SHORT COMMUNICATION

Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins

Valeria A. Risso,[1]* Jose A. Gavira,[2] Eric A. Gaucher,[3,4] and Jose M. Sanchez-Ruiz[1]*

[1] Departamento de Quimica Fisica, Facultad de Ciencias, Universidad de Granada, 18071, Spain

[2] Laboratorio de Estudios Cristalográficos, Instituto Andaluz de Ciencias de la Tierra (Consejo Superior de Investigaciones Científicas–Universidad de Granada), Avenida de las Palmeras 4, Armilla, Granada 18100, Spain

[3] School of Biology, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332

[4] Parker H. Petit Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30332

ABSTRACT

Consensus-sequence engineering has generated protein variants with enhanced stability, and sometimes, with modulated biological function. Consensus mutations are often interpreted as the introduction of ancestral amino acid residues. However, the precise relationship between consensus engineering and ancestral protein resurrection is not fully understood. Here, we report the properties of proteins encoded by consensus sequences derived from a multiple sequence alignment of extant, class A β-lactamases, as compared with the properties of ancient Precambrian β-lactamases resurrected in the laboratory. These comparisons considered primary sequence, secondary, and tertiary structure, as well as stability and catalysis against different antibiotics. Out of the three consensus variants generated, one could not be expressed and purified (likely due to misfolding and/or low stability) and only one displayed substantial stability having substrate promiscuity, although to a lower extent than ancient β-lactamases. These results: (i) highlight the phenotypic differences between consensus variants and laboratory resurrections of ancestral proteins; (ii) question interpretations of consensus proteins as phenotypic proxies of ancestral proteins; and (iii) support the notion that ancient proteins provide a robust approach toward the preparation of protein variants having large numbers of mutational changes while possessing unique biomolecular properties.

© 2014 Wiley Periodicals, Inc.

INTRODUCTION

Fifty years have passed since Pauling and Zuckerkandl[1] noted that the availability of sequence information contained in several extant (living) organisms could be used to derive a reasonable statistical estimate of the ancient sequence representing the common ancestor of the modern sequences of a given protein. Starting with advances in bioinformatics and molecular biology in the 1990's, together with the genomic-era availability of increasing

numbers of extant protein sequences, researches are now able to confidently generate laboratory proteins encoded by reconstructed ancestral sequences.[2–4] In recent years, some of these "molecular restoration studies" (to use the terminology originally employed by Pauling and Zuckerkand[l]) have been reported to successfully infer proteins on the order of billions of years ago, despite the very large number of substitutions involved.[5–8] Furthermore, laboratory resurrections of Precambrian proteins appear to consistently display interesting properties in terms of both stability and function. Large denaturation temperature enhancements of 30–35°C upon "traveling back in time" 2–4 billion years have been reported for elongation factors, thioredoxins and β-lactamases.[5–7] Also, laboratory resurrections of ~4 billion-year-old thioredoxins are highly active at acidic pH[6] while laboratory resurrections of 2–3 billion years old β-lactamases have been found to be moderately efficient promiscuous enzymes capable of hydrolyzing various antibiotics *in vitro* and able to endow modern microorganisms with significant levels of resistance toward the various antibiotics.[7] Certainly, some of these properties are informative about the environment surrounding ancient life. For instance, the large protein stability enhancements support the notion that Precambrian life was thermophilic,[6,7] while the activity/pH profiles for ~4 billion-year-old thioredoxins can be interpreted as evidence[6] of the acidic character of the primitive oceans (presumably linked to high levels of CO_2 in the primitive atmosphere). From an applied perspective, the unique and extreme properties reported for ancient proteins support the biotechnological potential when enhanced stability and promiscuity are highly desirable features in protein scaffolds to be used for enzyme engineering protocols.[7]

Generating a multiple sequence alignment (MSA) of extant sequences is the first step of any study attempting to infer ancient sequences or calculate a consensus sequence. Due to sequence conservation, it is reasonable to expect a substantial number of positions to have identical residues when comparing ancestral sequences with consensus sequences, but how these residues contribute to function is not always clear. Consensus mutations have been studied in many protein systems over the years and these mutations are sometimes found to be stabilizing[9–16] and, in some cases, capable of modulating biomolecular function.[17] Thus, the possibility arises that consensus mutations are responsible for the unique and extreme properties of laboratory resurrections of Precambrian proteins and that consensus variants are, in fact, good approximations of ancestral proteins. In practice, however, the precise relationship between consensus and ancestral proteins is not fully understood. While it is possible that consensus mutations may steer a protein sequence closer to that of its inferred ancestor,[18] our experimental studies have found significant phenotypic differences between consensus and ancestral

proteins.[5,19,20] In this study, we take advantage of our recent experimental studies on the laboratory resurrections of Precambrian β-lactamases to explore these issues in detail. Briefly (see paragraph below for details), we prepared variants encoded by the consensus of extant sequences under specific nodes in the phylogenetic tree of class A β-lactamases and we compare these consensus proteins to their inferred ancestral counterparts in terms of sequence, structure, stability and catalysis.

Starting with a MSA of 75 chromosomal sequences of extant class A β-lactamases, we recently performed an ancestral reconstruction analysis targeting several billion-year-old Precambrian nodes in the evolution of these proteins.[7] Specifically (see Fig. 1), we targeted the common ancestors of enterobacteria (ENCA), Gammproteobacteria (GPBCA), various Gram-negative bacteria (GNCA) and various Gram-positive and Gram-negative bacteria (PNCA). The encoded proteins were prepared in the laboratory and, as discussed above, those corresponding to 2–3 billion-year-old nodes were found to be highly stable and enzymatically promiscuous against various antibiotic substrates. We further determined the 3D-structure for the ancient proteins corresponding to the GNCA and ENCA nodes and found that they fold into the canonical lactamase fold despite extensive sequence differences compared with the extant lactamases. In this work, we extend these experimental studies to include protein variants encoded by several consensus sequences. Specifically (see Fig. 1), we analyze protein variants encoded by: (i) the consensus sequence of the 14 lactamases of the MSA that present the descendent modern sequences of the GPBCA node in the phylogenetic tree used; (ii) the consensus sequence of the 26 modern lactamases descended from the GNCA node in the tree; (iii) the consensus sequence of all 75 lactamases included in the MSA. Following our previous nomenclature, we refer to these proteins as GPB-CON, GN-CON, and ALL-CON, respectively. Note that both the reconstructed ancestral sequences and the consensus sequences are based on the same MSA, so that comparisons between them are appropriate.

MATERIALS AND METHODS

Consensus sequences were obtained from the MSA of 75 sequences of chromosomal β-lactamases we previously used for ancestral sequence reconstruction analysis (Supporting Information Fig. S1 in Risso *et al.*[7]) and from the two subsets of the MSA described above. As previously described (see Supporting Information in Risso *et al.*[7]), insertions occurring in a few sequences were deleted and the sequences were reduced to a common length of 262 amino acids by removing a few external residues in the highly flexible C-terminal domain. Consensus sequences (given in Fig. S1 of Supporting Information) were obtained with Bioedit (www.mbio.ncsu.edu) using a threshold frequency of 0%, which guarantees
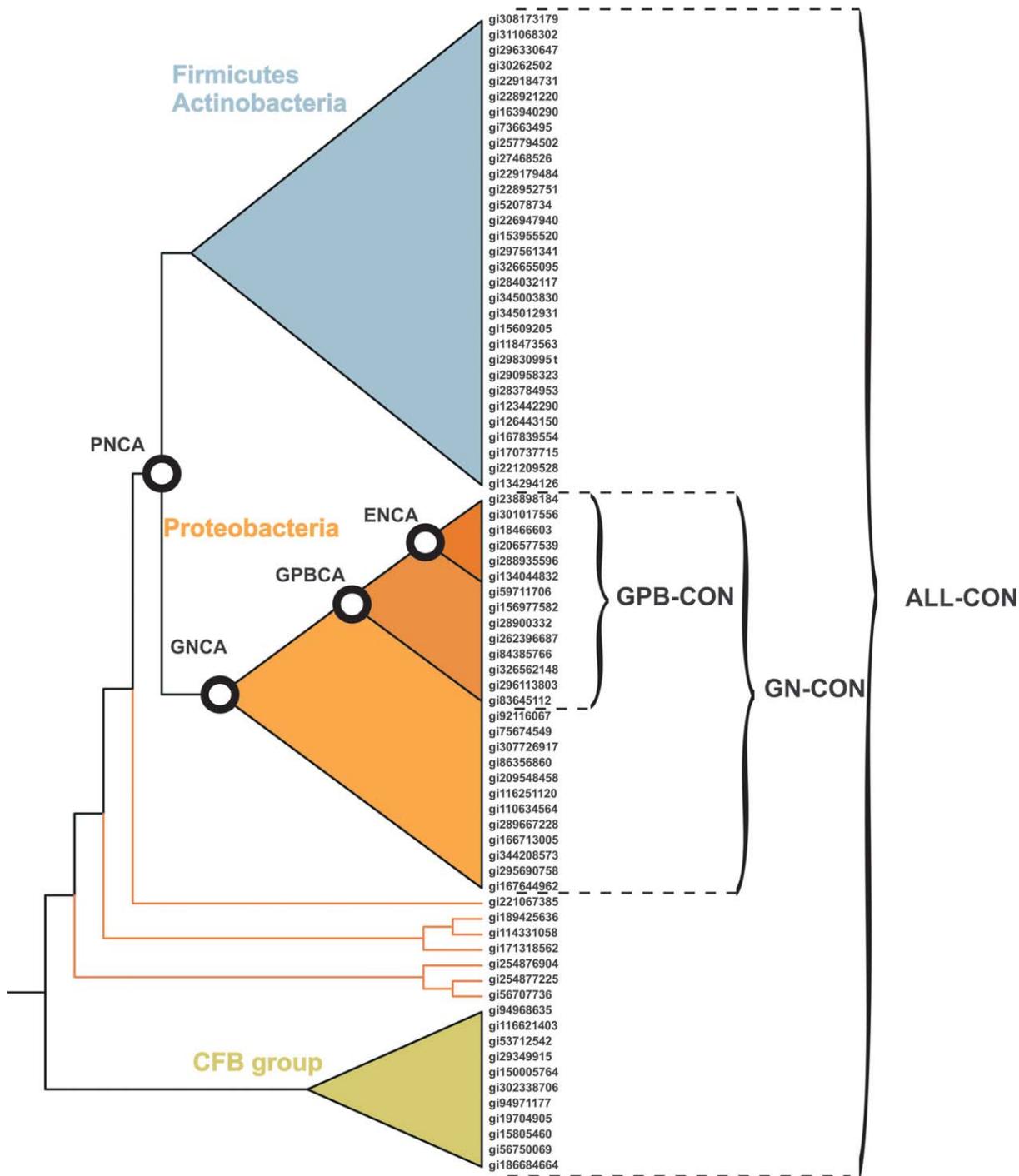
Figure 1

Genbank Identifiers for 75 extant sequences of class A β-lactamases and schematic representation of the phylogenetic tree derived from Bayesian analysis (see Risso *et al.*[7] for a detailed representation of the tree). The Precambrian nodes targeted in our previous study (Risso *et al.*[7]) on laboratory resurrection of Precambrian β-lactamases are labeled: ENCA (common ancestor of enterobacteria), GPBCA (common ancestor of Gammaproteobacteria), GNCA (common ancestor of various Gram-negative bacteria), and PNCA (common ancestor of various Gram-positive and Gram-negative bacteria). Curly brackets are used to indicate the sequence subsets used for the calculation of the consensus sequences GPB-CON, GN-CON, and ALL-CON. A sequence alignment including the reconstructed ancestral sequences for the ENCA, GPBCA, GNCA, and PNCA nodes, the consensus sequences GPB-CON, GN-CON, and ALL-CON, as well as the sequence of the extant TEM-1 β-lactamase, is given in Figure S1 of Supporting Information.

Table I
Sequence Differences Between the β-Lactamase Studied in This Work

	GPB-CON	GN-CON	ALL-CON
TEM-1	67/262	93/262	121/262
ENCA	42/262	67/262	106/262
	29/236	**50/236**	**86/236**
GPBCA	77/262	39/262	68/262
	21/195	**11/195**	**32/195**
GNCA	98/262	53/262	61/262
	72/223	**34/223**	**35/223**
PNCA	103/262	62/262	54/262
	71/223	**33/223**	**35/223**

Numbers in bold refer to robustly inferred positions in ancestral sequences (probability of the most probabilistic residue is ≥0.8 at these positions). First number in each pair refers to amino acids differences while the second number is the total number of residues.

that all positions in the consensus sequence are assigned the highest frequency amino acid.

Genes encoding the consensus lactamases were synthesized and codon optimized for expression in *E. coli* cells. Genes were cloned into a pET24 vector with kanamycin resistance (Novagen), transformed in *E. coli* BL21 (DE3) cells (Stratagene, CA) and the proteins were purified by osmotic shock and gel filtration chromatography as previously described (Risso *et al.*[7]). Protein purity was checked by SDS-PAGE. Lactamase activity in cellular extracts after induction was assessed by following spectrophotometrically (486 nm) the degradation of nitrocefin.[21] Enzyme kinetic parameters (K_m, k_{cat} and k_{cat}/K_m) for the lactamase-catalyzed hydrolysis of benzylpenicillin (BZ), cefotaxime (CTX), and ceftazidime (CAZ) at 25°C in 50 mM sodium phosphate buffer pH 7.0 were determined from the fitting of the Michaelis-Menten equation to profiles of initial rate versus substrate concentration (see Fig. S2 and Table S2 in Supporting Information). These profiles were obtained from the change in UV absorbance upon antibiotic hydrolysis as previously described (Risso *et al.*[7]). Thermal stability of the consensus lactamases was measured in 25 mM Hepes buffer pH 7.0 using a VP-Capillary differential scanning calorimeter [Microcal (GE Healthcare) Northhampton] and following the protocol described in detail (Risso *et al.*[7]).

Crystals of the consensus lactamases and the ancient lactamase corresponding to the PNCA node were grown using the counter-diffusion technique[22] in capillaries of 0.1 or 0.2 mm inner diameter. Good-quality diffracting crystals were obtained in condition C18 (25% PEG 4K, 0.2M NH$_4$ sulfate, 0.1M Na-acetate pH 4.60) of the GCB-CSK and box 4 (20% PEG 400, 15% PEG 4K, 10% PEG 8K, 0.1M Tris-HCl pH 7.0) of the PEG448-49 crystallization kits (Triana S&T, Granada, Spain) for ALL-CON and PNCA, respectively. In both cases, structures were determined by the molecular-replacement method using the GNCA structure (PDB ID 4B88) as search model from data collected at Soleil (beam-line Proxima-1) and ESRF

(beam-line BM14) synchrotron sources. Details on the crystal treatment, data collection, and statistics are summarized in Table S1 in Supporting Information.

RESULTS AND DISCUSSION

Comparison between consensus sequences and reconstructed ancestral sequences

Figure 1 shows the Genbank Identifiers of the 75 extant β-lactamase sequences we used in our recent sequence reconstruction analysis of Precambrian lactamases (Risso *et al.*[7]). To avoid complications due to recent evolution in the clinical antibiotic era, all sequences were chromosomally located (derived from a search in the NCBI database of complete genomes) thus no plasmidic lactamases were included, and we also made sure that none of the selected sequences came from clinical isolates (see Risso *et al.*[7] for details). Figure 1 also shows a schematic representation of the phylogenetic tree inferred from the 75 sequences using Bayesian analysis,[7] the Precambrian nodes originally resurrected[7] and the sets of sequences used to derive the three consensus proteins targeted in this work: the consensus of the 75 extant lactamases calculated from the complete MSA used for ancestral sequence reconstruction (ALL-CON), the consensus of the 26 extant sequences descended from the GNCA node (GN-CON) and the consensus of the 14 extant sequences descended from the GPBCA node (GPB-CON).

As expected, there is a substantial amount of sequence identity between consensus and ancestral sequences as seen from the alignment shown in Figure S1 of Supporting Information. However, site-specific residue differences between the consensus and ancestral sequences are notable (see the data collected in Table I). In particular, many residues differ between sites in which the ancestrally inferred residues are supported with high posterior probabilities (differences shown in bold in Table I are tabulated by taking into account only those positions in which the reconstructed ancestral residues have a posterior probability >0.8).

On the differences between consensus sequences and ancestral reconstructed sequences

Differences between consensus sequences and ancestral reconstructions should not come as a surprise, even if both analyses start with the same extant sequence information, as is the case here. It is reasonable to expect such differences because the two analyses handle extant sequence information in very different manners. Consensus is a simple amino acid counting procedure and the consensus sequence itself reflects the most counts at individual sites across a set of extant sequences. By contrast, ancestral reconstruction uses all extant sequence

information to construct a phylogenetic tree in which the branches have different lengths and, subsequently (see pages 45–46 in Liberles[3]), produces probabilities for different sets of ancestral states at all nodes using amino acid replacement probabilities that take branch lengths into account. In one instance, the use of joint probabilities implies that the most probabilistic ancestral states at a given node are linked to the ancestral states determined for all other nodes on tree including not only "younger descendant" nodes below, but also "older ancestor" nodes above. Furthermore, the lengths of the branches connecting the nodes are relevant as they are used to calculate amino acid replacement probabilities. In fact, most discrepancies between consensus residues and the inferred ancestral residues can be intuitively understood on the basis of these ideas, as it is illustrated by the two representative examples in figures S3 and S4 of the Supporting Information. Note that the nodes of interest in the trees of Supporting Information Figures S3 and S4 are typically connected to other nodes with identical ancestral states via short branches.

Differences between the consensus and ancestral approaches are likely to occur when a site contains a high amount of sequence diversity. In fact, a simple statistical analysis (Fig. S5 in the Supporting Information) shows that the consensus/ancestral discrepancies occur more often at positions with high Shannon entropy based on the MSA. Furthermore, since all the extant sequences are assumed to have evolved from a single ancestral sequence (or from a population of closely related sequences), high diversity should correspond to high evolutionary rates. Indeed, Shannon entropy values show a good correlation with the site-specific evolutionary rates calculated using an among-site rate variation model of the gamma distribution (Supporting Information Fig. S6). Thus, we expect that consensus/ancestral discrepancies are also likely to occur at sites having high rates of evolution for the lactamase family as well as other protein families (Fig. S7 in the Supporting Information).

The results presented above highlight that a certain number of differences between a consensus sequence and a corresponding reconstructed ancestral sequence (i.e., derived from the same extant sequence alignment) are to be expected, particularly at sites with high rates of evolution. A potentially more important issue, however, is whether such differences have a significant phenotypic impact or not. It is important to note at this point that ancestral resurrection studies do not claim that reconstructed sequences are precise representations of true and unique ancestral sequences but, rather, that the properties of the proteins encoded by the reconstructed sequences (the laboratory resurrected proteins) provide an acceptable approximation of the ancestral protein phenotype. This claim is reasonable when (i) similar phenotypic properties are displayed for proteins encoded by alternative reconstructions at a single given node (phenotypic robustness)

and (ii) the properties of the resurrected proteins lead to credible evolutionary narratives that are further supported by disparate scientific studies. In the case of our previous work on resurrected Precambrian β-lactamases,[7] phenotypic robustness was displayed based on measurements of additional sequences derived from a Monte Carlo sampling of the posterior probability distributions. Furthermore, the stability and catalytic features of the resurrected proteins led to a remarkably convincing evolutionary narrative. Laboratory resurrections of 2–3 billion-years-old β-lactamases were thus found to be highly stable and catalytically promiscuous, supporting the often proposed thermophilic character of ancestral life[5–7,20] and providing evidence for the evolutionary conversion of generalists (promiscuous enzymes) into specialists over the course of natural evolution. Evolutionary conversion of substrate-promiscuous β-lactamases into penicillin-specialists may be an adaptation to the development of penicillin antibiotics by other organisms and, interestingly, our experimental data for several laboratory-resurrected Precambrian β-lactamases suggested that the conversion occurred over a time span roughly matching the divergence time of fungi (about 1.2 billion years ago). It should be clear, in view of all this, that the crucial issue in the context of this work is whether the consensus/ancestral discrepancies have a phenotypic impact and, in particular, whether the properties of the proteins encoded by the consensus sequences lead to similarly convincing evolutionary narratives provided by the proteins obtained on the basis of ancestral sequence reconstruction. To experimentally address this issue, we have targeted the preparation in the laboratory of the proteins encoded by the ALL-CON, GN-CON, and GPB-CON consensus sequences and their characterization in terms of 3D-structure, stability, and catalysis. These efforts are described below.

Consensus variant purification

The consensus variants GPB-CON and ALL-CON could be easily purified to high yield using standard procedures (see Materials and Methods for details). However, all our attempts to obtain GN-CON were unsuccessful (no protein detected in SDS-PAGE of total extracts: Fig. 2) despite trying a variety of overexpression conditions using the standard strain BL21 (DE3). All combinations of various induction times (3 h, 5 h, and overnight), various induction temperatures (4, 16, 20, 30 and 37°C) and various culture media (LB and TB) were tested. We also tried the overexpression of GN-CON in the BL21 (DE3)pLysS strain that is recommended for overexpression of toxic proteins; still, no expression was detected in SDS-PAGE. Measurements of lactamase activity in cellular lysates obtained across different times after protein induction revealed a large increase of lactamase activity correlated with time for the ALL-CON and GPB-CON variants but only a negligible increase for the GN-CON variant
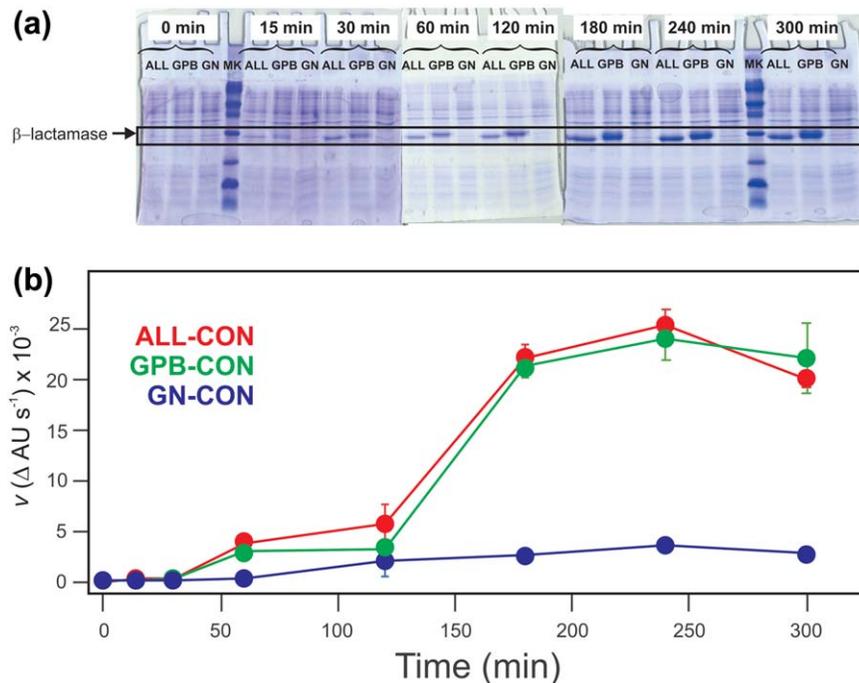
Figure 2

The amounts of consensus lactamases after induction in BL21 cells. (**a**) SDS-PAGE of total cellular extracts at different times after induction. GPB, GN, and ALL refer to the consensus variants GPB-CON, GN-CON, and ALL-CON. Lanes with the molecular weight markers are labeled MK. (**b**) lactamase activity (as measured by the rate of hydrolysis of nitrocefine) in the soluble fraction versus time after induction. Note that in both the SDS-PAGE and activity assays, very low levels of the consensus GN-CON lactamase are detected. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

(Fig. 2), a pattern that is consistent with the corresponding amount of protein as seen in SDS-PAGE. This suggests that the protein encoded by the GN-CON consensus sequence fails to substantially fold properly and/or it is highly unstable and becomes degraded quickly in the cell.

significant differences are observed in the active site region, suggesting that differences in catalytic features between these enzymes (see further below) could be related to dynamic factors that may not be apparent in static X-ray structures, as we have previously discussed.[7]
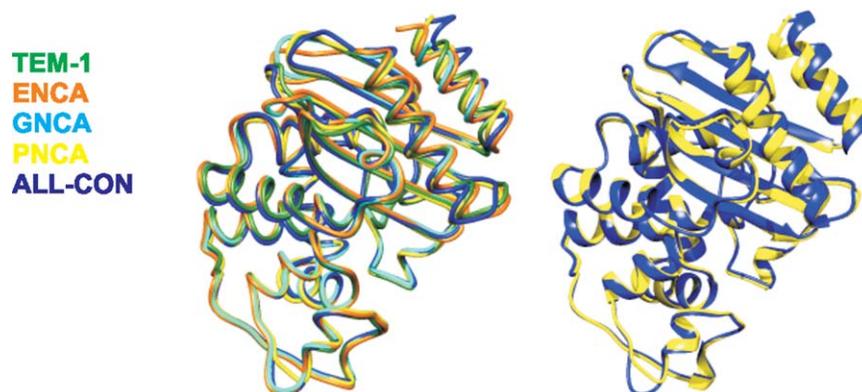
Structure

Diffraction quality crystals could not be generated for the GPB-CON variant (nor could crystals corresponding to the GPBCA ancestral protein from our previous work: Risso *et al.*[7]) despite trying both counter diffusion and vapour diffusion experiments with several crystallization screening kits. On the other hand, crystallization and 3D-structure determination to 1.8 and 2.2 Å resolution limit could be carried out for the PNCA and ALL-CON variants respectively. Figure 3 shows a comparison of the structure determined for the consensus ALL-CON variant along with the structures of an extant TEM-1 β-lactamase and three resurrected Precambrian proteins: GNCA and ENCA β-lactamases (structures reported in Risso *et al.*[7]), as well as the lactamase corresponding to the PNCA node (determined as part of this work). It is evident from Figure 3 that the proteins all share the canonical lactamase fold and only minor movements appear in both the terminal helix and some solvent-exposed loops. In fact, no

Stability

The consensus GPB-CON variant is generated from the subset of sequences in the MSA used by Risso *et al.*[7] that are descended from the node assigned to the last common ancestor of gamma-Proteobateria [GPBCA node in Fig. 1(a)]. It is reasonable, therefore, to compare the properties of this consensus protein with those of the resurrected lactamase corresponding to the ancestral node, as well as to extant proteins within this subset of sequences. In terms of stability [see DSC profiles in Fig. 4(a), GPB-CON lactamase is only slightly more stable than the extant TEM-1 lactamase from *E. coli* (denaturation temperature of 60°C versus 55°C, respectively]. Conversely, the GPB-CON protein has a denaturation temperature about 30° below that of the GPBCA lactamase.

The variant encoded by the consensus of all sequences in the MSA used by Risso *et al.*[7] (ALL-CON) has a denaturation temperature of 79°C. This demonstrates
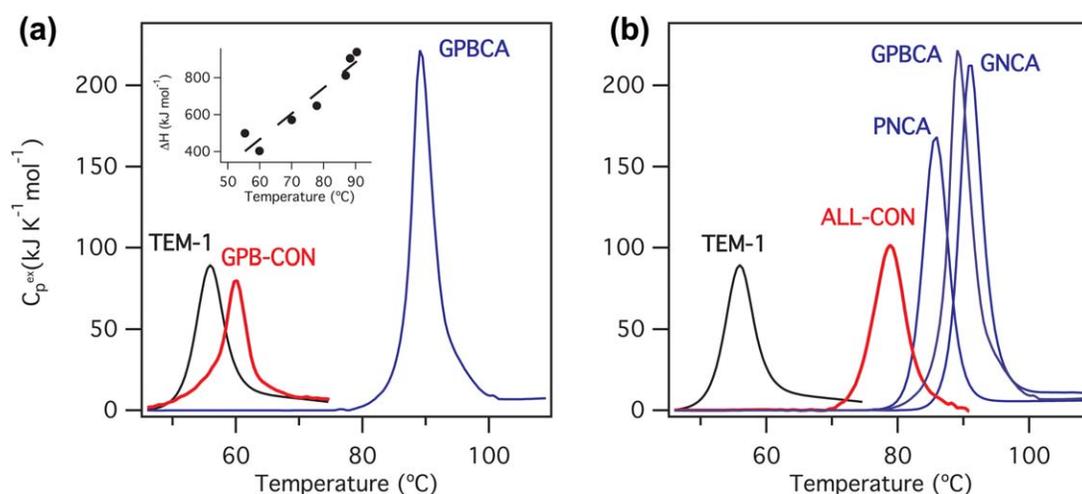
Figure 3

Structural comparison of the ALL-CON consensus lactamase, the extant TEM-1 β-lactamase (PDB code: 1BTL), and the laboratory resurrections corresponding to the ENCA, GNCA, and PNCA Precambrian nodes (Fig. 1). ENCA and GNCA structures (PDB codes: 3ZDJ and 4B88, respectively) were previously reported (Risso *et al.*[7]), while ALL-CON and PNCA structures (PDB codes: 4C75 and 4C6Y, respectively) have been determined in this work. Minor structural differences (in solvent exposed loops and the terminal helix) are labeled with arrows. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

that the ALL-CON protein has a denaturation temperature above most extant lactamases, but this value is still below the values for the ancestral lactamases dating to about 2–3 billion years old [see Fig. 4(b)]. The result that consensus engineering enhances protein stability but to a lesser extent than ancestral reconstruction does is in fact consistent with two published studies in which consensus proteins are compared with laboratory resurrections of ancestral proteins. Gaucher and coworkers[5,19] determined a denaturation temperature of 73°C for the reconstructed/resurrected elongation factor Tu of the last

bacterial common ancestor, while the corresponding consensus protein has a denaturation temperature of 60°C. Likewise, Akanuma *et al.*[20] have recently reported denaturation temperatures in the 100–110°C range for several laboratory resurrections corresponding to ~4 billion years-old nodes in the evolution of nucleoside diphosphate kinases, while a denaturation temperature of 84°C was reported for the corresponding consensus protein.

The enhanced thermostability for several laboratory resurrections of ancestral proteins corresponding to nodes of Precambrian age[5–7,20] supports that the notion

Figure 4

Differential scanning calorimetry thermograms for the denaturation of consensus lactamases, laboratory resurrections of Precambrian lactamases and the extant TEM-1 β-lactamase. (**a**) Comparison of the thermal stabilities of the consensus GPB-CON lactamase and the corresponding (see Fig. 1) GPBCA Precambrian laboratory resurrection. (**b**) Comparison of the thermal stabilities of the lactamase variant encoded by the consensus of all 75 extant sequences of Figure 1 with the laboratory resurrections of 1.5–3 billion years old lactamases. Differences in areas under the transitions reflect the expected temperature dependence of the denaturation enthalpy, as is shown in the Inset of panel A. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
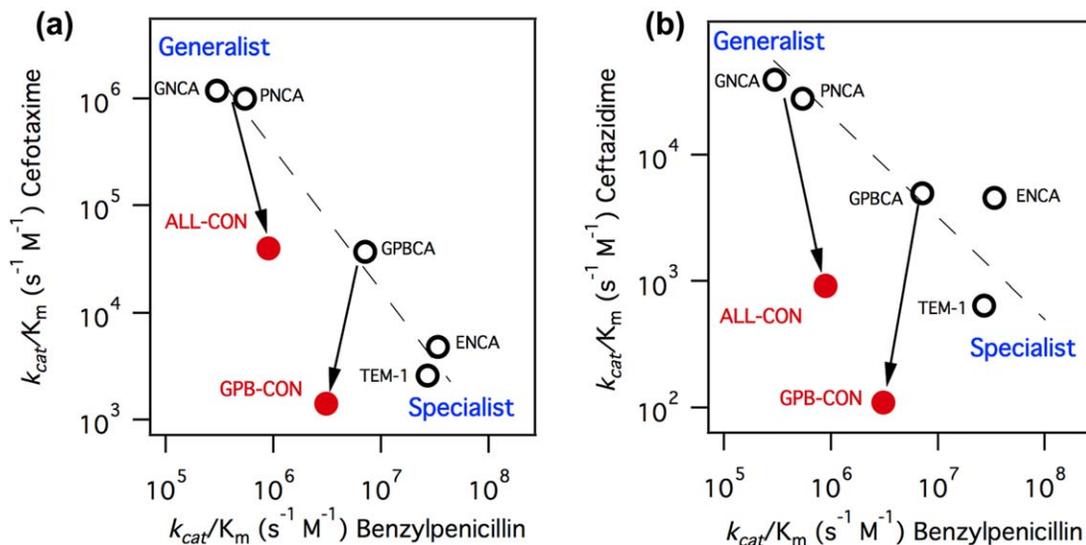
Figure 5

Plots of catalytic efficiency for the hydrolysis of third-generation antibiotics (CTX and CAZ) versus catalytic efficiency for the hydrolysis of BZ. Catalytic efficiency is measured by the Michaelis-Menten k_{cat}/K_M ratio. Values for TEM-1 lactamase and laboratory resurrections of Precambrian lactamases (ENCA, GPBCA, GNCA, and PNCA) are taken from Risso et al.[7] Values for the consensus GPB-CON and ALL-CON lactamases are determined in this work (see Fig. S2 in Supporting Information for fits of the Michaelis-Menten equation to the experimental catalysis data and supporting Information Table S3 for the K_M and k_{cat} values and their associated errors). The dashed line represents the generalist-specialist conversion defined by the data on the extant TEM-1 and the Precambrian resurrected lactamases. The consensus GPB-CON lactamase is completely off of the conversion line and is also a poorly efficient penicillin specialist (compare with its "ancestral partner," the GPBCA lactamase). The consensus ALL-CON lactamase is closer to the conversion line, at least for CTX, but does not yet reach the level of efficient substrate promiscuity as the ancestral GNCA and PNCA lactamases. Error associated to the catalytic efficiency values (given in Table S3 of Supporting Information) are in most cases smaller than the size of the data points in the plots shown here. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

that ancient life was thermophilic, that Precambrian proteins were, consequently, highly stable and that ancestral reconstruction, a broad-scale phylogenetic procedure, does provide a reasonable approximation to the ancestral protein phenotype. However, it appears that, in most cases, ancestral protein hyperstability will be captured to a lesser extent by proteins computed using the consensus approach, as shown by the ancestral/consensus comparisons reported for β-lactamases (this work), elongation factors Tu[5,19] and nucleoside diphosphate kinases.[20]

Catalysis

From our previous work, 2–3 billion year old β-lactamases displayed[7] a decreased catalytic efficiency for the degradation of penicillin antibiotics (such as BZ) compared with modern enzymes, but had an increased efficiency for degradation of third-generation antibiotics such as CTX and CAZ compared with modern enzymes. The ancestral catalytic efficiencies against these various antibiotics were determined to be roughly similar, and of the same order, as the catalytic efficiency of the average modern enzyme,[7,23] indicating that laboratory resurrections of 2–3 billion year old β-lactamases are moderately efficient as promiscuous enzymes. For instance, plots of catalytic efficiency towards a third-generation antibiotic

(CTX or CAZ) versus catalytic efficiency toward a penicillin antibiotic (BZ) provide an excellent illustration of the generalist-specialist conversion during the course of evolutionary history (Fig. 5). To determine the extent to what consensus β-lactamases capture the ancestral substrate promiscuity, we determined the Michaelis-Menten parameters (Fig. S2 and Table S3 in Supporting Information) for the degradation of BZ, CTX, and CAZ catalyzed by GPB-CON and ALL-CON (Fig. 5).

GPB-CON does not appear to be a promiscuous enzyme at all but, rather, a poorly efficient penicillin specialist, with a decreased level of catalytic efficiency for BZ (as compared with extant lactamases and even with the ancestral GPBCA lactamase) and negligible levels of activity toward third-generation antibiotics. In terms of catalytic efficiency, GPB-CON lactamase does not overlap with the generalist-(promiscuous)-to-specialist characteristic that defines the ancient Precambrian lactamases (Fig. 5). Conversely, the ALL-CON lactamase does show an increased efficiency toward third-generation antibiotics concomitant with a decrease of efficiency towards penicillin. Yet, ALL-CON lactamase does not reach the level of substrate promiscuity of the most ancient β-lactamases and it is an outlier in the generalist-specialist conversion line as defined by the resurrected proteins (Fig. 5).

CONCLUSIONS

Laboratory resurrections of ancestral proteins have often led to coherent evolutionary narratives that provide information about the environment surrounding ancient life and the adaptation of these biomolecules to changing environments over planetary time scales. Ancestral protein resurrection involves a large number of mutational changes and may lead to variants of biotechnological interest based on unique and/or extreme properties, especially when targeting billion-year-old proteins. These features are well illustrated by our recent work on laboratory resurrections of Precambrian β-lactamases[7] and Elongation Factors[5,19,24] in which highly stable and efficient promiscuous enzymes were obtained, thus providing support for the thermophilic character of Precambrian life and for the generalist-specialist conversion during natural evolution. Here, we have compared the properties of previously reported laboratory resurrections of Precambrian β-lactamases[7] with those of variants encoded by consensus sequences based on the same lactamase sequence alignment previously used for ancestral sequence reconstruction. As expected, the computed consensus sequences show substantial levels of identity with the Precambrian reconstructed sequences (Table I). This lends credence to the notion that back-to-consensus mutations may bring a modern sequence closer to that of an inferred ancestor[18] and provides some justification for the combined discussion of consensus and ancestral sequences in terms of protein stability.[25] However, the comparatively small number of amino acid differences between consensus and ancestral reconstructions (Table I) does have an impact on protein properties (i.e., substantial stability and catalytic impacts). We thus argue that the consensus approach fails to consistently achieve enhancements in stability and substrate promiscuity as seen with ancestral proteins on the order of billions-of-years old. In fact, out of the three consensus variants targeted, one of these could not be prepared in the laboratory (likely because of misfolding and low stability), one showed little stability enhancement and poor catalysis, and only one displayed clearly enhanced stability and substrate promiscuity albeit to a substantially lower extent than laboratory resurrections of 2–3 billion year old β-lactamases.

The consensus approach, as well as other sequence statistics approaches, as a starting point for protein engineering still holds value of course. Back-to-consensus mutations have been shown to be able to modulate protein properties (to enhance stability in most cases) in many systems[9–17] and correlation analyses of sequence alignments have been used to engineer allosteric communication,[26] to design sequences to fold to target structures[27] and to simultaneously enhance primary and promiscuous enzyme activities.[28] Nevertheless, there are clear differences between simple statistical analyses of sequence alignments and reconstructed ancestral proteins derived from the same alignments. Indeed, our results: (i) highlight the phenotypic differences between consensus variants and laboratory resurrections of ancestral proteins (of Precambrian age in particular); (ii) disfavor over-arching interpretations of consensus proteins as accurate phenotypic representations of ancestral proteins; (iii) support the notion that Precambrian resurrections are a more robust approach to the preparation of protein variants with large numbers of mutational changes and extreme properties. These results seem reasonable considering that ancestral sequence reconstruction is a broad-scale phylogenetic procedure that uses information of all the extant sequences to simultaneously and consistently reconstruct the sequences at all the nodes in the phylogenetic tree.

ACKNOWLEDGMENTS

REFERENCES

1. Pauling L, Zuckerkandl E. Chemical paleogenetics. Molecular "restoration studies" of extinct forms of life. Acta Chem Scan 1963; 17:S9–S16.
2. Benner SA, Sassi SO, Gaucher EA. Molecular paleoscience: systems biology from the past. Adv Enzymol Relat Areas Mol Biol 2007;75: 1–132.
3. Liberles DE, editor. Ancestral sequence reconstruction. Oxford: Oxford University Press; 2010. p 252.
4. Harms MJ, Thornton JW. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat Rev Genetics 2013;14:559–571.
5. Gaucher EA, Govindarajan S, Ganesh OK. Paleotemperature trend for Precambrian life inferred from resurrected proteins. Nature 2008;451:704–707.
6. Perez-Jimenez R, Ingles-Prieto A, Zhao ZM, Sanchez-Romero I, Alegre-Cebollada J, Kosuri P, Garcia-Manyes S, Kappock TJ, Tanokura M, Holmgren A, Sanchez-Ruiz JM, Gaucher EA, Fernandez JM. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nat Struct Mol Biol 2011;18:592–596.
7. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. J Am Chem Soc 2013;135:2899–2902.
8. Ingles-Prieto A, Ibarra-Molero B, Delgado-Delgado A, Perez-Jimenez R, Fernandez, JM, Gaucher EA, Sanchez-Ruiz JM, Gavira JA. Conservation of protein structure over four billion years. Structure 2013;21:1–8.
9. Steipe B, Schiller B, Plückthum A, Steinbacher S. Sequence statistics reliably predict stabilizing mutations in a protein domain. J Mol Biol 1994;240:182–192.
10. Lehman M, Pasamontes L, Lassen SF, Wyss M. The consensus concept for thermostability engineering of proteins. Biochim Biophys Acta 2000;1543:408–415.
11. Godoy-Ruiz R, Perez-Jimenez R, Ibarra-Molero B, Sanchez-Ruiz JM. Relation between protein stability, evolution and structure, as probed by carboxylic acid mutations. J Mol Biol 2004;336:313–318.

12. Godoy-Ruiz R, Ariza F, Rodriguez-Larrea, Perez-Jimenez R, Ibarra-Molero B, Sanchez-Ruiz JM. Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. J Mol Biol 2006;362:966–978.
13. Kohn A, Binz HK, Forrer P, Stumpp MT, Plückthun A, Grütter MG. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc Natl Acad Sci USA 2003:100:1700–1705.
14. Amin N, Liu AD, Ramer S, Aehle W, Meifer D, Metin M, Wong S, Gualfetti P, Schellenberger V. Construction of stabilized proteins by combinatorial consensus mutagenesis. Protein Eng Des Sel 2004;17:787–793.
15. Khersonsky O, Kiss G, Röthlisberger D, Dym O, ALbeck S, Houk KN, Baker D, Tawfik DS. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. Proc Natl Acad Sci USA 2012;109:10358–10363.
16. Durani V, Magliery TJ. Protein engineering and stabilization from sequence statistics: variation and covariation analysis. Methods Enzymol 2013;523:237–256.
17. Rodriguez-Larrea D, Perez-Jimenez R, Sanchez-Romero I, Delgado-Delgado A, Fernandez JM, Sanchez-Ruiz JM. Role of conservative mutations in protein multi-property adaptation. Biochem. J. 2010;429:243–249.
18. Bershtein S, Goldin K, Tawfik DS. Intense neutral drifts yield robust and evolvable consensus proteins. J Mol Biol 2008;379:1029–1044.
19. Cole MF, Gaucher EA. Utilizing natural diversity to evolve protein function: applications towards thermostability. Curr Opin Chem Biol 2011;15:399–406.
20. Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi, A. Experimental evidence for the thermophilicity of ancestral life. Proc Natl Acad Sci USA 2013; 110:11067–11072.
21. Risso VA, Acierno JP, Capaldi S, Monaco HL, Ermácora MR. X-ray evidence of a native state with increased compactness populated by tryptophan-less B. licheniformis β-lactamase. Protein Sci 2012;21:964–976.
22. Otálora F, Gavira JA, Ng JD, Garcia-Ruiz JM. Counterdifussion methods applied to protein crystallization. Prog Biophys Mol Biol 2009;101:26–37.
23. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 2011;50:4402–4410.
24. Zhou Y, Asahara H, Gaucher EA, Chong S. Reconstitution of translation from Thermus Thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. Nucleic Acid Res 2012;40:7932–7945.
25. Wijma Hj, Floor RJ, Janssen DB. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. Curr Opin Struc Biol 2013;23:588–594.
26. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R. Surface sites for engineering allosteric control in proteins. Science 2008;322:438–442.
27. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, Ranganathan R. Evolutionary information for specifying a protein fold. Nature 2005;437:512–518.
28. Garcia-Seisdedos H, Ibarra-Molero B, Sanchez-Ruiz JM. Probing the mutational interplay between primary and promiscuous protein functions: a computational-experimental approach. PLoS Comp Biol 2012;8:e1002558.