jmb Article

What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-offs

Eynat Dellus-Gur†, Agnes Toth-Petroczy†, Mikael Elias and Dan S. Tawfik

Department of Biological Chemistry, *Weizmann Institute of Science, Rehovot 76100, Israel*

Correspondence to Mikael Elias and Dan S. Tawfik: *mikael.elias@weizmann.ac.il; dan.tawfik@weizmann.ac.il*
http://dx.doi.org/10.1016/j.jmb.2013.03.033
Edited by A. Panchenko

Abstract

Protein evolvability includes two elements—robustness (or neutrality, mutations having no effect) and innovability (mutations readily inducing new functions). How are these two conflicting demands bridged? Does the ability to bridge them relate to the observation that certain folds, such as TIM barrels, accommodate numerous functions, whereas other folds support only one? Here, we hypothesize that the key to innovability is *polarity*—an active site composed of flexible, loosely packed loops alongside a well-separated, highly ordered scaffold. We show that highly stabilized variants of TEM-1 β-lactamase exhibit selective rigidification of the enzyme's scaffold while the active-site loops maintained their conformational plasticity. Polarity therefore results in stabilizing, compensatory mutations not trading off, but instead promoting the acquisition of new activities. Indeed, computational analysis indicates that in folds that accommodate only one function throughout evolution, for example, dihydrofolate reductase, ≥60% of the active-site residues belong to the scaffold. In contrast, folds associated with multiple functions such as the TIM barrel show high scaffold–active-site polarity (~20% of the active site comprises scaffold residues) and >2-fold higher rates of sequence divergence at active-site positions. Our work suggests structural measures of fold polarity that appear to be correlated with innovability, thereby providing new insights regarding protein evolution, design, and engineering.

Introduction

In silico and experimental analyses indicate that protein stability confers tolerance to mutations and thereby promotes protein evolvability.[1–6] Evolvability, however, has two components that are interlinked but also potentially contradictory.[7,8] The accumulation of mutations while maintaining the original function (neutrality, or genetic robustness) is one component of protein evolvability. The acquisition of new functions, termed here innovability, is another component (the term innovability was adopted from Ref. 8; in here, innovability relates to the divergence of novel protein functions, rather than the enhancement of latent, promiscuous activities). That stability promotes robustness and therefore neutral evolution, is well established and biophysically understood. A highly ordered, well-packed protein affords a higher stability margin, or threshold, and enables more destabilizing mutations to accumulate.[1–4,6,9] What about innovability, namely, the ability to evolve new functions? On the one hand, mutations that promote new functions tend to be destabilizing,[10–12] and thus, excess stability would promote innovability.[5] On the other hand, increased stability coincides with reduced conformational plasticity.[13,14] The acquisition of new functions, however, often depends on conformational plasticity—the coexistence of multiple structural conformers.[15,16] This is certainly the case with the adaptation of β-lactamases towards new antibiotics,[17,18] including TEM-1, the model enzyme studied here.[19–21] Thus, higher stability could also hamper innovability by reducing conformational plasticity.

It is often the case that activity and stability trade off.[10,22] Many enzymes were, however, dramatically stabilized without compromising their activity.[23,24] Do stability and innovability trade-off? Namely, could excess stability hamper the effect of function-modifying mutations? Does the existence or absence, of a stability–innovability trade-off relate to the protein's architecture?

Additionally, stability promotes evolvability only if stability is an additive, global parameter, whereby

stabilizing mutations in one region (e.g., a protein's scaffold) readily compensate for the destabilizing effects of mutations in other locations (e.g., in the active-site region). While this is the prevailing model,[4,5] can it be taken for granted? In some proteins, higher stability is mediated by mutations in residues that mediate function, suggesting that stability and function do trade off.[22] In other cases, the compensatory mutations are in direct contact with the function-modulating mutations, that is, local, specific suppressors[25] and not global ones.

We thus sought to explore certain aspects of stability–evolvability. Firstly, whether, and why, the increased rigidity conferred by stabilizing mutations does not affect the active site's conformational plasticity, and/or the ability to acquire new functions. Secondly, do the lack of stability–innovability trade-offs and the enhancement of evolvability as well as innovability by stabilizing mutations relate to the protein's architecture? Protein folds seem to dramatically differ in their innovability—TIM barrels, for example, underline >120 different enzyme families with different functions and with no apparent sequence identity. Hence, the TIM barrel fold exhibits high innovability and robustness. However, other folds, for example, dihydrofolate reductase (DHFR), are associated with only one enzymatic function. Is this a coincidence, or are certain protein architectures less prone to trade-offs and more amenable to functional innovation?

Results

TEM-1 β-lactamase is an established model for the emergence of new enzymatic specificities. Originally conferring resistance to naturally occurring penicillins, TEM-1 rapidly evolved to hydrolyze newly introduced synthetic antibiotics such as cephalosporins.[21] This adaptation involves mutations that modulate the configuration of active-site loops.[19] Hence, TEM-1's adaptation is mediated by conformational plasticity.[20] TEM-1's new-function mutations are also strongly destabilizing and are usually followed by stabilizing, compensatory mutations.[10] The latter were found to be consensus/ancestor substitutions[26]—that is, substitutions that take the protein closer to the sequence of its family consensus and/or predicted ancestor.[27,28] Using such substitutions, we generated highly stable variants of TEM-1 and studied their evolvability, in terms of robustness and innovability.

Directed evolutions of stabilized TEM-1 variants

We used consensus/ancestor substitutions known to have no effect, when individually introduced, on enzymatic activity[29–31] (Supplementary Table 1). These were introduced to wild-type TEM-1 to yield a library containing various combinations of consensus/ancestor mutations. To isolate stabilized variants, we used a strategy similar to the one applied by Hecky and Muller.[29] Two severely destabilizing point mutations (L76N and R222C)[26] were introduced, and the library was selected for growth on ampicillin for gain of TEM-1's antibiotic resistance function. Two TEM-1 variants from the selected library were characterized in detail (v3 and v13). They were chosen because they did not carry any random mutations and had a number and composition of stabilizing mutations that represented all the selected variants. Overall, 10 out of the 13 mutations included in the library were represented

Table 1. Kinetic and stability parameters of the TEM-1 variants

Variant	Antibiotic	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})	Apparent T_m (°C)
Wild type	Amp	71 ± 9	1857 ± 27	$26 \pm 1.3 \times 10^6$	55 ± 0.01
	Ctx	n.d.	n.d.	$0.1 \times 10^4 \pm 6$	
Stabilized v13	Amp	63 ± 13	1040 ± 21	$16.5 \pm 1.3 \times 10^6$	69 ± 0.03
	Ctx	n.d.	n.d.	$0.07 \times 10^4 \pm 10$	
Stabilized v3	Amp	186 ± 23	4536 ± 333	$24 \pm 1.8 \times 10^6$	71 ± 0.05
	Ctx	n.d.	n.d.	$0.06 \times 10^4 \pm 13$	
Wild type + G238S	Amp	26 ± 4	42 ± 0.4	$1.4 \pm 0.08 \times 10^6$	50 ± 0.02
	Ctx	403 ± 146	50 ± 3	$13 \pm 4 \times 10^4$	
Stabilized v13 + G238S	Amp	18 ± 5	72 ± 0.5	$4 \pm 0.36 \times 10^6$	70 ± 0.04
	Ctx	502 ± 65	75 ± 1.4	$14.5 \pm 1 \times 10^4$	
Stabilized v3 + G238S	Amp	88 ± 15	158 ± 13	$1.8 \pm 0.3 \times 10^6$	69 ± 0.04
	Ctx	440 ± 40	28 ± 1	$6.2 \pm 0.3 \times 10^4$	
Wild type + G238S + E104K	Amp	28 ± 8	32 ± 0.8	$1.1 \pm 0.12 \times 10^6$	50 ± 0.02
	Ctx	88.5 ± 15	40 ± 1.1	$45 \pm 4 \times 10^4$	
Stabilized v13 + G238S + E104K	Amp	20 ± 5	31 ± 0.5	$1.4 \pm 0.15 \times 10^6$	66 ± 0.05
	Ctx	78 ± 11.5	32 ± 0.8	$40 \pm 3.5 \times 10^4$	
Stabilized v3 + G238S + E104K	Amp	89 ± 15	160 ± 13	$1.8 \pm 0.3 \times 10^6$	67 ± 0.04
	Ctx	51 ± 5	62 ± 2.3	$112.5 \pm 4 \times 10^4$	

n.d., a linear dependency of initial rates on substrate concentrations was observed (substrate concentrations were limited by maximal absorbance), and only k_{cat}/K_M could be determined.
Amp, ampicillin; Ctx, cefotaxime.

by the combination of these two variants (Supplementary Table 1). Upon removal of the destabilizing mutations, v3 and v13 exhibited apparent melting temperatures (T_m) that are ~14 °C higher than that of wild-type TEM-1 (Table 1).

Activity and evolvability of the stabilized variants

To assess the ability of the stabilized TEM-1 variants to gain new functions, we examined the effects of two commonly observed active-site mutations—G238S and E104K.[32] These mutations confer resistance to third-generation cephalosporins such as cefotaxime[33,34] both in the clinic and in the laboratory evolution of TEM-1. G238S and E104K were incorporated into wild-type TEM-1, and into the stabilized variants v3 and v13, individually and in combination. Their effects on the enzymatic parameters and the enzyme's structure were examined.

The kinetic parameters were measured for both ampicillin, TEM-1's original substrate, and cefotaxime, the newly evolved substrate (Table 1). The stabilized variants and wild type exhibited nearly identical kinetic parameters with ampicillin. This indicates, as previously noted,[24] that the large excess of stability did not reduce enzymatic activity. The kinetic parameters were also similar with cefotaxime: the variations in k_{cat}/K_M were ≤4-fold

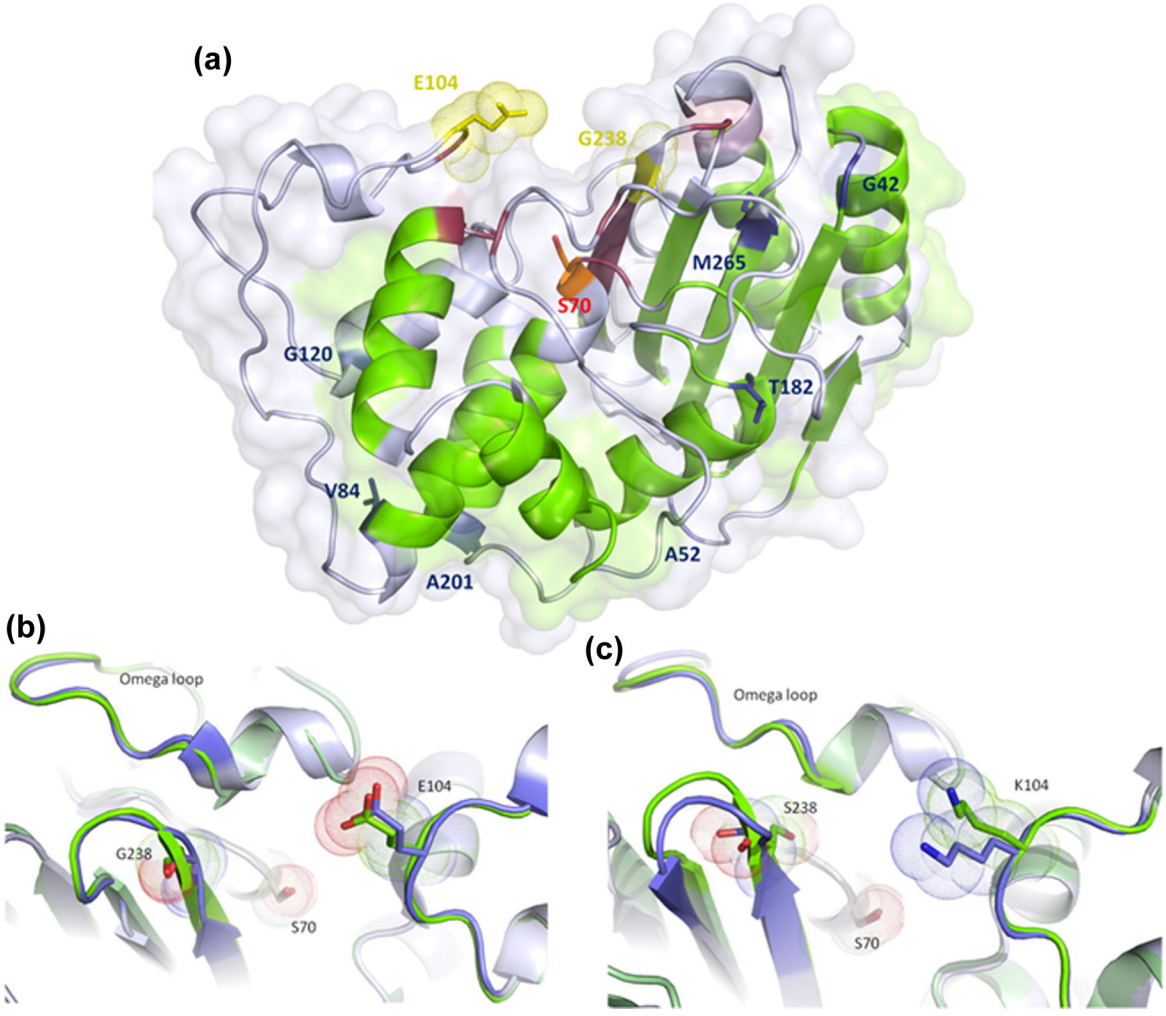

Fig. 1. TEM-1's stabilizing and new-function mutations. (a) Structure of TEM-1 variant v13 (PDB code: 4IBX) with the stabilizing mutations mapped as blue sticks. The scaffold, colored green, was assigned as the secondary-structure elements shared by all structures in the family (Class A β-lactamases; see Materials and Methods). The remaining, non-scaffold parts are in gray, and the active-site residues are in red (defined residues with one or more atoms at ≤4 Å distance from the substrate). The catalytic serine, S70, is in orange sticks. The residues bearing the new-function mutations (E104K and G238S) are presented in yellow sticks. (b) Superposition of the active-site loops of the stabilized v13 (green, PDB code: 4IBX) and wild-type TEM-1 (blue, PDB code: 1ZG4). (c) The active-site loops of v13 and wild type carrying the new-function mutations E104K and G238S (green, PDB code 4IBX; blue, PDB code 1HTZ, respectively). Residue numbers are according to wild-type TEM-1 structure.

with no systematic trends. It seems therefore that the incorporated stabilizing consensus/ancestor mutations did not limit the enzyme's ability to acquire a new activity, even upon combining seven different consensus/ancestor mutations. Further, wild-type TEM-1's stability is severely compromised by the new-function mutations, especially when G238S and E104K are combined.[10] In contrast, the stabilized variants carrying G238S and E104K exhibited much higher T_m values than wild type with no new-function mutations (Table 1). Thus, despite being far from TEM-1's active site (Fig. 1), the introduced consensus/ancestor, stabilizing substitutions acted as compensators for the new-function mutations (as shown before for individual stabilizing substitutions such as M182T[10,11]). The stabilizing substitutions did not compromise TEM-1's original activity, yet increased its ability to adopt new substrate specificity (i.e., increased its innovability).

Further, several of these stabilizing mutations, individually and in combination, have been shown to increase TEM-1 robustness towards random mutations.[6,26] Hence, the consensus/ancestor-stabilized TEM-1 variants exhibit higher stability, higher mutational robustness, and higher ability to acquire new functions. It therefore appears that in TEM-1, the apparently contradicting components of evolvability, that is, robustness and innovability, can be readily reconciled by global suppressors that promote both.

Structural analysis

To better understand the absence of trade-offs, we crystallized the stabilized variant v13 and the same variant carrying the new-function mutations G238S and E104K (2.68 and 2.2 Å resolution; Supplementary Table 2). As expected, the structures of the stabilized v13 are essentially identical with the published structures of wild-type TEM-1 (RMSD of 0.43 and 0.58 for the wild type and G238S and E104K mutants, respectively). The only exception is a slight shift in the loop that includes position 238 in

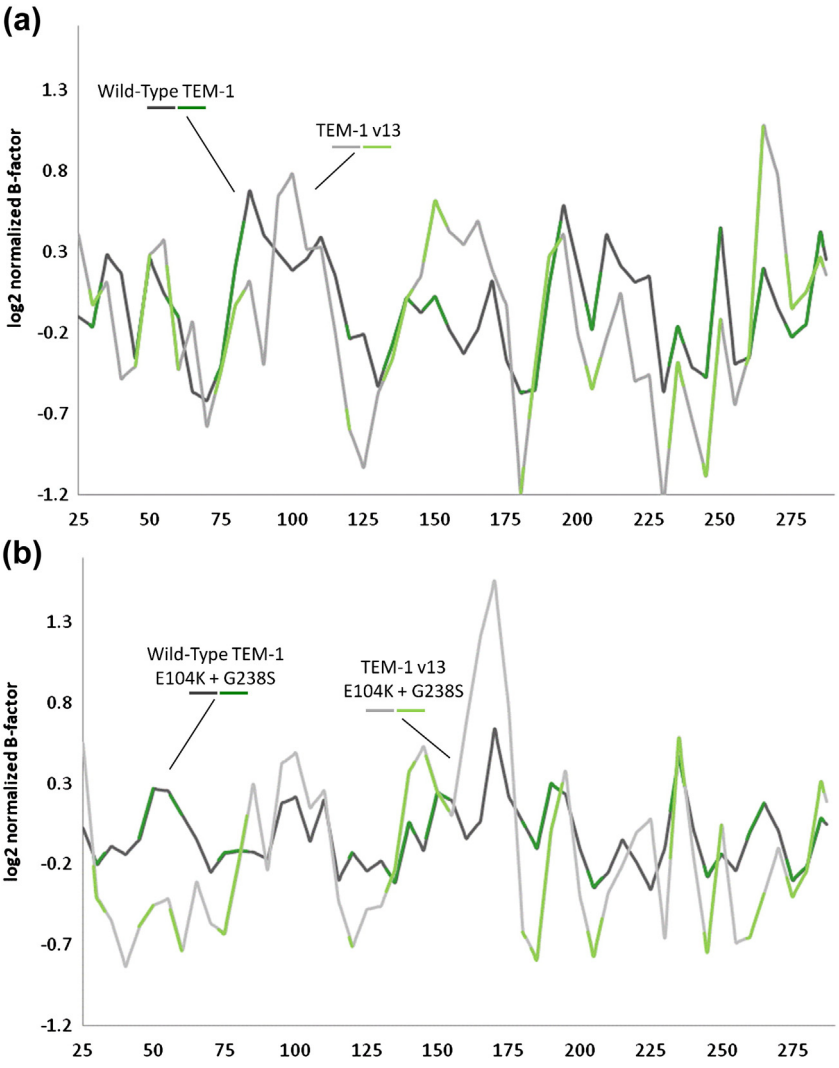

Fig. 2. Positional distributions of normalized *B*-factor values (the *x*-axis represents residue number). Shown, for clarity, are the log2 values of the normalized *B*-factors. The green parts of the lines mark scaffold residues, and gray parts indicate all other residues. (a) Wild-type TEM-1 (PDB code: 1ZG4, in dark gray and dark green) compared to stabilized v13 (PDB code: 4IBX, in light gray and green). (b) Wild-type TEM-1 and stabilized v13 carrying the new-function mutations E104K and G238S (PDB codes: 1HTZ and 4IBR).

the structures of the E104K + G238S mutants of wild type *versus* v13 (Fig. 1a and b).

The structural impacts of M182T[10,31] and L201P[31] (L201A in our variant) were previously analyzed, and these mutations are likely to have similar effects in v13. A structure containing T265M is firstly described here. Threonine 265 is a buried residue, and methionine may fit better within the core's hydrophobic environment. The stabilization mechanism of R120G is not evident from the structure. However, this and other consensus/ancestor substitutions improve TEM-1's kinetic stability.[26] The structural effects of the remaining substitutions (A42G, N52A, and I84V) are also not evident, but the mechanism of stabilization is not the aim of this study.

Mapping structural rigidity and flexibility

The structures enabled us to assess the relative rigidity of various parts of TEM-1's fold. We aimed to find support for the following notions: (i) The improved mutational robustness of variant v13 is associated with improved packing and a more rigid scaffold; (ii) the ability to evolve towards alternative antibiotics relates to the structural flexibility of TEM-1's active-site loops; (iii) the above two are independent—namely, that a better-packed and rigidified scaffold has no effect on the mobility of the active-site loops.

We examined the refined crystallographic temperature factors (*B*-factors) along TEM-1's polypeptide chain. *B*-factors (temperature or Debye–Waller factors) are obtained for each atom in a crystal structure. They describe the degree to which the electron density is scattered and reflects how mobile an atom is.[35,36] *B*-factors cannot be treated as absolute values, because they also depend on the refinement procedure, crystal packing, and other data parameters. They can, however, be normalized and thus used to compare the relative mobility of atoms within the same structure.[36] To minimize biases related to refinement parameters, we set the occupancy for all atoms to 100% and re-refined the published wild-type structures with the same protocol used to refine the newly obtained structures. Additionally, only main-chain atoms were included in the analysis. The resulting per-residue *B*-factors were normalized to the average *B*-factor of the entire protein.

The overall *B*-factor patterns of the stabilized v13 and of wild-type TEM-1, that is, the regions of relatively low *versus* high mobility, are similar (Fig. 2a). However, v13's *B*-factor values are generally more polarized than those of wild type (standard deviation values are 0.39 for wild type *versus* 0.58 for v13). Foremost, in v13, the differences between the protein's scaffold (in green) and non-scaffold parts (in gray), and particularly the active-site loops, are larger than in wild type. Towards this comparison, TEM-1's scaffold residues were defined as secondary-structure elements that are shared by all structures in the family (Fig. 1). Active-site residues were defined as all residues within a contacting distance from the substrate (see Materials and Methods for details).

The above observation can be interpreted in two ways (or by a combination of both): (i) The non-scaffold parts, including the active-site loops of the stabilized v13, exhibit higher mobility than in wild type and/or (ii) v13's scaffold is more rigid than that of the wild type. Given the markedly increased T_m of v13, and the identical kinetic parameters of wild type and v13, the second explanation is more likely. Further, thermal inactivation tests indicated that v13 maintained activity at higher temperatures than wild type, supporting the notion that its active-site loops are not more conformationally mobile than wild type's (Supplementary Fig. 1). Overall, it appears that stabilization resulted in v13 having a more rigid scaffold and core, while its active-site loops maintained a degree of flexibility that is similar to wild type's.

Fold polarity and evolvability

An even stronger trend of polarization is seen upon comparing the *B*-factors of v13 carrying the new-function mutations E104K and G238S to the published structure of wild-type TEM-1 carrying these mutations and only one stabilizing mutation, M182T.[20] The standard deviation for the *B*-factor patterns is 0.26 at the wild-type's background *versus* 0.57 at v13's, that is, >2-fold higher polarity (Fig. 2b). (Note that the term *polarity* is used here to describe a lower level of connectivity within the very same structural entity, or domain, whereas *modularity* is usually used in the context of multi-domain proteins, or other clearly separated structural elements.) These enzyme variants (wild type and v13 with E104K and G238S) exhibit essentially identical kinetic parameters but differ in stability (T_m for wild-type's background was 16 °C lower than at v13's; Table 1). This implies that the new-function mutations (E104K and G238S) primarily increased the flexibility of the active-site loops and that flexibility was increased in wild type and in v13 to a similar degree. Indeed, when binding site residues' *B*-factor is set to the same value in both of the structures, v13 scaffold is 1.3-fold more stable than the wild type's.

Due to its mobile active-site loops *versus* a highly ordered and tightly packed scaffold, v13's structure is highly polarized. Further, the stabilizing consensus/ancestor substitutions suppressed the destabilizing effects of TEM-1's new-function mutations via increased scaffold stability and rigidity. The stabilizing mutations did not, however, suppress the conformational plasticity of the active-site loops such that the new-function mutations exerted their

full effect. The differences in fold polarity between wild type and v13 can also be seen upon comparing the *B*-factor patterns of each variant with and without the new function mutations (Supplementary Fig. 5). In wild type, the increase in normalized *B*-factors for the loop regions upon the introduction of E104K and G238S is relatively small (e.g., in the omega-loop, in which conformational changes that relate to TEM-1's ability to hydrolyze cefotaxime are observed). The small change is likely to be the outcome of the core parts also showing increased mobility, consistent with the T_m lowering to 50 °C. In v13, however, the increased loop flexibility in the variant carrying E104K + G238S is clearly seen, primarily because the core remained rigid (consistent with a T_m of 66 °C).

Overall, it appears that the separation of the scaffold from the active-site loops, or polarity, as defined here, underlines TEM-1's evolvability. It enables TEM-1 to be robust to mutations as well as adaptable towards new functions. Are, then, folds with high polarity more evolvable than those exhibiting low polarity? Namely, would folds in which most functional residues are weakly coupled to the scaffold be more evolvable, in terms of mutational robustness and/or innovability?

Is fold polarity a common feature of innovable folds?

To further explore the polarity concept, we have attempted to develop a bioinformatic framework for analysis of polarity parameters and thus provide preliminary indications for the correlation of fold polarity and innovability. Unlike the micro scale (comparing different TEM-1 mutants), at the macro scale (comparing different folds), comparison of *B*-factors is irrelevant. We therefore developed generic measures of fold polarity. By our model, enzymes where a smaller fraction of the active-site residues are part of the scaffold (criterion i) and/or where the active-site residues are weakly bonded to the scaffold (criterion ii) would more readily adopt new functions. This is because the scaffold and the active site can accommodate mutations with fewer mutual constrains. In other words, high connectivity of the scaffold and the active-site residues is more likely to be associated with strong stability–activity and stability–innovability trade-offs, and hence with the need for local, specific compensatory mutations.

An exhaustive comparison of all folds presents inherent difficulties, including different birth ages (younger folds had less chance to diverge) and

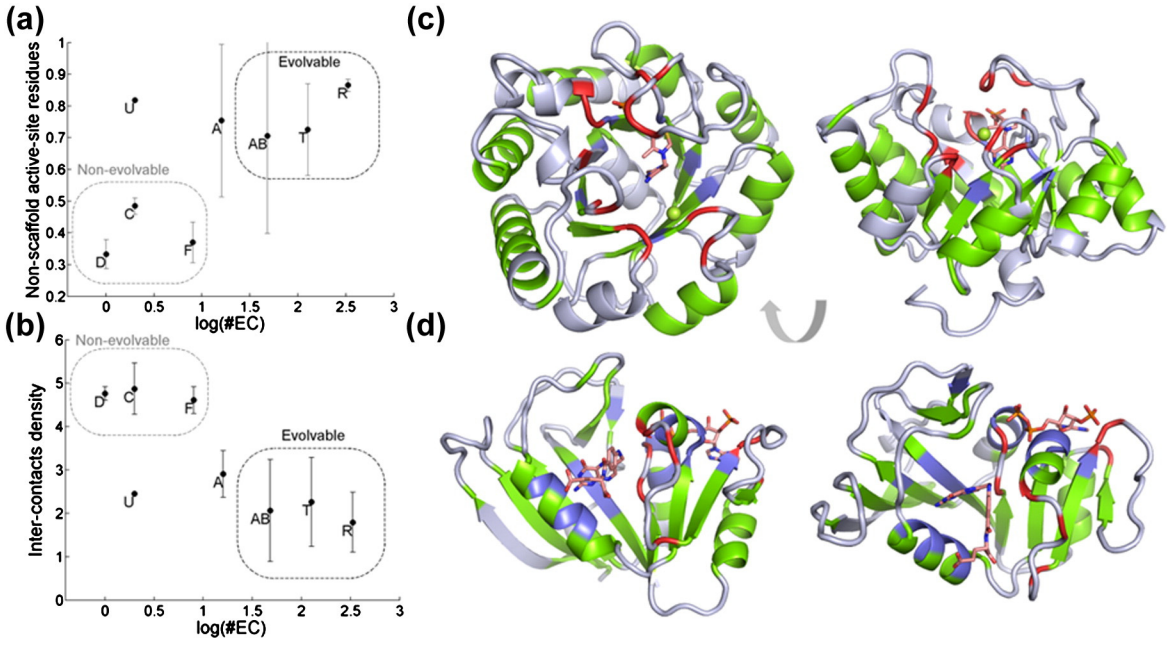

Fig. 3. Fold polarity and innovability. The *x*-axis denotes innovability as the number of different enzymatic functions (E.C. numbers) associated with a given fold (log10 scale). D, DHFR; C, chitinase; F, farnesyl transferase; U, UDG; A, aminopeptidase; AB, α–β plaits; T, TIM barrels; R, Rossmann fold. (a) Innovability as a function of the fraction of non-scaffold active-site residues (the number of active-site residues that are not part of the scaffold normalized by the total number of active-site residues). (b) Inter-contacts density (the number of contacts between scaffold and active-site residues normalized by the total number of active-site residues) as a function of innovability. (c) The TIM barrel exemplifies a highly innovable fold. Drawn is PDB code: 2TPS, thiamin phosphate synthase. Scaffold residues are colored green and substrates are shown as sticks. Active-site residues belonging to the scaffold are in blue, and non-scaffold active-site residues are in red. (d) A depiction following the same color scheme of a non-innovable fold exemplified by DHFR (PDB code: 1RA8).

biased sampling of known structures and functions. We therefore examined folds that are predicted to have existed in the last universal common ancestor (LUCA)[37] and thus are of similar age. Out of 87 LUCA folds, 63 folds possess enzymatic functions. We focused on single-domain enzymes, as functional diversification often relates to multi-domain arrangements.[38] And we addressed folds for which high-resolution structures in complex with a relevant ligand are available (for assigning active-site residues) and with enough representative examples of different functions and/or different species. Albeit, this filtering reduced our data set to eight different folds (Fig. 3; Supplementary Table 3).

As a measure of the ability to acquire new functions (innovability), we used the number of different E.C. (Enzyme Commission) numbers assigned to each fold. This parameter is biased by the nature of the reactions, and how easily they may diverge within a given active-site chemistry. It does, however, provide a systematic measure of the level of functional diversification,[39] certainly for the order-of-magnitude differences in functional diversity examined here. We tested the robustness of this definition by considering either four or three hierarchy levels in the E.C. number assignment and obtained essentially the same results (Supplementary Fig. 6). Thus, the following analyses were done by considering all four E.C. digits (e.g., TEM-1 β-lactamase, 3.5.2.6).

To measure the above-mentioned polarity criteria, a general way of defining the scaffold and active-site residues is needed. We defined the scaffold as the secondary-structure elements (helices, strands) shared by all enzymes belonging to a given fold. This criterion captured the essence of TEM-1's fold (Fig. 1) as well as of the LUCA folds (Supplementary Fig. 2). Active-site residues were defined as residues for which at least one atom is within 4 Å of the bound ligand. For all analyzed folds, the scaffold residues exhibited higher contact density values relative to the active-site ones (Supplementary Fig. 3). Thus, as expected, the scaffold represents the most tightly packed and highly ordered part, whereas the active-site residues are primarily located on the surface of the proteins and are more loosely packed.

Innovability correlates with polarity

Examination of the fraction of non-scaffold active-site residues (criterion i) revealed that in the highly innovable folds, on average, only 20% of the active-site residues comprise part of the scaffold. Conversely, with the exception of one fold [uracyl-DNA glycosilase (UDG)], folds that carry only a few or several different functions have >50% of the active-site residues as part of their scaffold (Fig. 3a). TEM-1 exhibits a mid-value with ~50% of the active-site residues belonging to the scaffold. Indeed, although TEM-1 rapidly adopts to hydrolyze different β-lactams, its fold is associated with only three different E.C. numbers (Supplementary Table 3). Analysis of criterion ii indicated that the most innovable folds in our data set exhibited approximately half of the inter-contacts density exhibited by non-innovable ones (Fig. 3b; 2.1 *versus* 4.7, on average; TEM-1 shows a mid-value of 3.3).

The above-noted trends were reproduced in an independently assembled data set of the yeast (*Saccharomyces cerevisiae*) enzymes that belong to the same folds, thus indicating that the polarity criteria are indeed fold specific (Supplementary Fig. 4). Overall, it appears that high polarity is a necessary requirement for innovability in single-domain enzymes (all highly evolvable folds show high polarity). Polarity, however, is clearly not the only criterion. Certain active-site chemistries might be, for example, more limited in their divergence ability, whereas other folds might adopt new

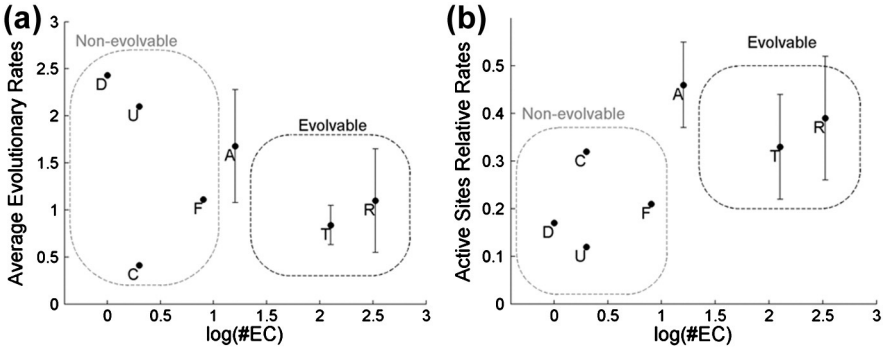

Fig. 4. Fold polarity and mutational robustness. (a) The average evolutionary rates of the analyzed folds (normalized number of amino acid exchanges per protein, per generation) and their functional diversification (innovability) do not seem to be correlated. (b) In contrast, the relative evolutionary rates of the active-site residues (normalized by the average rates per protein) correlate with innovability.

functions primarily through divergence in an auxiliary, 'cap' domain inserted within their core domain.[38]

Evolutionary rates and polarity

As noted in the Introduction, mutational robustness, or neutrality, promotes innovations, as mutations acquired as neutral in one context pave the way for new functions under changed circumstances.[7,8] This general principle has also been demonstrated for the acquisition of new enzymatic functions.[40] Neutrality, or robustness, is manifested in higher evolutionary rates—namely, higher number of amino acid exchanges per position per generation. We obtained for enzymes that belong to the analyzed LUCA folds the average evolutionary rates (per protein), as well as the rates per each position, using sets of orthologous sequences from closely related fungi species including *S. cerevisiae*. Interestingly, we found no correlation between the overall evolutionary rates (average rate per protein) and innovability (Fig. 4a). This supports the notion that mutational robustness alone is not a criterion for innovability. However, the relative evolutionary rates of the active-site residues—namely, rates for active-site residues normalized to the average protein rates, are nearly two folds higher for innovable folds (Fig. 4b). The higher divergence rates are not the outcome of the active-site residues of innovable folds having lower contact density (Supplementary Fig. 3) or of being more solvent exposed (Supplementary Fig. 7). Rather, the higher evolutionary rates seem to primarily correlate with the lower connectivity of the active-site residues and the scaffold (Fig. 3b). Thus, active-site residues of innovable folds drift much faster relative to their scaffolds, hence promoting the potential for evolving new functions without compromising the scaffold's integrity.

Discussion

There exists an inherent contradiction between robustness and innovability in evolution in general and in protein evolution in particular. Mutations should have minimal effects under a constant environment (robustness) but rapidly lead to adaptation towards new functions when the environment changes (innovability).[7,41] What might be the structural features that resolve these conflicting demands in proteins? We present a model by which protein folds that exhibit high scaffold–active-site polarity are highly evolvable in both respects. A well-ordered and tightly packed scaffold endows high global stability and thereby robustness to mutations. This property is primarily reflected in higher contact density and core rigidity and enables a protein to tolerate extensive sequence changes without losing its structural integrity.[1,42–44] Conversely, an active site with relatively low contact density (partial disorder) and high conformational plasticity, and that is also well separated from the protein's scaffold, facilitates the adaptation towards new functions.

The polarity model seems to apply at the micro scale, namely, to the comparison of wild-type TEM-1 and its stabilized variants. Substitutions taken from the consensus/ancestor sequence resulted in both higher robustness and higher adaptability. Structural analysis indicated the higher polarity of the stabilized variants: The enzyme's scaffold was stabilized and rigidified, while the active-site loops maintained their original conformational plasticity (Fig. 2). Under a constant environment, stabilized TEM-1 variants carrying mutations similar to those of v13 exhibit higher tolerance to mutations, that is, higher robustness.[6,26] However, upon encountering a new challenge, for example, third-generation antibiotics, their active site's conformational plasticity enables rapid adaptation. Further, in TEM-1, the scaffold's stability efficiently compensates for the stability losses associated with new-function mutations within the active-site loops.

Are laboratory experiments that examine evolutionary innovations within several years' timescale relevant to evolutionary processes that occurred along billions of years? Namely, can the fold polarity model be applied to the evolution of enzyme families and superfamilies? Our preliminary analysis indicates that the most innovable folds—TIM barrels and Rossmann folds[45]—that carry dozens of different enzymatic functions exhibit markedly high scaffold—active-site polarity (Fig. 3). In contrast, in folds such as DHFR that are associated with only one function, the active site and the scaffold are largely fused (Fig. 3). In such folds, the active site and scaffold effectively co-evolve, resulting in higher constraints and lower likelihood for the emergence of new functions. This feature is seen upon comparing evolutionary rates in orthologs, namely, rates of neutral evolution. In non-innovable, non-polar folds, there is a 'mutual freeze' whereby the active-site residues diverge very slowly. In polar folds, however, the low active-site–scaffold connectivity enables the active-site residues to drift (Fig. 4) and hence promote the potential for adaptation.[7,8,16,40] Indeed, enzyme superfamilies are defined by the usage of the same fold and few key active-site residues that are typically part of the scaffold (e.g., at the tips of the β-strands that comprise the TIM barrel's central tunnel). The remaining active-site residues dramatically vary between individual families, not only in amino acid composition but also in backbone (e.g., the huge diversity of active-site loop lengths and conformations in TIM barrels). It is worth noting that functional divergence within families and superfamilies is driven by gene duplication. Duplication initially reduces the selection pressure owing to higher protein dose. However, the trade-off between

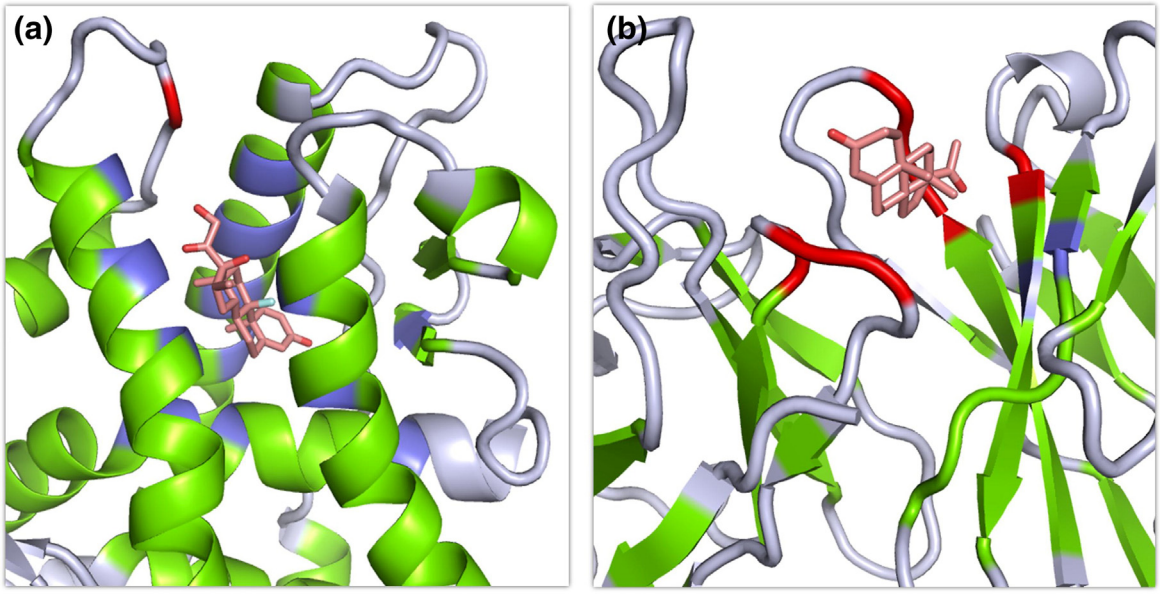

Fig. 5. Fold polarity and innovability in ligand-binding proteins. The coloring scheme follows Fig. 3. (a) Low-polarity folds exemplified by the glucocorticoid receptor in complex with dexamethasone (PDB code: 3GN8). Over 90% of the ligand-contacting residues are part of the receptor's scaffold. This fold is accordingly associated with only one function. (b) A progesterone-binding antibody uses mainly loop, non-scaffold residues for binding (PDB code: 1DBB).

stability and activity remains, and eventually, as more mutations accumulate, the inability to compensate for stability losses will impair the acquisition of a new function.

Compensation for stability losses associated with new-function mutations and adaptation might also be specific and local in low-polarity folds, and such folds might therefore be more prone to stability–activity trade-offs. This seems to be the case with Im7[22] and with glucocorticoid receptors.[25] Im7's binding surface comprises an integral part of its scaffold. The steroid binding site of glucocorticoid receptors consists of 90% scaffold residues, and the inter-contacts density is accordingly high (~7; Fig. 5a).

To conclude, the structural features that promote protein evolvability are only partially understood. Robustness (evolutionary rates) seems to be correlated with highly ordered structures.[1,42–44] However, structural complexity and intrinsic disorder were also found to be correlated with evolvability.[42] We surmise that the coexistence of order and disorder within the very same protein domain promotes evolvability, in terms of both robustness and innovability. Modularity, or 'division of labor', is a key contributor to evolution, as for example in the assembly of multi-domain proteins by 'copy and paste' of existing domains. Along the same vein, folds in which a relatively high fraction of active-site residues is weakly connected to the scaffold (i.e., with high polarity) might more readily acquire new functions.

Finally, high scaffold–active-site polarity criteria might also apply to protein engineering and design. Take, for example, antibodies in comparison to glucocorticoid receptors (Fig. 5). These possess fundamentally different architectures—only 25% of the binding site of a progesterone-binding antibody belongs to the scaffold *versus* 90% for glucocorticoid receptors. Indeed, antibodies represent an optimal fold for engineering, readily allowing, for example, loop grafting, an endeavor that is largely unattainable in other proteins. The criteria described here may therefore apply to the design of completely novel folds and towards the identification of natural folds and/or specific natural proteins that might be more amenable to engineering of new functions.

Materials and Methods

Library construction

Sequences of class A β-lactamases with >40% identity to TEM-1 (or TEM-116 in Lahey's database[46]) were retrieved as described in Ref. 26 and aligned using Tcoffee‡. The tree and ancestors were obtained as previously described.[26] Stabilizing mutations identified in a neutral drift of TEM-1[30] and in previously described stabilized TEM-1 variants[29] were compared to the alignment, and mutations that appeared both in the alignment and previous work were chosen. Altogether,

13 stabilizing mutations known from previous works and also identified as consensus/ancestor mutations[26] were included in the library (Supplementary Table 1). Introduction of one stabilizing consensus/ancestor mutation usually has a modest effect on stability, and certain mutations, or combinations of mutations, may be deleterious. Therefore, a combinatorial library was built using the ISOR method.[47] Variants in the unselected library carried an average of 6 consensus/ancestor mutations per variant and 1.6 random mutations incorporated by the PCR (Supplementary Table 1).

Selection of stabilized variants

The library was based on plasmid-encoded TEM-1[4,26] carrying two severely destabilizing mutations L76N and R222C.[26,48] In the absence of any stabilizing, compensatory mutations, this TEM-1 variant conferred no resistance to ampicillin (compared to >2500 mg/ml for wild-type TEM-1). A high-efficiency cloning method using type II restriction enzymes[26] was used to maintain library diversity and reduce contamination of the library with variants lacking the destabilizing mutations. To ensure diversity ($\geq 10^6$ individual clones), we plated samples from the transformed bacteria with serial dilutions on agar plates with chloramphenicol (the plasmid's resistance marker) and no ampicillin. Subsequently, the library was selected on varying concentrations of ampicillin (0, 100, 250, 500, 1000, and 1500 μg/ml). Twenty-one of the $\sim 10^3$ surviving clones were sequenced. All sequenced variants contained the originally incorporated destabilizing mutations, and thus the presence of the false positives is assumed to be close to zero. All 13 stabilizing mutations included in the library were represented in the 21 surviving clones, at frequencies ranging from 100% (M182T) to 14% (H153R; Supplementary Table 1).

Characterization of TEM-1 variants

Protein purification, thermal denaturation, and enzymatic measurements were performed as previously described.[26,49] Temperature dependency was determined by measuring enzymatic activity at temperatures between 30 °C and 80 °C. TEM-1 variants (0.5 nM) were mixed with the colorimetric β-lactam Centa ($[S]_0$ = 100 μM ≫ K_M) in 200 mM potassium phosphate buffer, pH 7.0, and the increase in absorbance at 405 nm was monitored. Enzymes and substrates were pre-incubated for 5 min at each temperature. Rates of product formation with no enzyme were monitored at all temperatures and subtracted from the enzymatic reaction rates.

Crystallization

Concentrated solutions of TEM-1 variants (60 mg/ml) in 25 mM Tris, pH 8.4, and 100 mM NaCl were applied. Crystallization was performed using the hanging-drop vapor diffusion method. Equal volumes (0.5 μl) of protein and reservoir solutions were mixed, and the resulting drops were equilibrated at 293 K against a 400-μl reservoir solution made of 9% (wt/vol) polyethylene glycol (PEG) 8000, 100 mM 4-morpholineethanesulfonic acid, pH 6.2, and 200 mM $Ca(OAc)_2$ for v13, and 11% (wt/vol) PEG 8000, 100 mM 4-morpholineethanesulfonic acid buffer, pH 6.7, 200 mM $Ca(OAc)_2$, and 10 μM Zn for v13 carrying the E104K and G238S mutations. Drops were seeded with microcrystals to obtain high-quality diffracting crystals.

Data collection and refinement

Crystals were transferred into a cryo-protectant solution containing the reservoir solution and 25% PEG 600 for 1 min and then flash-cooled in liquid nitrogen. X-ray diffraction data for the crystal of v13 carrying G238S and E104K mutations (2.2 Å resolution) were collected at 100 K on a Rigaku R-AXIS IV++ imaging plate area detector mounted on a Rigaku RU-H3R generator with CuKα radiation focused by Osmic confocal mirrors. The v13 data set was collected at 100 K with synchrotron radiation at ID29-1 beam line (European Synchrotron Radiation Facility, Grenoble, France) using a Pilatus 6M detector. Data were integrated and scaled using XDS and XSCALE.[50] Molecular replacement for the v13 carrying G238S and E104K was performed using MOLREP[51] with the structure of wild-type TEM-1 [Protein Data Bank (PDB) code: 1ZG4] as the starting model. For v13, however, molecular replacement using this model failed, both with the predicted space group (*P*222) and with enantiomorphs (using MOLREP and Phaser[52]). The initial model was therefore obtained by performing molecular replacement with MOLREP in *P*1 space group using the wild-type structure as model. Initially, 8 molecules were placed (R_{work} = 39.3%, R_{free} = 44.4% after refinement) by MOLREP, and 7 additional (total = 15; R_{work} = 33.9%, R_{free} = 39.5% after refinement) by using Phaser with the previous molecular replacement output model as fixed. A second refinement cycle enabled the reconstruction of the complete unit cell in *P*1 (20 molecules; R_{work} = 27.9%, R_{free} = 32.9% after refinement). The unit cell reduction was performed using Zanuda§ to the space group $P2_12_12$ with 5 protein molecules per asymmetric unit. Manual model improvement was performed using Coot,[53] and the refinement was performed using REFMAC5.[51] A TLS refinement strategy was used for refining the v13 structure. Details of data collection and refinement statistics are provided in Supplementary Table 2. Figures depicting structures were prepared with PyMOL.

Normalized *B*-factors

The occupancies of all residues in all analyzed structures used were changed to 1, and the structures were re-refined using REFMAC5.[51] The per-residue *B*-factor was obtained by averaging the values for its main-chain atoms and normalizing to the average *B*-factor of the structure. TEM-1's scaffold residues and active-site residues were defined as described in the next section.

Folds polarity and innovability

We adopted the protein structure classification of CATH database v3.4‖[54] and defined folds as the ensemble of structures with the same class, architecture, and topology (the first three digits in CATH codes). We restricted our analysis to folds with only one annotated CATH topology (i.e., single-domain proteins) and counted all the distinct E.C.

numbers reported for enzymes with known structures belonging to each fold. We defined scaffold as the structural units that are preserved along evolution and shared by all proteins that belong to a given fold. Structures belonging to each fold were aligned with MUSTANG[55] and the conserved secondary-structure elements in the resulting structural alignment were identified. Helices, sheets, and other elements such as loops were assigned using the dssp program.[50] Scaffold positions were defined as positions with ≥70% conservation of secondary structure in the structural alignment (e.g., ≥70% of sequences have helix assigned to a given position). The 140 domains assigned to the LUCA[37] correspond to 87 folds. However, only 8 folds of the 87 were found to have enough representative examples of single-domain enzymes with high-resolution crystal structures (<2.5 Å) with bound substrates or other relevant ligands: TIM barrel (CATH code: 3.20.20), Rossmann (3.40.50), α–β plaits (3.30.70), aminopeptidase (3.40.630), farnesyl transferase (1.10.600), UDG (3.40.470), DHFR (3.40.430), and chitinase (3.10.50). Overall, our data set comprised 43 PDB structures representing 8 different folds and 28 different E.C. numbers (Supplementary Table 3; Supplementary Fig. 2). In case of non-evolvable folds, structures from different organisms were collected that represent up to ~80% sequence divergence (typically from bacteria to mammals). A subset of proteins ($N = 30$) that were crystallized with relevant ligands (cofactors, substrate, or transition-state analogs) were used to annotate the active-site residues (as residues having at least one atom within 4 Å from the bound ligand). Contact density was calculated using csu[56] to assign contacts per each scaffold and active-site residue.

Evolutionary rates

We collected all *S. cerevisiae* enzymes that belong to the analyzed LUCA folds, with different E.C. numbers and known structures including bound ligands (no example of α–β plaits were found; hence, only 7 out of the 8 folds were included; Supplementary Table 4). We calculated the positional evolutionary rates as described using orthologous sequences from 10 fungi species.[57] The average per-protein rates, and positional rates, were derived from alignments and trees of fungi orthologs, using the Bayesian rate estimation method of the Rate4-Site program, as previously described.[57,58] The positional rates indicate how fast a given position evolves relative to the average rate for the entire phylogeny across all sites.

Accession codes

Atomic coordinates and structure factors have been deposited with the following PDB accession codes: 4IBR and 4IBX.

Acknowledgements

The notion of local *versus* global compensation stemmed from an inspiring symposium at New England BioLabs, and following discussions with Joe Thornton. We are very grateful to Ruth Nussinov, Devin Trudeau, Evandro Ferrada, and Emmanuel Levy for valuable discussions and comments. We thank Prof. J. Pelletier for the TEM-1 expression plasmid. Financial support by the Israel Science Foundation and the Meil de Botton Fund is gratefully acknowledged. D.S.T. is the incumbent of the Nella and Leon Benoziyo Professorial Chair.

Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.03.033

Received 11 February 2013;
Received in revised form 18 March 2013;
Accepted 24 March 2013
Available online 28 March 2013

Keywords:
protein evolution;
enzyme evolution;
protein disorder;
protein folds

† E.D.-G. and A.T.-P. contributed equally to this work.
‡ http://www.tcoffee.org/
§ http://www.ysbl.york.ac.uk/YSBLPrograms/index.jsp
‖ www.cathdb.info

Abbreviations used:
DHFR, dihydrofolate reductase; LUCA, last universal common ancestor; PEG, polyethylene glycol; PDB, Protein Data Bank; UDG, uracyl-DNA glycosilase.

References

1. Bornberg-Bauer, E. & Chan, H. S. (1999). Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. *Proc. Natl Acad. Sci. USA*, **96**, 10689–10694.
2. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **6**, 678–687.
3. Shakhnovich, B. E., Deeds, E., Delisi, C. & Shakhnovich, E. (2005). Protein structure and evolutionary history determine sequence space topology. *Genome Res.* **15**, 385–392.
4. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*, **444**, 929–932.
5. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. (2006). Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA*, **103**, 5869–5874.
6. Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C. & Arnold, F. H. (2005). Thermodynamic prediction of protein neutrality. *Proc. Natl Acad. Sci. USA*, **102**, 606–611.

7. Wagner, A. (2005). *Robustness and Evolvability in Living Systems.* Princeton University Press, Princeton, NJ.
8. Wagner, A. (2011). *The Origins of Evolutionary Innovations.* Oxford University Press, Oxford.
9. Rorick, M. M. & Wagner, G. P. (2011). Protein structural modularity and robustness are associated with evolvability. *Genome Biol. Evol.* **3**, 456–475.
10. Wang, X., Minasov, G. & Shoichet, B. K. (2002). Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85–95.
11. Marciano, D. C., Pennington, J. M., Wang, X., Wang, J., Chen, Y., Thomas, V. L. *et al.* (2008). Genetic and structural characterization of an L201P global suppressor substitution in TEM-1 beta-lactamase. *J. Mol. Biol.* **384**, 151–164.
12. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. (2008). How protein stability and new functions trade off. *PLoS Comput Biol*, **4**, e1000002.
13. Vihinen, M. (1987). Relationship of protein flexibility to thermostability. *Protein Eng.* **1**, 477–480.
14. Reetz, M. T., Carballeira, J. D. & Vogel, A. (2006). Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed. Engl.* **45**, 7745–7751.
15. Tokuriki, N. & Tawfik, D. S. (2009). Protein dynamism and evolvability. *Science*, **324**, 203–207.
16. Wroe, R., Chan, H. S. & Bornberg-Bauer, E. (2007). A structural model of latent evolutionary potentials underlying neutral networks in proteins. *HFSP J.* **1**, 79–87.
17. Chen, Y., Delmas, J., Sirot, J., Shoichet, B. & Bonnet, R. (2005). Atomic resolution structures of CTX-M beta-lactamases: extended spectrum activities from increased mobility and decreased stability. *J. Mol. Biol.* **348**, 349–362.
18. Tomatis, P. E., Fabiane, S. M., Simona, F., Carloni, P., Sutton, B. J. & Vila, A. J. (2008). Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. *Proc. Natl Acad. Sci. USA*, **105**, 20605–20610.
19. Maveyraud, L., Saves, I., Burlet-Schiltz, O., Swaren, P., Masson, J. M., Delaire, M. *et al.* (1996). Structural basis of extended spectrum TEM beta-lactamases. Crystallographic, kinetic, and mass spectrometric investigations of enzyme mutants. *J. Biol. Chem.* **271**, 10482–10489.
20. Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. & Stevens, R. C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat. Struct. Biol.* **8**, 238–242.
21. De Wals, P. Y., Doucet, N. & Pelletier, J. N. (2009). High tolerance to simultaneous active-site mutations in TEM-1 beta-lactamase: distinct mutational paths provide more generalized beta-lactam recognition. *Protein Sci.* **18**, 147–160.
22. Foit, L., Morgan, G. J., Kern, M. J., Steimer, L. R., von Hacht, A. A., Titchmarsh, J. *et al.* (2009). Optimizing protein stability in vivo. *Mol. Cell*, **36**, 861–871.
23. Reetz, M. T. & Carballeira, J. D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903.
24. Arnold, F. H., Wintrode, P. L., Miyazaki, K. & Gershenson, A. (2001). How enzymes adapt: lessons from directed evolution. *Trends Biochem. Sci.* **26**, 100–106.
25. Carroll, S. M., Ortlund, E. A. & Thornton, J. W. (2011). Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS Genet.* **7**, e1002117.
26. Bershtein, S., Goldin, K. & Tawfik, D. S. (2008). Intense neutral drifts yield robust and evolvable consensus proteins. *J. Mol. Biol.* **379**, 1029–1044.
27. Steipe, B., Schiller, B., Pluckthun, A. & Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240**, 188–192.
28. Watanabe, K., Ohkuri, T., Yokobori, S. & Yamagishi, A. (2006). Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J. Mol. Biol.* **355**, 664–674.
29. Hecky, J. & Muller, K. M. (2005). Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of beta-lactamase. *Biochemistry*, **44**, 12640–12654.
30. Bershtein, S. (2007). Understanding the underlining mechanism of random genetic drift by experimental evolution. Ph.D. thesis, Weizmann Institute of Science.
31. Kather, I., Jakob, R. P., Dobbek, H. & Schmid, F. X. (2008). Increased folding stability of TEM-1 beta-lactamase by in vitro selection. *J. Mol. Biol.* **383**, 238–251.
32. Salverda, M. L., de Visser, J. A. & Barlow, M. (2010). Natural evolution of TEM-1 beta-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol. Rev.* **34**, 1015–1036.
33. Cantu, C., 3rd & Palzkill, T. (1998). The role of residue 238 of TEM-1 beta-lactamase in the hydrolysis of extended-spectrum antibiotics. *J. Biol. Chem.* **273**, 26603–26609.
34. Petit, A., Maveyraud, L., Lenfant, F., Samama, J. P., Labia, R. & Masson, J. M. (1995). Multiple substitutions at position 104 of beta-lactamase TEM-1: assessing the role of this residue in substrate specificity. *Biochem. J.* **305**, 33–40.
35. Parthasarathy, S. & Murthy, M. R. (1997). Analysis of temperature factor distribution in high-resolution protein structures. *Protein Sci.* **6**, 2561–2567.
36. Yuan, Z., Bailey, T. L. & Teasdale, R. D. (2005). Prediction of protein *B*-factor profiles. *Proteins*, **58**, 905–912.
37. Ranea, J. A., Sillero, A., Thornton, J. M. & Orengo, C. A. (2006). Protein superfamily evolution and the last universal common ancestor (LUCA). *J. Mol. Evol.* **63**, 513–525.
38. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. (2006). Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol.* **361**, 1003–1034.
39. Khersonsky, O. & Tawfik, D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471–505.

40. Amitai, G., Gupta, R. D. & Tawfik, D. S. (2007). Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* **1**, 67–78.
41. Kirschner, M. & Gerhart, J. (1998). Evolvability. *Proc. Natl Acad. Sci. USA*, **95**, 8420–8427.
42. Ferrada, E. & Wagner, A. (2008). Protein robustness promotes evolutionary innovations on large evolutionary time-scales. *Proc. Biol. Sci.* **275**, 1595–1602.
43. Bloom, J. D., Drummond, D. A., Arnold, F. H. & Wilke, C. O. (2006). Structural determinants of the rate of protein evolution in yeast. *Mol. Biol. Evol.* **23**, 1751–1761.
44. England, J. L. & Shakhnovich, E. I. (2003). Structural determinant of protein designability. *Phys. Rev. Lett.* **90**, 218101.
45. Zhang, C. & DeLisi, C. (2001). Protein folds: molecular systematics in three dimensions. *Cell. Mol. Life Sci.* **58**, 72–79.
46. Jeong, S. H., Bae, I. K., Lee, J. H., Sohn, S. G., Kang, G. H., Jeon, G. J. *et al.* (2004). Molecular characterization of extended-spectrum beta-lactamases produced by clinical isolates of *Klebsiella pneumoniae* and *Escherichia coli* from a Korean nationwide survey. *J. Clin. Microbiol.* **42**, 2902–2906.
47. Herman, A. & Tawfik, D. S. (2007). Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20**, 219–226.
48. Petrosino, J. F., Baker, M. & Palzkill, T. (1999). Susceptibility of beta-lactamase to core amino acid substitutions. *Protein Eng.* **12**, 761–769.
49. Salverda, M. L., Dellus, E., Gorter, F. A., Debets, A. J., van der Oost, J., Hoekstra, R. F. *et al.* (2011). Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **7**, e1001321.
50. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
51. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**. (1994)., 760–763.
52. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
53. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
54. Cuff, A. L., Sillitoe, I., Lewis, T., Clegg, A. B., Rentzsch, R., Furnham, N. *et al.* (2011). Extending CATH: increasing coverage of the protein structure universe and linking structure with function. *Nucleic Acids Res.* **39**, D420–D426.
55. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. (2006). MUSTANG: a multiple structural alignment algorithm. *Proteins*, **64**, 559–574.
56. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. (1999). Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15**, 327–332.
57. Toth-Petroczy, A. & Tawfik, D. S. (2011). Slow protein evolutionary rates are dictated by surface-core association. *Proc. Natl Acad. Sci. USA*, **108**, 11151–11156.
58. Mayrose, I., Graur, D., Ben-Tal, N. & Pupko, T. (2004). Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol. Biol. Evol.* **21**, 1781–1791.