namely, its allowance for programming protein libraries in order to obtain proteins conditionally skewed in properties of charge, hydrophobicity, or secondary structure. For example, by skewing the proportions of the KL and RT (positively charged) or DI and EL (negatively charged) dicodons in the total dicodon pool, codon shuffling can create libraries that display charge preferences. Additionally, the complete removal of either one of the charged dicodon sets from the pool would result in proteins that are exclusively negatively or positively charged.

For the choice of *M. tuberculosis* protein target, we drew our attention to the recent seminal contributions of Rubin and coworkers in the area of identification of essential *M. tuberculosis* proteins (42). Using an incisively elegant transposon-based methodology called TRASH, Rubin and coworkers have drawn up lists of proteins that are essential to *M. tuberculosis*, albeit the term “essential” is used in an assortment of contexts, ranging from general growth of the pathogen to its growth inside a macrophage, etc. (38, 43, 44). Based on an inspection of cutoff ratios (which characterize the degree of essentiality) and isoelectric points (pI) of the essential proteins, we narrowed down our choice to the protein HupB (also called Hlp, or histone-like protein). This protein has a low cutoff ratio (an indication of a high probability of essentiality) and a very high pI of 12.5 (a good bearing for instigation of a charge-based interaction). Moreover, the behavior and properties of mycobacterial HupB have been extensively characterized (26, 33). HupB has been shown to play an important role in cellular adhesion of mycobacteria, an event that is crucial for gaining entry into macrophages and other host cells (48). Additionally, HupB has been shown to be hyperinduced during anaerobically induced dormancy as well as under cold shock stress (25). Under normal mycobacterial growth conditions, HupB is preferentially expressed during the exponential phase of growth (46). Knockout of the *hupB* gene in mycobacteria results in an inability to withstand cold shock stress, as well as the inability to bind to Schwann cells (25, 48). It is also hypothesized that HupB binds to DNA and plays a crucial role in the regulation of cellular growth (26, 46). The primary sequence characteristics of HupB also merit a brief mention. It is a reasonably sized protein (MW, 22 kDa) of strikingly different N- and C-terminal halves. While the N-terminal half of the protein has a more or less neutral charge and shows strong homology to histone-like proteins of other organisms like *E. coli* and *Streptomyces coelicolor*, the C-terminal half is highly positively charged and shows homology only to proteins of only mycobacteria, for which the sequence identity is upwards of 80% (90% with the *M. smegmatis* histone-like protein) (Fig. 2A to C). There is as yet no structural information available for HupB. Nonetheless, three-dimensional models obtained through threading bring out clearly the degree of dissimilarity between the two halves of HupB (Fig. 2D). Therefore, de novo protein binders that target the C-terminal half of this protein could also bind to similar proteins across the mycobacterial spectrum. Given that our target protein HupB is very positively charged, one could envisage the construction of a very negatively charged library for the two-hybrid experiments. However, selection of binders through a charge-based interaction has not been previously reported. As a proof of principle, we designed an experiment to test this very concept.

Charge-based selection of de novo protein partners. In the bacterial two-hybrid experiment, the bait and prey proteins are fused with λ cI repressor protein and the α subunit of RNA polymerase, respectively, or vice versa. Any worthwhile interaction between the two accelerates transcription of a marker gene, ultimately resulting in the selection of cells wherein the interaction is occurring. The extent of interaction can be scored by a variety of means, most commonly through the appearance of a blue colony when the marker gene is *lacZ*. First, we altered the expression plasmids pBT and pTRG,

A.

B.

FIG. 3. (A) A plate showing the in vivo two-hybrid results obtained in the course of this study. The 16 individual colony streaks displayed are of strains that have the indicated combinations of pBT and pTRG plasmids and are numbered as follows: 1, pBTLGF2 plus pTRGGAL11^P (positive control); 2, pBTnn plus pTRGnn (negative control); 3, pBTnnLvL7+ plus pTRGnn; 4, pTRGnnLvL7- plus pBTnn; 5, pBTnnLvL7+ plus pTRGnnLvL7-; 6, pTRGnnC4 plus pBTnn; 7, pBTnnHupB plus pTRGnn; 8, pTRGnnC4 plus pBTnnHupB; 9, pTRGnnHup2 plus pBTnn; 10, pBTnnC4 plus pTRGnn; 11, pBTnnHup2 plus pTRGnnC4; 12, pTRGnnHup2 plus pBTnnC4; 13, pBTnnHup2 plus pTRGnn; 14, pTRGnnHup2 plus pBTnnC5; 15, pBTnnC5 plus pTRGnn; 16, two-hybrid reporter strain. (B) β -Galactosidase (gal) liquid assay of cell lysates of each cotransformant strain after induction with 20 μ M IPTG. Numbers 1 to 16 correspond to the cotransformants shown in panel A and are as described above. Gray bars represent cotransformant strains that displayed blue color on the plate (see panel A).

which carry the λ cI and RNAP α subunit-encoding genes, so that our codon-shuffled libraries could be inserted downstream and in frame to the genes mentioned above. This resulted in λ cI-fused and RNAP α subunit-fused protein libraries (see Materials and Methods). We then created two sets of highly

Protein	Sequence	Length (aa)	MW (kDa)	pI	SS (%)		
					H	S	R
LvL7+	SGRHVF FGANVSTKLSTSTR TMHMKLSTMHSTMHRTSTRTGAKLRTS TRTSTRTMHMHSTQLMHSTRTNVRTMHSTKLMHGARTMHRTMHGASTR TKL FKHVAAA	106	11.98	13.04	42	16	42
LvL7-	SGRHVF KPGSTPGDISTEL MHDIDISTDIDIDIDIDIDIDISTELD IELDIMHDIDIDIELDIMHELDIDIDIDIMHELDIMHDIIYVDIMHD IDISTDIELDIDIDIDIF KHVAAA	120	13.67	3.72	41	08	51
LvL3+	SGRHVF KMHNVRTSTM MHMKLSTRTKLMHGAPGSTMHRTSTRTGAQLS TKLRTMHRTGAQLKLRTKLRTSTRTKLKL FKHVAAA	86	9.77	12.96	42	18	40
LvL3-	SGRHVF KELDIQLMHDIDIEL DIDIDQLSTSTMHDIDIELDIDIMHELD IF KHVAAA	56	6.47	4.30	49	18	33
C4	SGRHVF KPGMHYVMHGASTM HDIDINVSTMHPGDIMHDMHNVPGSTD IDIELELDIMHPGMHSTMHPGDIGADIDIMHDI FKHVAAA	88	9.70	5.43	24	20	56
C5	SGRHVF KELDIELDIDIDIDID IDIELMHDIQLDIELDIDIMHMD IPGDIDIMHSTMHQLDIELMHPGDIDIQ LFKHVAAA	84	9.70	4.10	42	06	52
C6	SGRHVF KGADIDIQLDIDISTG ADIMHMMHDMHDMHNVMHDMHHE LDIQLMHGADIDIQLELDIQLDIMHCAMHSTD IQLFKHVAAA	90	10.27	4.96	54	18	28

FIG. 4. List of codon-shuffled de novo proteins described in this work. The primary sequence characteristics as well as consensus secondary structure predictions are also shown. aa, amino acid; pI, isoelectric point; SS, secondary structure; H, helix; S, sheet; R, random. The seven-amino-acid-long N- and C-terminal portions of the sequences (shown in gray) are the hairpin-scaffold sequences that flank each de novo protein. Proteins C4 through C6 are *M. tuberculosis* HupB binders.

charged libraries. In one set, the proportions of DI and EL dicodons were increased by three and five times with respect to the other dicodons, while at the same time, the KL and RT dicodons were completely removed from the pool. Correspondingly, in the other set, the proportions of KL and RT were increased by three and five times with respect to the other dicodons, while DI and EL were completely absent. Using the codon-shuffling method described previously (35), we obtained a total library size of approximately 10⁵ CFU in both cases. The oppositely charged libraries were cotransformed in the two-hybrid reporter strain. We found fairly equal prevalences of blue and white colonies. As mentioned earlier, a blue colony is indicative of an interaction having occurred, while a white colony indicates the absence of interaction. We isolated the pTRG and pBT plasmids from one representative colony (LvL7) that displayed the same intensity of blue as the positive control (Fig. 3A, colony streak 5). The isolated plasmids were used to again cotransform the reporter strain to rule out the possibility of a false positive. DNA sequencing of the plasmids showed incorporation of the codon-shuffled fragments downstream of the λ cI and RNAP α subunit-encoding genes, respectively. The corresponding protein sequence, along with primary protein characteristics, is shown in Fig. 4. The proteins, labeled LvL7+ and LvL7−, carry opposite charges, as reflected by their extreme pIs. Both proteins could routinely be overexpressed and purified as thioredoxin fusion partners (Fig. 5A and B). MALDI–time-of-flight analysis of the proteins confirmed their molecular masses with a deviation of 0.001% from the expected values while circular-dichroism data on the two proteins indicated good secondary structure formation, with one protein (LvL7−) calculated to hold almost 40% beta-sheet content (results not shown). The interaction between LvL7+

and LvL7− was further confirmed by an ion-exchange pull-down assay. Protein LvL7+ bound strongly to the negatively charged SP matrix and was seen to pull down protein LvL7− (Fig. 5C). On the other hand, we also isolated the pBT and pTRG plasmids from a representative white colony. Again, DNA sequencing showed both plasmids to be carrying codon-shuffled fragments; the predicted protein sequences showed the bait and target to be highly charged (one negatively and the other positively) (Fig. 4), yet the two proteins (LvL3+ and LvL3−) were not interacting in vivo. This represents a direct confirmation of the hypothesis that while a charge-based interaction allows two proteins to come together, the strength and stability of the interaction depends on electronic-steering, induced-fit, and surface hydrophobic types of interactions between the two proteins. Indeed, X-ray or nuclear magnetic resonance-based structural studies may shed more light on what allows LvL7 and not the LvL3 proteins to interact. Meanwhile, for the present objective of isolating binders against *M. tuberculosis* HupB, the results of the LvL experiments aptly illustrate the judgment of choosing protein partners on a charge-based selection, given the extremely high charge of the target.

Isolation of protein binders against the HupB protein. Our next objective was to isolate de novo protein binders against *M. tuberculosis* HupB and then explore the possibility that these binders inhibit the functioning of HupB in its natural mycobacterial environment. The full-length *hupB* gene was PCR amplified from *M. tuberculosis* H37Rv genomic DNA and inserted just downstream and in frame to the λ cI and RNAP α subunit-encoding genes in the two-hybrid pBT and pTRG plasmids, respectively. Separately, the *hupB* gene was cloned as a GST fusion to check for expression of the protein in *E. coli*

FIG. 5. (A) SDS-PAGE showing the expression and purification of TRX-LvL7+ protein. Lane 1, Pharmacia low-molecular-mass markers (94, 67, 43, 30, and 20 kDa); lane 2, IPTG-induced *E. coli* TOP10/pThiohisL7+ total cell lysate; lane 3, flowthrough; lane 4, wash 1; lane 5, wash 2; lane 6, elution of TRX-LvL7+ (MW, 24.3 kDa) with buffer containing 250 mM NaCl; lane 7, elution of TRX-LvL7+ with buffer containing 500 mM NaCl; lane 8, elution of TRX-LvL7+ with buffer containing 750 mM NaCl. (B) SDS-PAGE showing the expression and purification of TRX-LvL7- protein. Lane 1, IPTG-induced *E. coli* TOP10/pThiohisL7- total cell lysate; lane 2, Pharmacia low-molecular-mass markers (67, 43, 30, 20, and 14 kDa); lane 3, IPTG-induced *E. coli* TOP10/pThiohisL7- total cell lysate; lane 4, flowthrough; lane 5, wash 1; lane 6, wash 2; lane 7, elution of TRX-LvL7- (MW, 26.0 kDa) with buffer containing 500 mM NaCl; lane 8, elution of TRX-LvL7- with buffer containing 500 mM NaCl. (C) SDS-PAGE illustrating the in vitro ion-exchange pull-down assay involving the two LvL proteins TRX-LvL7+ and TRX-LvL7-. Lane 1, SP-Sepharose beads; lane 2, Pharmacia low-molecular-mass markers (94, 67, 43, 30, 20, and 14 kDa); lane 3, Q-Sepharose beads plus TRX alone (MW, 12.37 kDa); lane 4, Q-Sepharose beads plus TRX-LvL7- (MW, 26.0 kDa; indicated by the gray arrow); lane 5, SP-Sepharose beads plus TRX alone; lane 6, SP-Sepharose beads plus TRX-LvL7+ (MW, 24.3 kDa; indicated by the black arrow); lane 7, SP-Sepharose beads plus TRX-LvL7-; lane 8, SP-Sepharose beads-TRX-LvL7+ plus TRX-LvL7-; lane 9, SP-Sepharose beads plus TRX-LvL7+-TRX-LvL7-; lane 10, SP-Sepharose beads-TRX-LvL7+ plus TRX. A point of note is the appearance of LvL proteins not at their expected positions on SDS-PAGE gels but rather at positions that indicate a greater MW. This behavior is routinely observed for highly charged proteins (32). Indeed, MALDI-time-of-flight analysis of a sample of the purified TRX-LvL7+ protein (lane 6) gives an MW value of 24.3 kDa, identical to the expected value.

(Fig. 6). It was previously mentioned that not only the entire HupB protein but also the C-terminal half, which shows exclusive homology to only mycobacterial proteins, represents an attractive target. In light of this, we also created a truncated version of the *hupB* gene, coding for just the C-terminal half, and fused it with the λ cI and RNAP α subunit-encoding genes of pBT and pTRG. The C-terminal protein, termed Hup2, was also seen to be expressed in *E. coli* as a GST fusion protein (Fig. 6). The two-hybrid plasmids carrying the two targets, *hupB* and *hup2*, were cotransformed with a range of negatively charged libraries wherein the KL and RT dicodons were removed and the proportions of DI and EL were gradually skewed, from one to five times with respect to the other 10 dicodons. In all cases, the library sizes ranged from 10^4 (when normal competent cells were used) to 10^6 (when electrocompetent cells were used) CFU. We were able to isolate three separate intensely blue-colored colonies, C4, C5, and C6 (C4 in HupB experiments and the other two in Hup2 experiments). The pBT and pTRG plasmids were isolated from each colony and used to freshly cotransform the reporter strain to rule out false positives. Additionally, the codon-shuffled DNA frag-

ments were PCR amplified from these plasmids and cloned afresh into pBT and pTRG plasmids and experiments repeated. All three clones reproducibly gave blue colonies (Fig. 3A). The amino acid sequence and primary characteristics of the binders are shown in Fig. 4. All proteins show a very high negative charge but nonetheless share little homology with each other. Consensus secondary structure predictions on the three binders mirror their sequence dissimilarity by showing differing compositions of alpha-helical and beta-sheet folds. We then carried out a beta-galactosidase liquid assay to determine the strength of interactions in vivo for the obtained binders (Fig. 3B). As the results indicate, all three show strong interactions compared to the positive control (the positive control proteins Gal11P and LGF-2 display very strong interactions, with a K_d of 10^{-7} M) (1). Lastly, confirmation of the binder-target interaction was further confirmed via an in vitro GST pull-down assay, wherein GST-C4 was shown to pull down the recombinant Hup2 protein (Fig. 7). Despite our best efforts, we found that the purified GST-C4 protein underwent some amount of degradation, which resulted in the presence of cleaved GST along with GST-C4. However, because of the

FIG. 6. SDS-PAGE showing the expression of *M. tuberculosis* HupB protein and its recombinant derivatives in various expression vectors. Lane 1, Pharmacia low-molecular-mass marker (94, 67, 43, 30, 20, and 14 kDa); lane 2, uninduced *E. coli* TOP10/pGEX4T3-HupB; lane 3, IPTG-induced *E. coli* TOP10/pGEX4T3-HupB (GST-HupB fusion protein expected MW of 49.0 kDa; indicated by a black arrow); lane 4, L-arabinose-induced *E. coli* TOP10/pBAD-HupB (HupB protein expected MW of 23.96 kDa; indicated by a black arrow); lane 5, IPTG-induced *E. coli* TOP10/pMalc2x-Hup2 (MBP-Hup2 fusion protein expected MW of 55.06 kDa; indicated by a black arrow); lane 6, L-arabinose-induced *E. coli* TOP10/pBAD-Hup2r (recombinant Hup2 protein expected MW of 13.23 kDa; indicated by a black arrow); lane 7, IPTG-induced *E. coli* TOP10/pThiohisann-Hup2 (TRX-Hup2 fusion protein expected MW of 24.07 kDa; indicated by a black arrow).

design of the in vitro pull-down assay, the presence of cleaved GST did not affect the outcome of the experiment, as it was shown that purified recombinant Hup2 protein does not bind GST but only GST-C4.

We also found that all proteins bound equally strongly to both HupB and Hup2 (results not shown), thereby indicating that these binders probably target the C-terminal part of HupB. In addition, the obtained binder was also shown to interact strongly with the full-length *M. smegmatis* HupB (re-

sults not shown). The exact regions of binder-target contacts can be determined only from structural studies of C4, preferably when it is bound to its target. Helpfully, C4 is a soluble protein and can be expressed and purified as a GST fusion (Fig. 7). While such experiments would undoubtedly shed light on the nature of C4-HupB interactions (information that would be crucial for designing small-molecule analogs), it was important to first examine whether the obtained de novo binders were able to affect the functioning of HupB itself in mycobacteria.

Effect of protein binder C4 on mycobacterial growth and colony count. To explore the effects of de novo protein binders on mycobacteria, we chose to base our studies on the non-pathogenic *Mycobacterium smegmatis*. The availability of the complete genome sequences of both *M. tuberculosis* and *M. smegmatis* has pointed to a significant homology between the two proteomes (8). For example, *M. smegmatis* HupB is almost 90% identical to the *M. tuberculosis* HupB (Fig. 2A). Further, the *M. smegmatis* HupB has been extensively studied and characterized (25, 46). Consequently, the gene corresponding to one of the protein binders (C4) was cloned in a mycobacterial shuttle vector pSD5 (kind gift of Anil Tyagi). Plasmid pSD5 contains a strong heat shock protein 60 promoter element of mycobacterial origin that helps in the overexpression of foreign and mycobacterial genes in *M. smegmatis* (11). Figure 8A shows the growth curve of *M. smegmatis*/pSD5C4, with *M. smegmatis*/pSD5 and *M. smegmatis*/pSD5GFP strains serving as controls. As can be seen, the reduction in growth is considerable in the exponential phase with it petering out at the stationary-phase level. Interestingly, this is very consistent with the timing of HupB expression in mycobacteria. Previous reports have indicated that HupB is expressed chiefly in the exponential phase of mycobacterial growth (46). While no concrete reason has been attributed to this phenomenon, it has been speculated that the possible role played by HupB in growth regulation is the cause (25, 46). It is tempting to at-

FIG. 7. SDS-PAGE showing the expression and purification of GST-C4 fusion protein. Also shown in the same gel is the in vitro pull-down assay with the recombinant Hup2 protein (lanes 11 and 12). Lane 1, Pharmacia low-molecular-mass marker (94, 67, 43, 30, 20, and 14 kDa); lane 2, uninduced *E. coli* TOP10/C4pGEX4T3; lane 3, IPTG-induced *E. coli* TOP10/C4pGEX4T3 (GST-C4 fusion protein of MW 35.57 kDa; indicated by a black arrow); lane 4, pellet fraction of GST-C4 fusion protein; lane 5, soluble fraction of GST-C4 fusion protein; lane 6, flowthrough; lane 7, wash 1; lane 8, wash 2; lane 9, eluted fraction of GST-C4 fusion protein; lane 10, purified recombinant Hup2 protein (MW of 13.2 kDa; indicated by a dashed arrow); lane 11, GST protein (MW of 27.79 kDa) plus recombinant Hup2 protein (note the absence of a Hup2 band); lane 12, GST-C4 fusion protein plus recombinant Hup2 protein (note the presence of Hup2 band, indicated by a dashed arrow).

FIG. 8. Expression of de novo protein binders in mycobacteria. (A) Growth curve analysis of *M. smegmatis* (Msm) strains with plasmids pSD5, pSD5GFP, and pSD5C4 grown at 37°C. (B) Analysis of CFU obtained during growth of strains mentioned in the description of panel A. Gray bars, *M. smegmatis*/pSD5C4; white bars, *M. smegmatis*/pSD5GFP; white and dotted bars, *M. smegmatis*/pSD5; gray and dotted bars, wild-type *M. smegmatis*. (C) Estimation of CFU as the percent increases of *M. smegmatis*/pSD5 and *M. smegmatis*/pSD5C4 in response to the cold shock stress. White bars, *M. smegmatis*/pSD5; gray bars, *M. smegmatis*/pSD5C4.

tribute the observed reduction in the growth of *M. smegmatis*/pSD5C4 to the in vivo binding of C4 to *M. smegmatis* HupB, especially as the reduction in growth goes on to result in a reduction in *M. smegmatis* colony count (Fig. 8B). This last observation unquestionably relates the expression of C4 with the altered behavior of *M. smegmatis* growth.

Expression of one of the de novo protein binders in *M. smegmatis* was confirmed by visualization of GFP-binder fusion proteins in growing cultures of *M. smegmatis* (results not shown). As the inserts are cloned upstream of the GFP gene in pSD5GFP, the GFP fusion proteins act as an indicator for the viable expression of the de novo proteins in *M. smegmatis* (GFP can only be made after the C series proteins have been successfully translated). Additionally, in order to rule out excision of the wild-type *M. smegmatis* hupB through an inadvertent recombination event, the presence of both the binder gene and the *M. smegmatis* hupB gene was confirmed by gene-specific primers (results not shown).

One of the functions attributed to the HupB protein is that it facilitates the maintenance of mycobacteria under cold shock stress (46). We decided to investigate whether the de novo protein binder, by virtue of its interaction with *M. smegmatis* HupB, is able to reduce the stress tolerance. Exponentially

growing cultures of *M. smegmatis*/pSD5C4 as well as *M. smegmatis*/pSD5 were subjected to the cold shock stress (see Materials and Methods) and their growth monitored over a period of time (Fig. 8C). Herein, we altered the design of our experiment in contrast to an earlier study that simply measured CFU at two different temperatures, the favorable temperature of 37°C and the cold shock temperature of 10°C (46). To ascertain the viability of mycobacteria poststress, we periodically removed cells growing at 10°C and plated them at the favorable temperature (of 37°C), as this would demonstrate the growth viability of cells once they had been subjected to a cold shock (said to mirror the anaerobically induced dormancy commonly witnessed for mycobacteria). As can be seen from the figure, we found a noteworthy decrease in the colony counts of the recombinant *M. smegmatis* strain compared to the wild type, hinting at the possibility that *M. smegmatis* HupB was being directly targeted by the protein binder.

Lastly, we also carried out transmission electron microscopy of recombinant and wild-type *M. smegmatis* in order to investigate whether the expression of protein binders results in differences of a physical nature, e.g., in cell morphology. The results indicate noticeable differences in cell morphology between the wild-type and recombinant strains, with the latter

FIG. 9. TEM pictures of *M. smegmatis*/pSD5 (A and B) and *M. smegmatis*/pSD5C4 (C to F) strains. Arrows depict septum (A, B, D, E, and F), disintegration of cell wall (C), or both (D).

displaying unequal distribution of daughter cells across the septum (Fig. 9). This is indicative of an aberration of some cell division functions (2). Interestingly, TEM analysis of many *M. tuberculosis* and *E. coli* strains with proteins normally associated with cell division, elongation, and septum formation (e.g., FtsZ, PknF, and PknA) knocked out shows similar cell morphologies, thereby further underlining the role of HupB in the regulation of cell division (12, 54).

Effect of protein binder C4 on HupB transcription. In order to further ascertain the effect of a protein binder on mycobacteria, the transcription levels of the *hupB* gene in wild-type *M. smegmatis* (with pSD5) as well as in the recombinant strain (*M. smegmatis*/pSD5C4) were investigated. As can be seen in Fig. 10, there is a significant reduction in the level of *hupB* transcript in cells that express the protein binder (lane 3) compared

with levels when the binder is not present (lane 4). The transcriptional profile experiment provides further evidence that the binder targets the HupB protein. It is noteworthy to mention that HupB belongs to a family of proteins that bind DNA

FIG. 10. Transcription profile of *hupB* in wild-type and recombinant strains. Lane 1, *M. smegmatis*/pSD5C4 total RNA (20 μ g); lane 2, *M. smegmatis*/pSD5 total RNA (20 μ g); lane 3, *hupB* transcript of *M. smegmatis*/pSD5C4, exponential phase; lane 4, *hupB* transcript of *M. smegmatis*/pSD5, exponential phase.

and influence cell regulation, and many of these act as positive autoregulators of their own transcription (20, 50). Under the circumstances in which such proteins are inactivated by other proteins or small-molecule drugs, the transcription levels of their corresponding genes are reduced (17, 18, 50). In the present scenario, reduction in the transcription level of *M. smegmatis* *hupB* in cells in which the protein binder C4 is being expressed provides indirect confirmation for an in vivo binder-HupB interaction.

In summary, our observations indicate that the expression of the de novo protein binder in *M. smegmatis* causes (i) a reduction of growth of the recombinant strain in the exponential phase, (ii) a decrease in the transcription levels of the *hupB* gene, (iii) a decrease in the overall colony count, (iv) a reduction in the capacity of the recombinant strain to withstand cold shock stress, and (v) changes in cell morphology.

CONCLUSIONS

In this report, we have described the isolation and selection of de novo protein binders against an essential protein of *Mycobacterium tuberculosis*. Our study employed the novel codon-shuffling method for creating de novo protein libraries. A proof-of-principle experiment that involved bringing together two de novo protein libraries carrying opposite charges brought forth the advantages of the codon-shuffling method in unearthing strongly interacting proteins of various lengths and secondary structures. Next, based on target inspection, we programmed the method to generate skewed libraries that were predetermined for a charge-based interaction. The target, HupB, was selected on the basis of its essentiality to the growth and persistence of *M. tuberculosis*. The obtained protein binders, by virtue of their strong binding to the target, were also shown to adversely affect the regular functioning of HupB in *M. smegmatis*.

These findings, we believe, have laid the groundwork for attacking other essential *M. tuberculosis* targets mentioned in the many lists described by Rubin and coworkers. This approach would entail an inspection of target protein properties (pI, MW, and other secondary structure characteristics), followed by the programming of de novo libraries to have properties that are complementary to those of the target protein. Imminently, the 200-odd culture-filtrate proteins of *M. tuberculosis* appear as an attractive pool for target selection. This is because the generated protein binders would bind to the target extracellularly, as opposed to them having to penetrate the *M. tuberculosis* cell wall to carry out their action (as would have to happen in the case of HupB, a cytoplasmic protein), although recent developments point to some peptide tags being capable of transporting our protein binders to their eventual intracellular destination (14). On the other hand, one may envisage a scenario where a particular target gene is not picked, rather a library of in-frame *M. tuberculosis* gene fragments is hit against variably programmed codon-shuffled protein libraries, resulting not only in the isolation of protein binders but also in scores of targets with which the former bind strongly. In other words, the binders as well as the targets would be simultaneously selected; follow-up experiments could then be performed on particular targets from this pool, based on their

importance or essentiality. Such experiments are currently under way in our laboratory.

The absence of a reliable vaccine and new frontline drugs, coupled with increasing prevalence of multiple-drug-resistant *M. tuberculosis* strains, demands urgent attention. Rational, part-rational, or nonrational approaches towards the design of de novo proteins that target *M. tuberculosis* proteins (one at a time or indiscriminately) can, we believe, earnestly augment the conventional natural/semisynthetic small-molecule-based antituberculosis drug discovery efforts, even though matters of protein delivery, creation of peptidomimetics, and structure-based small-molecule design (all issues that would follow the initial protein/peptide leads) still remain largely at an exploratory stage (30). It is hoped that buildup of a collection of protein binders that show potent in vivo inhibitory activity would encourage further research in the areas mentioned above.

ACKNOWLEDGMENTS

We thank David McMurray, Texas A&M University College of Medicine, for critically reading the manuscript. We also thank Anil Tyagi for the kind gift of plasmid pSD5, Dinkar Sahal for his help in generating CD data, and Sandeep Mehta for analysis of TEM data.

This work was funded by internal grants of ICGEB and IGIB. A. Rao, G. Ram, and K. Kumar thank CSIR, India, for financial support.

REFERENCES

1. Barberis, A., J. Pearlberg, N. Simkovich, S. Farrell, P. Reinagel, C. Bamdad, G. Sigal, and M. Ptashne. 1995. Contact with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* **81**:359–368.
2. Begg, K. J., and W. D. Donachie. 1985. Cell shape and division in *Escherichia coli*: experiments with shape and division mutants. *J. Bacteriol.* **163**:615–622.
3. Blum, J. H., S. L. Dove, A. Hochschild, and J. J. Mekalanos. 2000. Isolation of peptide aptamers that inhibit intracellular processes. *Proc. Natl. Acad. Sci. USA* **97**:2241–2246.
4. Butz, K., C. Denk, B. Fitcher, I. Crnkovic-Mertens, A. Ullmann, C. H. Schroder, and F. Hoppe-Seyler. 2001. Peptide aptamers targeting the hepatitis B virus core protein: a new class of molecules with antiviral activity. *Oncogene* **20**:6579–6586.
5. Chattopadhyay, A., S. A. Tate, R. W. Beswick, S. D. Wagner, and P. K. Ferrigno. 2005. A peptide aptamer to antagonize BCL-6 function. *Oncogene* **25**:2223–2233.
6. Chopra, S., and A. Ranganathan. 2003. Protein evolution by “codon shuffling”: a novel method for generating highly variant mutant libraries by assembly of hexamer DNA duplexes. *Chem. Biol.* **10**:917–926.
7. Colas, P., B. Cohen, T. Jessen, I. Grishina, J. McCoy, and R. Brent. 1996. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**:548–550.
8. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.
9. Corey, M. J., and E. Corey. 1996. On the failure of de novo-designed peptides as biocatalysts. *Proc. Natl. Acad. Sci. USA* **93**:11428–11434.
10. Dahiyat, B. I., and S. L. Mayo. 1997. De novo protein design: fully automated sequence selection. *Science* **278**:82–87.
11. DasGupta, S. K., S. Jain, D. Kaushal, and A. K. Tyagi. 1998. Expression systems for study of mycobacterial gene regulation and development of recombinant BCG vaccines. *Biochem. Biophys. Res. Commun.* **246**:797–804.
12. Deol, P., R. Vohra, A. K. Saini, A. Singh, H. Chandra, P. Chopra, T. K. Das, A. K. Tyagi, and Y. Singh. 2005. Role of *Mycobacterium tuberculosis* Ser/Thr kinase PknF: implications in glucose transport and cell division. *J. Bacteriol.* **187**:3415–3420.
13. Dove, S. L., J. K. Joung, and A. Hochschild. 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386**:627–630.
14. Eiden, L. E. 2005. Fusion polypeptides that inhibit exocytosis: fusing aptamer and cell-penetrating peptide technologies and pharmacologies. *Mol. Pharmacol.* **67**:980–982.

15. Gazzard, B. 2001. Tuberculosis, HIV and the developing world. *Clin. Med.* **1**:62–68.
16. Glickman, S. W., E. B. Rasiel, C. D. Hamilton, A. Kubataev, and K. A. Schulman. 2006. A portfolio model of drug development for tuberculosis. *Science* **311**:1246–1247.
17. Grass, J. A., M. E. Boyer, S. Pal, J. Wu, M. J. Weiss, and E. H. Bresnick. 2003. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc. Natl. Acad. Sci. USA* **100**:8811–8816.
18. Hashimoto, Y., A. Q. Khan, and T. Ezaki. 1996. Positive autoregulation of *vipR* expression in *ViaB* region-encoded Vi antigen of *Salmonella typhi*. *J. Bacteriol.* **178**:1430–1436.
19. Hatfull, F., and W. R. Jacobs, Jr. 2000. Molecular genetics of mycobacteria. American Society for Microbiology, Washington, DC.
20. Haydel, S. E., W. H. Benjamin, N. E. Dunlap, and J. E. Clark-Curtiss. 2002. Expression, autoregulation, and DNA binding properties of the *Mycobacterium tuberculosis* TrcR response regulator. *J. Bacteriol.* **184**:2192–2203.
21. Kamtekar, S., J. M. Schiffer, H. Xiong, J. M. Babik, and M. H. Hecht. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**:1680–1685.
22. Keefe, A. D., and J. W. Szostak. 2001. Functional proteins from a random-sequence library. *Nature* **410**:715–718.
23. Kiel, C., T. Selzer, Y. Shaul, G. Schreiber, and C. Herrmann. 2004. Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. *Proc. Natl. Acad. Sci. USA* **101**:9223–9228.
24. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
25. Lee, B. H., B. Murugasu-Oei, and T. Dick. 1998. Upregulation of a histone-like protein in dormant *Mycobacterium smegmatis*. *Mol. Gen. Genet.* **260**:475–479.
26. Marques, M. A., S. Mahapatra, E. N. Sarno, S. Santos, J. S. Spencer, P. J. Brennan, and M. C. Pessolani. 2001. Further biochemical characterization of *Mycobacterium leprae* laminin-binding proteins. *Braz. J. Med. Biol. Res.* **34**:463–470.
27. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28. Miller, L. P., J. T. Crawford, and T. M. Shinnick. 1994. The *rpoB* gene of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **38**:805–811.
29. Moffet, D. A., and M. H. Hecht. 2001. De novo proteins from combinatorial libraries. *Chem. Rev.* **101**:3191–3203.
30. Ohlstein, E. H., R. R. Ruffolo, and J. D. Elliott. 2000. Drug discovery in the next millennium. *Annu. Rev. Pharmacol. Toxicol.* **40**:177–191.
31. Onyebujoh, P., A. Zumla, I. Ribeiro, R. Rustomjee, P. Mwaba, M. Gomes, and J. M. Grange. 2005. Treatment of tuberculosis: present status and future prospects. *Bull. W. H. O.* **83**:857–865.
32. Panyim, S., and R. Chalkley. 1971. The molecular weights of vertebrate histones exploiting a modified sodium dodecyl sulfate electrophoretic method. *J. Biol. Chem.* **246**:7557–7560.
33. Prabhakar, S., P. S. Annapurna, N. K. Jain, A. B. Dey, J. S. Tyagi, and H. K. Prasad. 1998. Identification of an immunogenic histone-like protein (HLPMT) of *Mycobacterium tuberculosis*. *Tuber. Lung Dis.* **79**:43–53.
34. Prijambada, I. D., T. Yomo, F. Tanaka, T. Kawama, K. Yamamoto, A. Hasegawa, Y. Shima, S. Negoro, and I. Urabe. 1996. Solubility of artificial proteins with random sequences. *FEBS Lett.* **382**:21–25.
35. Rao, A., S. Chopra, G. Ram, A. Gupta, and A. Ranganathan. 2005. Application of the “codon-shuffling” method. Synthesis and selection of *de novo* proteins as antibacterials. *J. Biol. Chem.* **280**:23605–23614.
36. Rao, A., and A. Ranganathan. 2004. Interaction studies on proteins encoded by the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. *Mol. Genet. Genomics* **272**:571–579.
37. Regan, L., and W. F. DeGrado. 1988. Characterization of a helical protein designed from first principles. *Science* **241**:976–978.
38. Rengarajan, J., B. R. Bloom, and E. J. Rubin. 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. USA* **102**:8327–8332.
39. Roberts, R. W. 1999. Totally *in vitro* protein selection using mRNA-protein fusions and ribosome display. *Curr. Opin. Chem. Biol.* **3**:268–273.
40. Rodi, D. J., and L. Makowski. 1999. Phage-display technology-finding a needle in a vast molecular haystack. *Curr. Opin. Biotechnol.* **10**:87–93.
41. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
42. Sassetti, C. M., D. H. Boyd, and E. J. Rubin. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl. Acad. Sci. USA* **98**:12712–12717.
43. Sassetti, C. M., D. H. Boyd, and E. J. Rubin. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48**:77–84.
44. Sassetti, C. M., and E. J. Rubin. 2003. Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. USA* **100**:12989–12994.
45. Schreiber, G., and A. R. Fersht. 1996. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* **3**:427–431.
46. Shires, K., and L. Steyn. 2001. The cold-shock stress response in *Mycobacterium smegmatis* induces the expression of a histone-like protein. *Mol. Microbiol.* **39**:994–1009.
47. Sinha, N., and S. J. Smith-Gill. 2002. Electrostatics in protein binding and function. *Curr. Protein Pept. Sci.* **3**:601–614.
48. Soares de Lima, C., L. Zulianello, M. A. Marques, H. Kim, M. I. Portugal, S. L. Antunes, F. D. Menozzi, T. H. Ottenhoff, P. J. Brennan, and M. C. Pessolani. 2005. Mapping the laminin-binding and adhesive domain of the cell surface-associated Hlp/LBP protein from *Mycobacterium leprae*. *Microbes. Infect.* **7**:1097–1109.
49. Taylor, S. V., K. U. Walter, P. Kast, and D. Hilvert. 2001. Searching sequence space for protein catalysts. *Proc. Natl. Acad. Sci. USA* **98**:10596–10601.
50. Turgay, K., J. Hahn, J. Burghoorn, and D. Dubnau. 1998. Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J.* **17**:6730–6738.
51. Wade, R. C., R. R. Gabdouliline, S. K. Ludemann, and V. Lounnas. 1998. Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. *Proc. Natl. Acad. Sci. USA* **95**:5942–5949.
52. Williams, B. G., R. Granich, L. S. Chauhan, N. S. Dharmshaktu, and C. Dye. 2005. The impact of HIV/AIDS on the control of tuberculosis in India. *Proc. Natl. Acad. Sci. USA* **102**:9619–9624.
53. Woodman, R., J. T. Yeh, S. Laurenson, and P. Ko-Ferrigno. 2005. Design and validation of a neutral protein scaffold for the presentation of peptide aptamers. *J. Mol. Biol.* **352**:1118–1133.
54. Zhang, C. C. 1993. A gene encoding a protein related to eukaryotic protein kinases from the filamentous heterocystous cyanobacterium *Anabaena PCC7120*. *Proc. Natl. Acad. Sci. USA* **90**:11840–11844.