 doi:10.20944/preprints202203.0246.v1

Article

Local and Global Protein Interactions Contribute to Residue Entrenchment in Beta-Lactamase TEM-1

André Birgy[1,2], Mélanie Magnan[1], Claire Amaris Hobson[1], Matteo Figliuzzi[3], Karine Panigoni[1], Cyrielle Codde[1], Olivier Tenaillon[1,#*] and Hervé Jacquier[1,4,#*]

1. IAME, UMR 1137, INSERM, Université de Paris, Paris, France.
2. Service de Microbiologie, Hôpital Robert-Debré, AP-HP, Paris, France.
3. UPMC, Institut de Calcul et de la Simulation, Sorbonne Universités, Paris, France Computational and Quantitative Biology, UPMC, UMR 7238, Sorbonne Universités, Paris, France Computational and Quantitative Biology, CNRS, UMR 7238, Paris, France.
4. Service de Bactériologie-Hygiène, Groupe Hospitalier Saint-Louis – Lariboisiére - Fernand Widal, AP-HP, Paris, France.
#These authors contributed equally to this work.
* Correspondence: OT Olivier.tenaillon@inserm.fr ; HJ herve.jacquier@aphp.fr

Abstract: Due to their rapid evolution and their impact on health care, beta-lactamases, protein degrading beta-lactam antibiotics, are used as generic model of protein evolution. Therefore, we investigated the mutation effects in two distant beta-lactamases TEM-1 and CTX-M-15. Interestingly we found a site with a complex pattern of genetic interactions. Mutation G251W in TEM-1 is inactivating the protein's function just as the reciprocal mutation W251G in CTXM-15. Phylogenetic analysis revealed that mutation G has been entrenched in TEM-1 background: while rarely observed throughout the phylogeny it is essential in TEM-1. Using a rescue experiment in TEM-1 G251W mutant, we could identify sites that alleviates the deviation from G to W. While few of these mutations could potentially involve local interactions, most of them were found on distant residues in the 3D structure. Many well-known mutations having an impact on protein stability, such as M182T, were recovered. Our results therefore suggest that entrenchment of an amino acid may rely on diffuse interactions among multiple sites with a major impact on protein stability.

Keywords: entrenchment; protein stability; TEM-1 beta-lactamase; CTX-M-15 beta-lactamase; M182T mutation

1. Introduction

Understanding the constraints that shape protein evolution is not only a matter of basic science but also of public health. In the field of medical microbiology, some determinants of resistance to anti-infective agents slowly evolved during millions of years. However, some mutations have been selected under antimicrobial selective pressure during these last decades. The diversification of beta-lactamases, which confer resistance to beta-lactams antibiotics in bacteria, or of proteases, targeted by anti-protease chemical used in HIV treatment [1], lead not only to an increased complexity of medical care but even to therapeutic impasses[2,3] both of which can result in fatal issues.

The chemical and physical properties of a protein, defined by the amino acids' interactions in three-dimensional space, determine its function. Currently, most of the set of environmental constraints acting on

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

these residues remain unknown, and their impact on the protein's functions is complex. Analysis comparing homologous protein sequences enables to better identify the constraints acting on the protein's residues. A conserved residue over a long evolutionary time period signs a strong constraint. Historically, amino acids substitutions were deduced from empirical models comparing protein sequences [4,5]. The hypothesis that evolution was identical and independent at all locations is the foundation of phenomenological substitution models. However, it is now acknowledged that evolution depends on interactions between residues and that the independence of each site cannot be affirmed [6,7]. This observation therefore leads to the notion of epistasis that is to say of interactions between mutations: the effect of two mutations combined may be different from that of the sum of the mutations taken separately. Hence, mutation effects may differ according to the genetic background where they emerge[6,8], in other words their effect can be context dependent. These genetic interactions, termed epistasis, have been deeply investigated in the literature, and the following vocabulary has been proposed. Pairs of mutations have magnitude epistasis if the combined mutations have a greater effect than that expected, or sign epistasis if at least one of the mutations shifts from deleterious to beneficial according to the presence of the other mutations. Interactions between mutations may be local, and termed local epistasis[9–12], or may operate through properties inherent of the whole protein, and be qualified as global epistasis[8,13,14] .

An extreme case of epistasis is referred to as compensated pathogenic deviation (CPD)[15]. Here, a pathogenic amino acid variant in one species is observed fixed in the orthologous protein (sharing the same function specificities) of another species. As an example, in *BBS4* gene (a gene involved in ciliopathies in humans), H165 mutation is highly conserved in different species, whereas N165 is found in Humans. Interestingly, N165H mutation is associated with ciliopathies in Humans, whereas this mutation should be expected as neutral [16]. Hence, this mutation has probably been associated with compensating coevolving residues, and would be deleterious in the absence of associated substitutions. This hypothesis termed entrenchment suggests that proteins tend to balance the presence of an amino acid at a position, through replacements at other positions[17].

The beta-lactamases are a good model for protein evolution for the following reasons: (i) high diversification in the wild, selected or not by antimicrobial pressure, (ii) ease of genetic manipulation and phenotypic measurements in the laboratory, (iii) knowledge on thermodynamic and structural data. Among the beta-lactamases, TEM-1 is of particular interest because of its wide distribution in commensal and pathogenic bacteria and its use in evolutionary models, as mentioned above. CTX-M-15 is another interesting beta-lactamase. It has an extended-spectrum and has been very successful among of infectious agents over the past two decades. Its evolutionary characteristics have been however explored to a lesser extent than those of TEM-1 [18,19].

In our study, we analyzed sequence alignments of beta-lactamases, and performed large mutant libraries of TEM-1 and CTX-M-15, to search for the presence of entrenchment in the historical allele TEM-1.

2. Results

2.1. Distribution of Single Mutant MICs of two beta-lactamases reveals a single opposite effect

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

In search of context-dependency, we first screened the effects of mutations in the two orthologous beta-lactamases, TEM-1 and CTX-M-15. The distribution of mutation effects of CTX-M-15 was investigated by determining mutation effects of individual mutants recovered by random mutagenesis (see Methods). Of 2304 mutants, 404 single missense mutants were obtained for which the amoxicillin (the main substrate of class A beta-lactamases) Minimum Inhibitory Concentration (MIC) was determined. We compared these results with the mutation effects of 5041 single missense TEM-1 mutants in Firnberg's previous work [20].

We aimed to find mutations, with marked context-dependency. We, therefore, focused on mutations resulting in a switch of a given residue to the one found in the alternative protein in the same site. We found 236 single-mutants of TEM-1 (Firnberg's study [20] and 25 single-mutants of CTX-M-15 (this study, sanger data), that mutated towards an equivalent amino acid in CTX-M-15, and in TEM-1, respectively (Figure 1). Among all these mutations, we noticed one with a strong signature of context dependency. On residue 251, TEM-1 has a glycine (G), and CTX-M-15 has a tryptophan (W). When switching G by W, in TEM-1, or inversely W by G in CTX-M-15, the proteins lose their function (Figure 1). Interestingly, the mutation behaves as a CPD in the two backgrounds. This observation raises the question of how can one sequence evolve from one to the other.

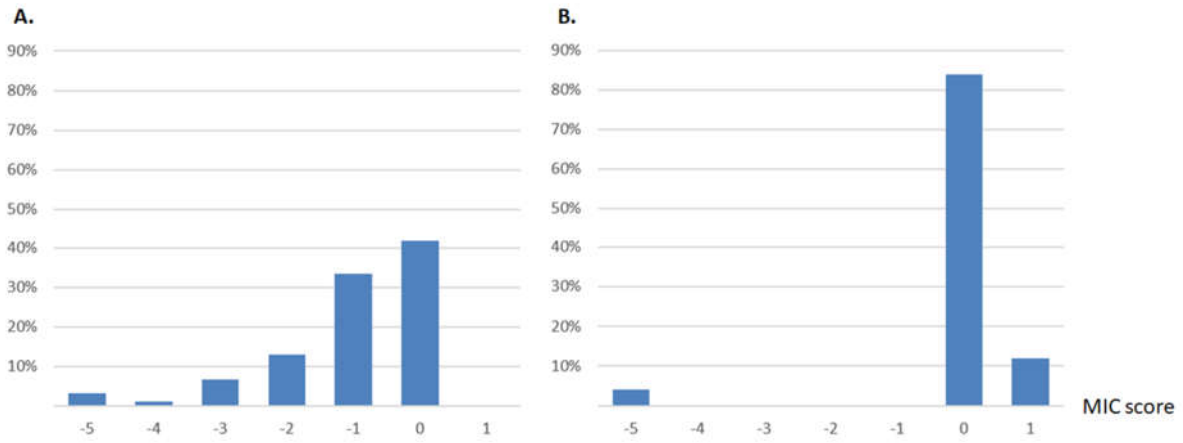


Figure 1. A. Effects of mutations on MIC score, when changing TEM-1 residue with that of CTX-M-15 (n=236, dataset from Firnberg *et al.* [20]). B. Effects of mutations on MIC score, when changing CTX-M-15 residue with that of TEM-1 (n=24). MIC score corresponds to $\log_2(MIC_{mutant}/MIC_{WT})$. For each collection, one single mutant harbored an inactivating mutation (MIC score = -5): G251W in TEM-1 and W251G in CTX-M-15.

2.2. Phylogeny and local environment of site 251 highlights a potential entrenched site

To better understand the tolerance of site 251 to mutations, and its conservation through the phylogeny of beta-lactamases, we used an alignment of 157 class A beta-lactamases [21]. Most beta-lactamases, such as CTX-M-15, harbor a tryptophan at this site (n=130/157), a tyrosine (n=8/157), a glycine (n=8/157), a glutamine (3/157), a lysine (n=2/157), an arginine (n=1/157), a threonine (n=1/157), or an isoleucine (n=1/157) (Figure 2). Beta-lactamases containing G251 belong to a monophyletic group of 8 enzymes: TEM-1, OKP-A, PLA-1, OKP-B, LEN-1, SHV-1, ORN-1, and LAP-1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

The deleterious effect of W251G in CTX-M-15 can therefore be suggested by phylogenetic data following a simple independent model. Indeed, the effect of a mutation at a given position can be predicted by the frequency of this amino acid at this same position. Based on the tree (Figure 2), mutations from W towards G are rare and, thus, more likely to be deleterious. Conversely, with high conservation of Tryptophan at position 251, the mutation G251W should not be deleterious in TEM-1. This observation suggests high epistatic interactions with the residue 251 and that G251 could have evolved as a potential entrenched site.

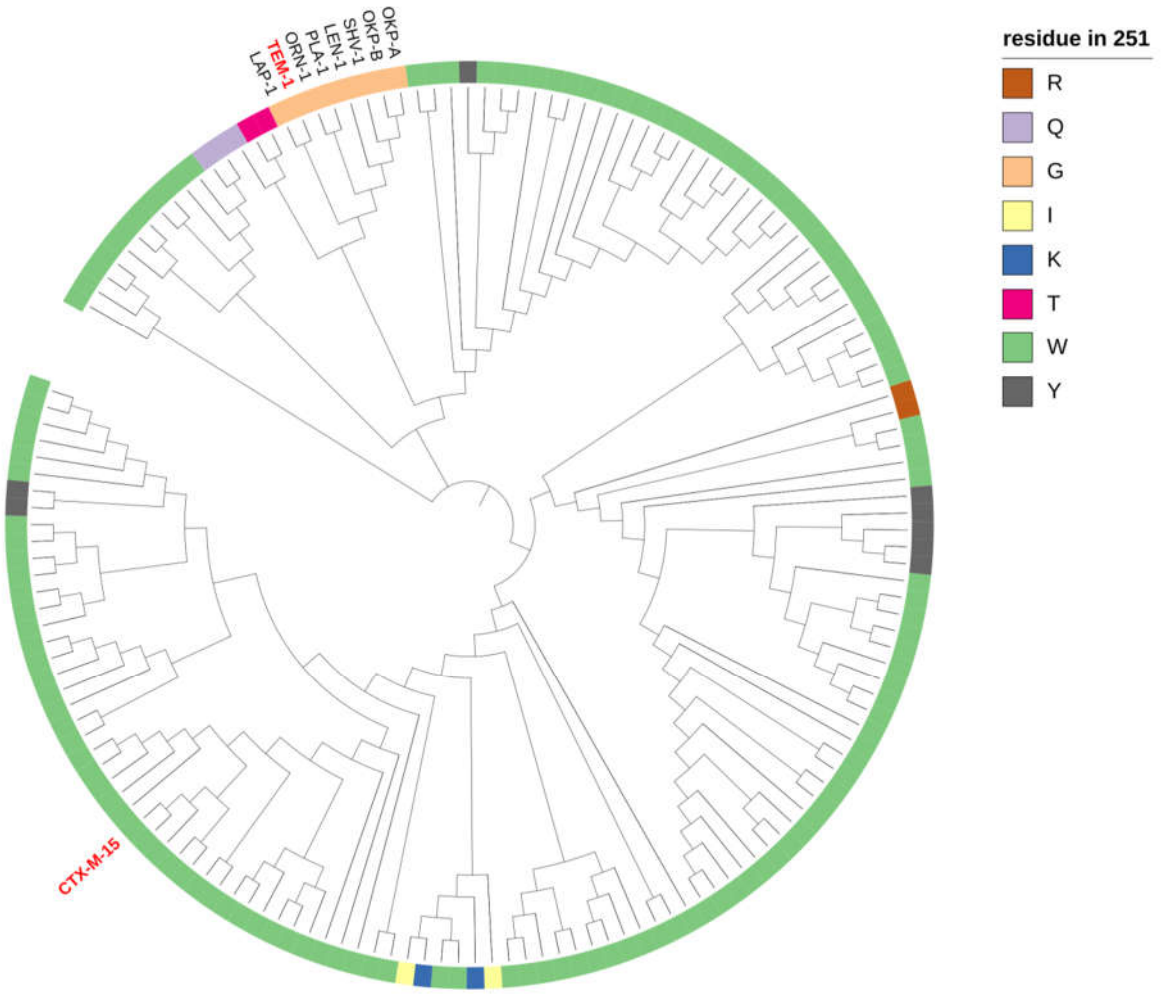


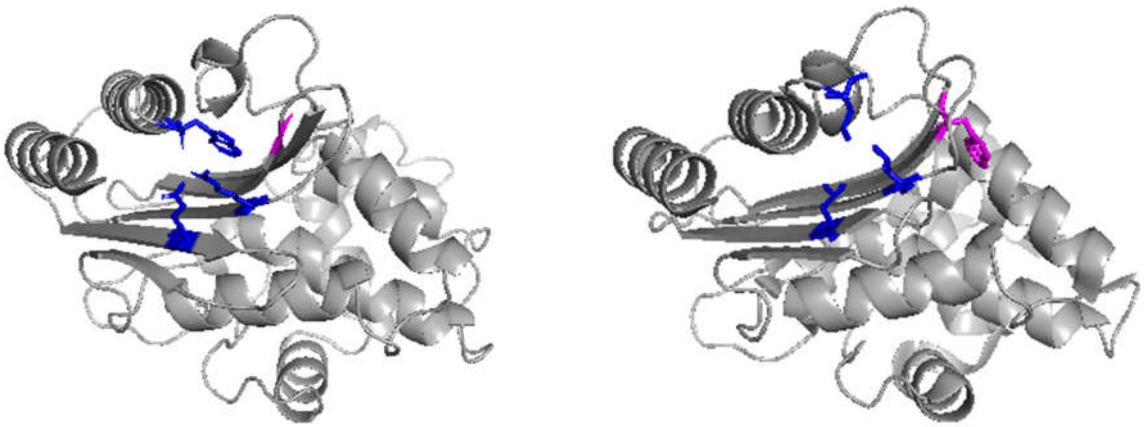
Figure 2. Phylogenetic tree of 157 class A beta-lactamases [21]. The different colors indicate the diversity at position 251. Beta-lactamases with a G251 are labeled, and TEM-1 and CTX-M-15 beta-lactamases are labeled in red.

2.3. Analysis of conserved sites suggests local co-evolution

To go further, we searched for sites that could have co-evolved with site 251 by targeting other amino acids exclusively present in the 8 beta-lactamases with a G251. We found 3 other sites: E48, R259, and W290. Interestingly, these 3 residues are close in the protein (Figure 3, Supplementary Table 2). This could suggest local epistasis compensating deleterious effect of glycine in 251. As observed in experimental data and suggested by phylogeny analysis [20,21], these 3 sites are more tolerant to mutations than G251. Yet, when replacing the amino acids with the ones of CTX-M-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

15 at the same site, in these four sites, we observe a moderate decrease in fitness (Table 1). Hence this cluster of sites may mutually contribute to the entrenchment of site 251.

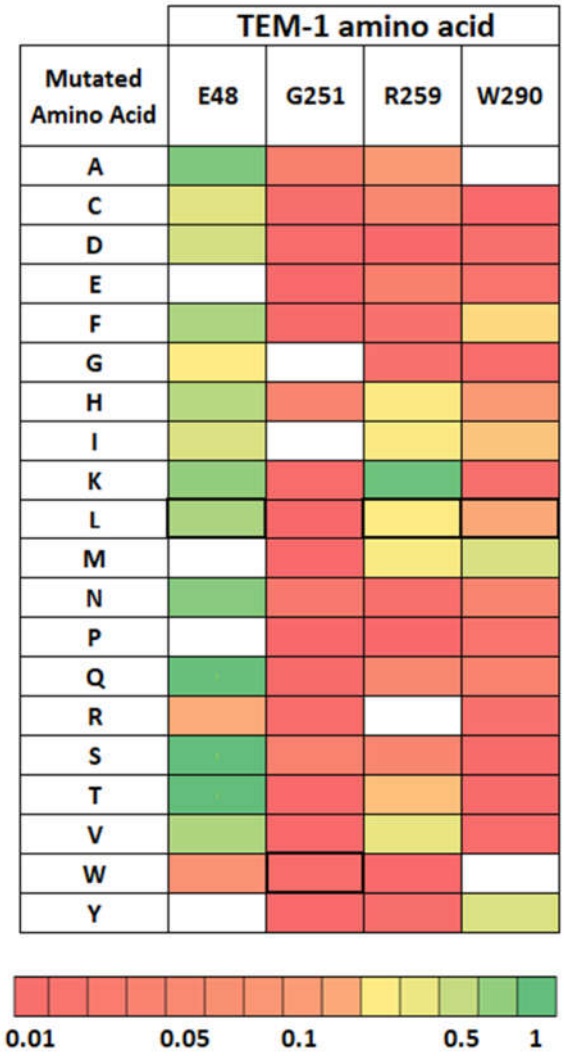


(**a**) (**b**)

Figure 3. A. 3D structure of TEM-1 (pdb entry 1BTL). The site G251 is shown in pink and the 3 residues exclusively associated with G251 (E48, R259, and W290) are indicated in blue. B. 3D structure of CTX-M-15 (pdb entry 4HBT). The site W251 is shown in pink the 3 residues associated with W251 (L48, L259, and L290) are indicated in blue.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

Table 1. Selective mutational effect of mutations in the 5 amino acids found exclusively in the G251 group (n=8) (E48, G251, R259, and W290). Fitness was obtained using Firnberg's data on TEM-1 [20]. A fitness of 1 (green) corresponds to the fitness of the wild-type allele (TEM-1). A fitness of 0 (red) corresponds to a minimum fitness. The color gradient from green to red corresponds to the fitness gradient. The boxed mutations correspond to amino acids present in CTX-M-15.

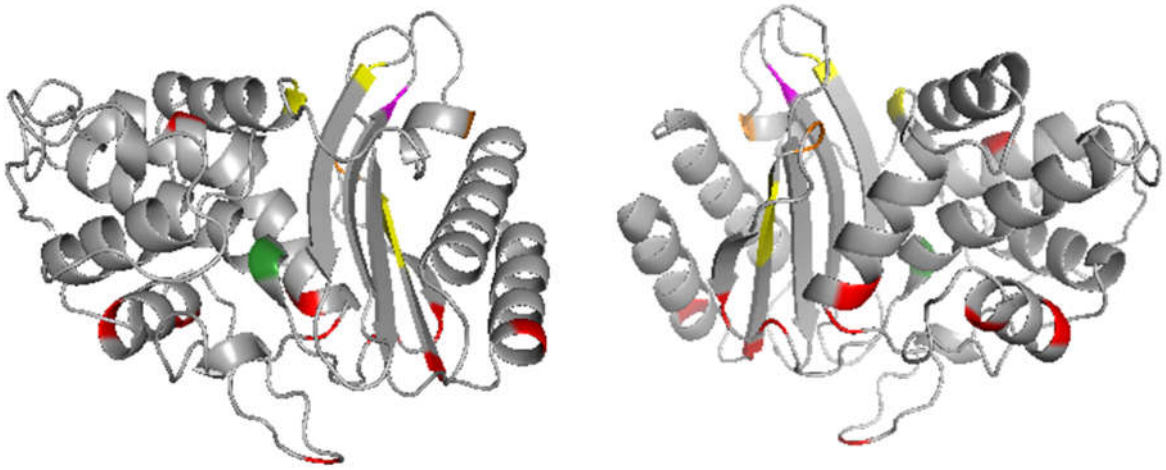


2.4. Distribution of compensating mutations in TEM G251W displays a various pattern of epistatic interactions

To unravel the sites contributing to the CPD state of Glycine at position 251 in TEM-1, we explored the distribution of the compensating mutations in the inactivated mutant G251W after a quick selection on amoxicillin. We performed a comprehensive mutagenesis using a pool of mutagenic primers containing a degenerated codon NNS for every site of the protein. To avoid reversion to the wild-type protein, we did not use the mutagenic primers targeting site 251. (See methods).

After filtering insertions, deletions and low-quality sequences, 13 millions of reliable reads were retained. We defined a cut-off of a 2-fold increase to define mutation as compensatory (see methods) and found 50 compensatory mutations, corresponding to 0.89 % (50/5869) of all possible mutations (Supplementary Table 2). The 50 compensatory mutations we identified were distributed among 23 different sites, including known compensating hotspots (Figure 4, Supplementary Text).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1



(**a**) (**b**)

Figure 4. Two opposite faces of the protein are shown (figures (a) and (b)). Sites with compensating mutations on G251W-TEM variant (pdb entry: 1 BTL). The active site (S70) is indicated in green and site 251 in cyan. Mutations with local effects (<13 Å from W251) are indicated in yellow and those with a global effect in red. Mutations located <13 Å from W251 but previously described with a global stabilizing effect on the protein are indicated in orange.

2.4.1. Global versus local epistasis

Based on the 3D structure of the protein, the mutations were globally distributed on the whole protein (Figure 4), including in some sites close to site 251, with probable local effects (Figure 4). Considering local interactions (<13 Å), we identified compensating mutations at sites 47-49, 51, 52, 212, 224, and 230. However, a general effect on protein stability could also be at play. Indeed, some local amino acids at 51 [22,23], 52 [24], and 224 [25,26] sites are also known to increase the protein's stability (also see Supplementary text).

2.4.2. Mutations with impact on global stability of the protein

In a more global way, many compensating sites could have an effect on protein stability. Residues 51, 52, 63, 147, 153, 182, 224, and 275 are involved, or supposed to be, in the global stability of the protein [8,13,19,22–27], while the role of other residues (31, 38, 42, 175 and 184) on stabilization properties is less clear but suggested[19,24,28] (also see Supplementary text).

Interestingly, this approach allowed us to display original patterns, described here for the first time. For example, the M182T mutation has been largely described as a high effect compensating mutation impacting stability. Here, we highlighted other mutations at this site, such as M182C, M182F, M182I, M182L, and M182S. This diversity of compensating mutations at a remote site underlies the complexity of compensation and the gain of such comprehensive and quantitative approaches.

2.4.3. Recovery mutations in TEM-1 and consensus sequences of beta-lactamases

We then looked at the residues for which the amino acid found in TEM-1 is an exception compared to the other beta-lactamases. We found 5 residues: V31, N52, E58, H153, and M182. Interestingly, within these 5 residues, 4 were selected (80%) in our rescue study after selection with

 doi:10.20944/preprints202203.0246.v1

antibiotics (31, 52, 153, and 182). Concerning residue 31, 5 substitutions were selected (A, Q, M, R, and K). All these amino acids were frequent at site 31 in the alignment of 157 beta-lactamases we used, except M31. Concerning residues 153 and 182 that are particularly conserved within beta-lactamases, some substitutions extremely frequent in the alignment of beta-lactamases were selected in our experiments (notably H153R, M182T and M182S). This suggests that rescue possibilities are in the direction of beta-lactamase's consensus. This shows that the evolution towards TEM-1 is the source of the epistasis pattern observed.

2.4.4. Steps of entrenchment of the residue G251 in TEM-1

On the basis of sequence alignments and of our experiments, we purpose how G251 could have been entrenched in TEM-1 (Figure 5). We particularly focused on sites 48, 259, and 290 (which co-evolved with G251) and on sites 153 and 182. As mentioned above, these latter are particularly of interest because H153 and M182 are only present in TEM-1, and because R153 and T/S182 are highly represented in class A beta-lactamases and described as stabilizing mutations in TEM-1.

The consensus sequence obtained with the alignment of 157 beta-lactamases associates small or hydrophobic amino acids (Valine, Alanine, Leucine, or Isoleucine) in sites 48, 259 and 290, W251, and amino acids R153 and T182.

G251 is strictly present in the monophyletic group of 8 beta-lactamases close to TEM-1, where it is associated with E48, R159, and W290. From a structural point of view, these sites are close to each other. Moreover, we determined experimentally that G251 was incompatible in the CTX-M-15 background. Hence, our results strongly suggest an entrenchment of G251 due to coevolution with residues E48, R159, and W290.

TEM-1 is the only beta-lactamase that harbors H153 and M182. Moreover, we determined experimentally that W251 was incompatible in the TEM-1 background, but that H153R and M182T are compensatory mutations in the TEM-1 G251W variant. Altogether, these observations strongly suggest that TEM-1 derives from its ancestor by mutations R153H and T182M, which decreased the global stability of the enzyme, and prevents later on the transition towards W251.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

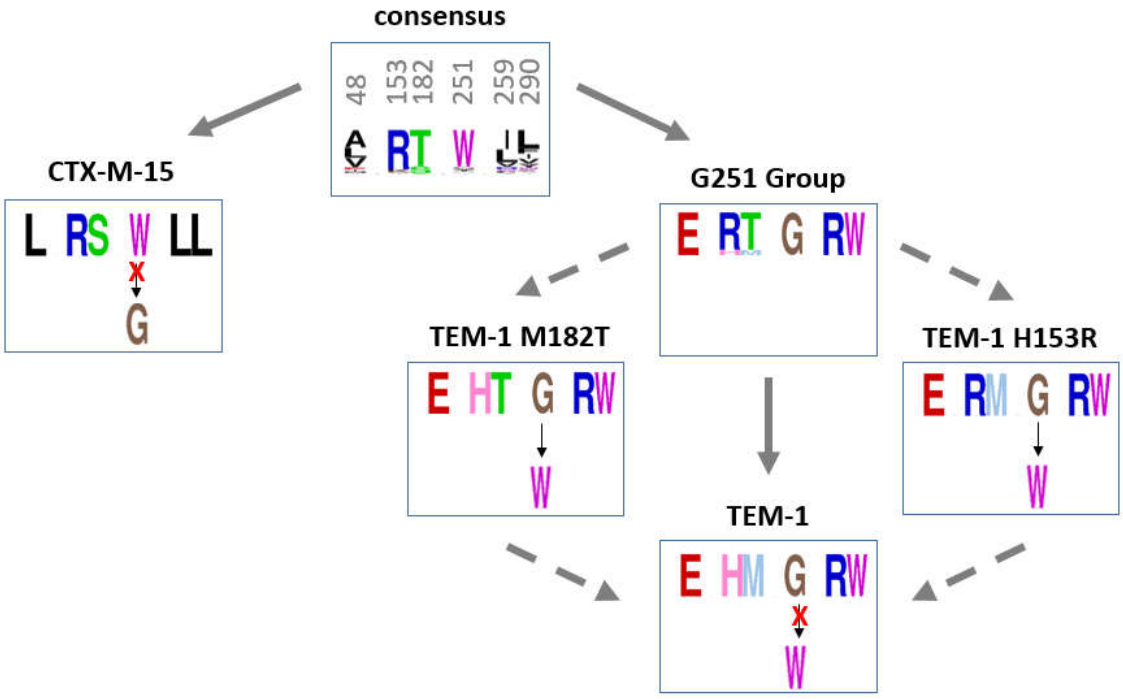


Figure 5. Steps of entrenchment of the residue G251 in TEM-1. The 3 sites (48, 259, and 290) that coevolve with G251, and two sites involved in compensating mutations (153 and 182) are represented beside site 251. The ability of each beta-lactamase to tolerate W251 or G251 is represented when it has been validated experimentally.

3. Discussion

Here, we used beta-lactamases as a model to study the context-dependency of mutation effects and proposed different steps leading to the entrenchment of a position in a protein. Looking at mutant libraries of TEM-1 and CTX-M-15, we found that switching residue 251 in the two backgrounds was highly deleterious in both of them. This observation suggests a complex evolutionary scenario to account for the evolution between these two sequences.

Focusing on that mutation, we then explored beta-lactamase multiple alignments and observed that W251 was highly represented and that G251 was limited to a monophyletic group of 8 beta-lactamases. In this group, G251 was associated with three residues that coevolved with this residue. To go further, we performed large libraries of mutants of TEM-1 G251W to investigate putative compensatory mutations. We identified a few mutations that were potentially involved in local interactions and many that were not physically close to the position of residue 251 in the 3D structure but associated with a stabilizing effect on TEM-1. The availability of several stabilizing mutations in the wild-type background of TEM-1 and the fact that TEM-1 alleles at these sites are not dominant suggests that TEM-1 has accumulated destabilizing mutations. Several scenarios could explain this pattern. Due to strong selection, the TEM-1 initial allele that spread through *E. coli* population may have been sampled from a succession of unlikely events and may have therefore accumulated by chance some destabilizing mutations. The presence of a destabilizing methionine at position 182 of TEM-1, a unique feature of all beta-lactamases could result from such bottlenecks. It could also be possible that due to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

stabilizing selection acting on protein stability, there is only a subset of the stabilizing mutations that are present at a given time in every beta-lactamase[29] . The presence of any destabilizing mutation is then making those stabilizing mutations beneficial. Finally, the diminishing return of stabilizing mutation effects on protein stability limits the selection for increased stability and makes the maintenance of all possible stabilizing mutations in the gene impossible [30]. In all cases, it seems that a common way to compensate for the mutation G251W is to improve protein stability.

The fact that globally stabilizing mutations are recruited to compensate for the cost of G251W mutation suggests that it affects the protein mostly through stability. In the case of flu virus, it has been shown that destabilizing mutations could be fixed if at the time of their appearance the protein was highly stable[7]. This scenario does not however seem to be sufficient to explain the mutation presence of a G in TEM-1 in our case as the mutation from G to W is also deleterious. Our data suggest that the G in TEM-1 has been entrenched such that its presence is tolerated in TEM-1 but not in distant relatives. Many other changes along the protein sequence could contribute to that entrenchment, but our compensatory approach found mostly distant mutations involved particularly in protein stability. Interestingly, none of the amino acids that are specifically associated with the presence of a G at position 251 were found to be involved in the compensation despite their proximity to the residue and their own signature of entrenchment as reversion of several of these were deleterious. This suggests that no single local interaction is responsible for the entrenchment of the G at position 251, and that the entrenchment may rely on diffuse interactions among multiple sites.

Overall, even though our data are not exhaustive, they reveal some complex patterns of protein evolution. While local interactions allow presumably an amino acid to be entrenched, after long enough, the reversion to the native state is mostly compensated by mutations having global effects on protein stability. Epistatic interactions and protein evolution may therefore be driven by both local and global effects.

4. Materials and Methods

Distribution of mutation effects of CTX-M-15

CTX-M-15 library Construction

CTX-M-15 mutants were constructed using GeneMorph II Random Mutagenesis Kit (Stratagene) to obtain an average of one mutation per gene, as previously described[8]. Briefly, The mutagenized amplicons were cloned into a modified pUC19 plasmid containing the pMB1 origin of replication from pBR322, *Nco*I and *Not*I flanking the start and stop codons of CTX-M-15 ORF, and gentamicin resistance gene (*aac*C4) at the *Xba*I site. The ligation products were transformed into ElectroMax DH10B-T1 Phage Resistant *E. coli* Competent Cells (Invitrogen, Fisher Scientific) and plated on Luria–Bertani agar supplemented with gentamicin (20 mg/L). A total of 2304 randomly picked CTX-M-15 mutants were stored into 384-well microplates and sequenced by Sanger method.

MIC Measurements

The MIC was measured by a standard agar dilution method on Mueller Hinton (MH) agar plates containing a growing concentration of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

amoxicillin (0, 250, 500, 1000, 2000, 4000, 8,000, and 16,000 mg/L). After 18 h of incubation at 37 °C, the MIC was defined as the first concentration of amoxicillin inhibiting the growth of bacteria.

MIC Score

For each mutant, MIC was computed as the median of three independent MIC measurements. MIC score is computed as $\log_2$(MIC). For amino acid changes that were found several times in the library as single amino acid changes, the average MIC score was retained.

Library of compensating mutations in G251W TEM variants

Constructs

Strains: *E. coli* strains used in this study:

XL1-Blue (Agilent, Santa Clara, CA) recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacI1qZΔM15 Tn10 (Tetr)];

CJ236 (new England biolabs) FΔ(HindIII)::cat (Tra+ Pil+ CamR)/ ung-1 relA1 dut-1 thi-1 spoT1 mcrA;

DH5α (Invitrogen) F– Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1

Plasmids: Phagemid pSkunk3-TEM-1 was obtained graciously from Elad Firnberg and Marc Ostermeier. Phagemids pSkunk3-TEM-G251W (containing the G251W) was created by mutagenesis from pSkunk3-TEM-1 by single step Pfunkel mutagenesis [31]. (Primers used are listed in supplementary table 3)

Pfunkel Mutagenesis

ssDNA production

Uracil-containing single strand DNA of pSkunk3-TEM-G251W was produced as published by Firnberg et al.[20] and Kowalsky et al.[32] (except the final centrifugation step which was performed at 26200 xG for 1h at 4°C. DNA was quantified using the Qubit® ssDNA Assay Kit (ThermoFisher Scientific).

Single step Pfunkel mutagenesis

Mutagenesis was done as previously described with 1 µg of ssDNA used as matrix in a total volume of 100 µl. The only difference is the elongation step which was of 15 min. So, the reaction cycling conditions were 95°C for 3 min, followed by 55°C for 90 sec, 68°C for 15 min and 45°C for 15 min[32]. We used the innuPREP PCRpure Kit (analytik jena) to purify DNA and eluted in 15 µl of distilled DNAse/RNAse free water. 2 µl were then electroporated in 20 µl of DH5alpha electrocompetent cells and then incubated with 500µl of LB media for 1 hour at 37°C with shaking at 250 rpm. The transformation was plated on LB agar with 50 µg/ml streptomycin and then incubated overnight at 37°C.

PCR verification and Sanger sequencing were done on isolated colonies using primers TEM-pSKUNK-DIM-F and TEM-pSKUNK-DIM-R (Supplementary table 3) and mutants were stocked in LB-glycerol 40% after an overnight culture at 37°C in LB media containing 50 µg/ml streptomycin.

Comprehensive mutagenesis pfunkel

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

Primers containing an NNS degenerated codon (N is either A, T, G or C ; S is either G or C) were designed using PfunkelMatlab algorithm for comprehensive codon mutagenesis as detailled in Firnberg et al. [20] (Supplementary table 3). One oligo is designed for each codon position replacing the codon by NNS. TEM gene was virtually tiled in 4 equal parts of 70-80 AA. Mutagenic primers were pooled according to this division. A separate Pfunkel reaction was performed for each region. Protocols were performed as published by Kowalsky et *al.*[32] and Firnberg et *al*. [20]. Purification was carried out using the innuPREP PCR pure Kit (analytik jena) and DNA was eluted in 15µl of distilled water. Two µl were electroporated in 20 µl of DH5alpha electrocompetent cells and then incubated with 500 µl of LB media for 1 hour at 37°C with shaking at 250 rpm. The transformation was plated on LB agar with 50 µg/ml streptomycin and then incubated overnight at 37°C. A pool of about 50 000 colonies per tiles were scrapped from LB agar plates (245 mm X 245 mm, Greiner bio-one) in LB broth and frozen at -80°C in LB/glycerol 40%. After pooling all colonies together, plasmids were extracted from aliquot using plasmid miniprep (Qiagen, Valencia, CA).

Selection experiments

TEM G251W: For each regions, 1ml of the frozen libraries cells stocks was cultured in MH broth at 37°C with shaking at 250 rpm from to OD_{600} 0.4 without antibiotics (named T0). Then, 3.2 ml of this first culture were used to re-inoculate 96.8 ml MH broth supplemented with 0.25 g/l of amoxicillin until OD_{600} 0.2. Half of these cultures was pelleted for freezing at -80°C and half was washed 2 consecutive times with sterile physiological serum and growth overnight in fresh MH medium without antibiotics. Finally, half of this overnight culture was used for plasmid extraction and half was pelleted and re-suspended in LB-glycerol 40% for freezing at -80°C.

Library Preparation and Deep Sequencing

Deep sequencing was used to obtain count data of each variant in the population after selection with amoxicillin on a MiSeq using V3 kit 2x300 paired-end (Illumina technology). A two-step PCR method was used to amplify independently the corresponding part of the gene and to add the lllumina sequencing adaptor and barcode sequences. In details, plasmid DNA concentration was determined using qubit fluorometric quantification (ThermoFisher scientific) and normalized to 2.5 ng/µl. 12.5 ng of DNA were used for the 1st PCR using specific primers (depending on the tile of the gene) and allowed the attachment of an adaptor that is necessary for the 2nd PCR. Between the specific primer and the adaptor, a variable number of degenerated nucleotides (0, 1, 2 or 3 for TEM primers) were inserted in order to increase the diversity of DNA during the sequencing part on the MiSeq and to improve crosstalk and phasing calculations (supplementary table 3).

Kapa Hifi Hotstart Ready Mix PCR Kit polymerase (Kapa Biosystems) was used for amplifications. The reaction cycling conditions were 95°C for 30 sec, followed by 12 cycles of 95°C for 10 sec, 55°C for 30 sec, 68°C for 30 sec and a final extension at 68°C for 5 min.

After gel purification using Qiagen gel extraction kit (Valencia, CA), DNA was quantified using Qubit Fluorometric quantification and DNA concentration was normalized. The 2nd PCR was performed using 5 ng of DNA using primers commercialized by Illumina in the Nextera Index Kit

allowing the dual indexing and the addition of the remainder of the sequencing primer annealing site along with the annealing site for the Illumina flow cell. The reaction cycling conditions were the same as previously but only 11 cycles were performed using Kapa Hifi Hotstart Ready Mix PCR Kit polymerase (Kapa Biosystems).

After gel purification with Qiagen gel extraction kit (Valencia, CA), quantification using qPCR kappa Hifi Hotstart (Kapa Biosystems) on a Light cycler 480 Roche was performed with reaction cycling conditions of 95°C for 5 min, followed by 35 cycles of 95°C for 30 sec and 60°C for 45 sec as specified by Kapa Biosystems. All libraries corresponding to different regions of the amoxicillin concentration selection were finally pooled in equal quantity and 12 pM were loaded on the MiSeq with a mix of 10% PhiX DNA (PhiX Control v3, Illumina) as sequencing control.

Sequence analysis: Mutations were detected on the reads and their frequency computed relative to that of wild-type sequences for each region of the gene. Technically the wild-type sequence should have mostly disappeared due to the treatment, but we found out that due to off-site compensatory mutations, most alleles could survive the treatment as they harbored a fraction of mutants resisting the treatment. However, as the off-set mutation rate was high enough to produce a reliable compensation for most mutants, the survival of alleles was supposed to be proportional to their real impact on survival. Alleles that increased in frequency by a two-fold factor were then considered as compensatory mutation.

Author Contributions: "Conceptualization, AB, OT and HJ; methodology, OT and HJ.; validation, AB, MF, MW, OT and HJ.; formal analysis, AB, MF, CC, HJ.; investigation, AB, MM, CAH, KP, CC and HJ.; resources, OT.; data curation, AB, MF, MM, OT and HJ.; writing—original draft preparation, AB and HJ.; writing—review and editing, AB, MM, CAH, MF, KP, CC, MW, OT and HJ.; visualization, AB, OT and HJ.; supervision, OT and HJ.; project administration, OT.; funding acquisition, OT. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the European Research Council (ERC) under the European Union's 7th Framework Program, ERC Grant 310944 (to O.T.), the French Agence Nationale pour la Recherche ANR GeWiEp (ANR-18-CE35-0005-01 to O.T.), the French Fondation pour la Recherche Medicale (EQU201903007848, to O.T.).

Acknowledgments: We thank Elad Firnberg and Marc Ostermeier for kindly providing Phagemid pSkunk3-TEM-1. We thank Timothy Whitehead and Caitlin Kowalsky for their advice and helpful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Flynn, W.F.; Haldane, A.; Torbett, B.E.; Levy, R.M. Inference of Epistatic Effects Leading to Entrenchment and Drug Resistance in HIV-1 Protease. *Mol Biol Evol* **2017**, *34*, 1291–1306, doi:10.1093/molbev/msx095.
2. de Visser, J.A.G.M.; Krug, J. Empirical Fitness Landscapes and the Predictability of Evolution. *Nat Rev Genet* **2014**, *15*, 480–490, doi:10.1038/nrg3744.
3. Schenk, M.F.; Szendro, I.G.; Salverda, M.L.M.; Krug, J.; de Visser, J.A.G.M. Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. *Mol Biol Evol* **2013**, *30*, 1779–1787, doi:10.1093/molbev/mst096.
4. Whelan, S.; Goldman, N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol Biol Evol* **2001**, *18*, 691–699, doi:10.1093/oxfordjournals.molbev.a003851.
5. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The Rapid Generation of Mutation Data Matrices from Protein Sequences. *Comput Appl Biosci* **1992**, *8*, 275–282, doi:10.1093/bioinformatics/8.3.275.

6. Figliuzzi, M.; Jacquier, H.; Schug, A.; Tenaillon, O.; Weigt, M. Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1. *Mol Biol Evol* **2016**, *33*, 268–280, doi:10.1093/molbev/msv211.
7. Gong, L.I.; Suchard, M.A.; Bloom, J.D. Stability-Mediated Epistasis Constrains the Evolution of an Influenza Protein. *Elife* **2013**, *2*, e00631, doi:10.7554/eLife.00631.
8. Jacquier, H.; Birgy, A.; Le Nagard, H.; Mechulam, Y.; Schmitt, E.; Glodt, J.; Bercot, B.; Petit, E.; Poulain, J.; Barnaud, G.; et al. Capturing the Mutational Landscape of the Beta-Lactamase TEM-1. *Proc Natl Acad Sci U S A* **2013**, *110*, 13067–13072, doi:10.1073/pnas.1215206110.
9. Dickinson, B.C.; Leconte, A.M.; Allen, B.; Esvelt, K.M.; Liu, D.R. Experimental Interrogation of the Path Dependence and Stochasticity of Protein Evolution Using Phage-Assisted Continuous Evolution. *Proc Natl Acad Sci U S A* **2013**, *110*, 9007–9012, doi:10.1073/pnas.1220670110.
10. Kaltenbach, M.; Jackson, C.J.; Campbell, E.C.; Hollfelder, F.; Tokuriki, N. Reverse Evolution Leads to Genotypic Incompatibility despite Functional and Active Site Convergence. *Elife* **2015**, *4*, doi:10.7554/eLife.06492.
11. Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **2009**, *138*, 774–786, doi:10.1016/j.cell.2009.07.038.
12. Dellus-Gur, E.; Elias, M.; Caselli, E.; Prati, F.; Salverda, M.L.M.; de Visser, J.A.G.M.; Fraser, J.S.; Tawfik, D.S. Negative Epistasis and Evolvability in TEM-1 β-Lactamase--The Thin Line between an Enzyme's Conformational Freedom and Disorder. *J Mol Biol* **2015**, *427*, 2396–2409, doi:10.1016/j.jmb.2015.05.011.
13. Huang, W.; Palzkill, T. A Natural Polymorphism in Beta-Lactamase Is a Global Suppressor. *Proc Natl Acad Sci U S A* **1997**, *94*, 8801–8806, doi:10.1073/pnas.94.16.8801.
14. Bershtein, S.; Segal, M.; Bekerman, R.; Tokuriki, N.; Tawfik, D.S. Robustness-Epistasis Link Shapes the Fitness Landscape of a Randomly Drifting Protein. *Nature* **2006**, *444*, 929–932, doi:10.1038/nature05385.
15. Kondrashov, A.S.; Sunyaev, S.; Kondrashov, F.A. Dobzhansky-Muller Incompatibilities in Protein Evolution. *Proc Natl Acad Sci U S A* **2002**, *99*, 14878–14883, doi:10.1073/pnas.232565499.
16. Jordan, D.M.; Frangakis, S.G.; Golzio, C.; Cassa, C.A.; Kurtzberg, J.; Task Force for Neonatal Genomics; Davis, E.E.; Sunyaev, S.R.; Katsanis, N. Identification of Cis-Suppression of Human Disease Mutations by Comparative Genomics. *Nature* **2015**, *524*, 225–229, doi:10.1038/nature14497.
17. Shah, P.; McCandlish, D.M.; Plotkin, J.B. Contingency and Entrenchment in Protein Evolution under Purifying Selection. *Proc Natl Acad Sci U S A* **2015**, *112*, E3226-3235, doi:10.1073/pnas.1412933112.
18. Cantón, R.; González-Alba, J.M.; Galán, J.C. CTX-M Enzymes: Origin and Diffusion. *Front Microbiol* **2012**, *3*, 110, doi:10.3389/fmicb.2012.00110.
19. Salverda, M.L.M.; De Visser, J.A.G.M.; Barlow, M. Natural Evolution of TEM-1 β-Lactamase: Experimental Reconstruction and Clinical Relevance. *FEMS Microbiol Rev* **2010**, *34*, 1015–1036, doi:10.1111/j.1574-6976.2010.00222.x.
20. Firnberg, E.; Labonte, J.W.; Gray, J.J.; Ostermeier, M. A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol Biol Evol* **2014**, *31*, 1581–1592, doi:10.1093/molbev/msu081.
21. Marciano, D.C.; Brown, N.G.; Palzkill, T. Analysis of the Plasticity of Location of the Arg244 Positive Charge within the Active Site of the TEM-1 Beta-Lactamase. *Protein Sci* **2009**, *18*, 2080–2089, doi:10.1002/pro.220.
22. Caporale, B.; Franceschini, N.; Perilli, M.; Segatore, B.; Rossolini, G.M.; Amicosante, G. Biochemical Characterization of Laboratory Mutants of Extended-Spectrum Beta-Lactamase TEM-60. *Antimicrob Agents Chemother* **2004**, *48*, 3579–3582, doi:10.1128/AAC.48.9.3579-3582.2004.
23. Osuna, J.; Pérez-Blancas, A.; Soberón, X. Improving a Circularly Permuted TEM-1 Beta-Lactamase by Directed Evolution. *Protein Eng* **2002**, *15*, 463–470, doi:10.1093/protein/15.6.463.
24. Hecky, J.; Müller, K.M. Structural Perturbation and Compensation by Directed Evolution at Physiological Temperature Leads to Thermostabilization of Beta-Lactamase. *Biochemistry* **2005**, *44*, 12640–12654, doi:10.1021/bi0501885.
25. Bershtein, S.; Goldin, K.; Tawfik, D.S. Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J Mol Biol* **2008**, *379*, 1029–1044, doi:10.1016/j.jmb.2008.04.024.
26. Kather, I.; Jakob, R.P.; Dobbek, H.; Schmid, F.X. Increased Folding Stability of TEM-1 Beta-Lactamase by in Vitro Selection. *J Mol Biol* **2008**, *383*, 238–251, doi:10.1016/j.jmb.2008.07.082.
27. Bloom, J.D.; Silberg, J.J.; Wilke, C.O.; Drummond, D.A.; Adami, C.; Arnold, F.H. Thermodynamic Prediction of Protein Neutrality. *Proc Natl Acad Sci U S A* **2005**, *102*, 606–611, doi:10.1073/pnas.0406744102.
28. Barlow, M.; Hall, B.G. Predicting Evolutionary Potential: In Vitro Evolution Accurately Reproduces Natural Evolution of the Tem Beta-Lactamase. *Genetics* **2002**, *160*, 823–832, doi:10.1093/genetics/160.3.823.
29. DePristo, M.A.; Weinreich, D.M.; Hartl, D.L. Missense Meanderings in Sequence Space: A Biophysical View of Protein Evolution. *Nat Rev Genet* **2005**, *6*, 678–687, doi:10.1038/nrg1672.

30. Wylie, C.S.; Shakhnovich, E.I. A Biophysical Protein Folding Model Accounts for Most Mutational Fitness Effects in Viruses. *Proc Natl Acad Sci U S A* **2011**, *108*, 9916–9921, doi:10.1073/pnas.1017572108.
31. Firnberg, E.; Ostermeier, M. PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLoS One* **2012**, *7*, e52031, doi:10.1371/journal.pone.0052031.
32. Kowalsky, C.A.; Klesmith, J.R.; Stapleton, J.A.; Kelly, V.; Reichkitzer, N.; Whitehead, T.A. High-Resolution Sequence-Function Mapping of Full-Length Proteins. *PLoS One* **2015**, *10*, e0118193, doi:10.1371/journal.pone.0118193.
33. Blazquez, J.; Morosini, M.I.; Negri, M.C.; Gonzalez-Leiza, M.; Baquero, F. Single Amino Acid Replacements at Positions Altered in Naturally Occurring Extended-Spectrum TEM Beta-Lactamases. *Antimicrob Agents Chemother* **1995**, *39*, 145–149, doi:10.1128/AAC.39.1.145.
34. Stemmer, W.P. Rapid Evolution of a Protein in Vitro by DNA Shuffling. *Nature* **1994**, *370*, 389–391, doi:10.1038/370389a0.
35. Ortlund, E.A.; Bridgham, J.T.; Redinbo, M.R.; Thornton, J.W. Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science* **2007**, *317*, 1544–1548, doi:10.1126/science.1142819.
36. Sideraki, V.; Huang, W.; Palzkill, T.; Gilbert, H.F. A Secondary Drug Resistance Mutation of TEM-1 Beta-Lactamase That Suppresses Misfolding and Aggregation. *Proc Natl Acad Sci U S A* **2001**, *98*, 283–288, doi:10.1073/pnas.98.1.283.
37. Chaïbi, E.B.; Sirot, D.; Paul, G.; Labia, R. Inhibitor-Resistant TEM Beta-Lactamases: Phenotypic, Genetic and Biochemical Characteristics. *J Antimicrob Chemother* **1999**, *43*, 447–458, doi:10.1093/jac/43.4.447.

 doi:10.20944/preprints202203.0246.v1

Supplementary Materials:

Supplementary text:
Details on sites and substitutions selected after antibiotic selection

V31
The amino acid of site 31 is a valine and is located in the first alpha-helix of the enzyme. The distance to site 251 is 22 Å. Accessibility to solvent is 53 (popMusic prediction). We find 5 different mutants after antibiotic selection that have restored, at least partially the hydrolytic capacities of the enzyme. Selected mutations are V31A, V31Q, V31M, V31R, and V31K with selection coefficient between 9 and 31, V31K being the most selected. To determine if these are mutations having an impact on stability, we used popMusic predictions. A negative ΔΔG (kcal/mol) makes possible to consider that they are stabilizing mutations. Predicted ΔΔG are -0.27, -0.66, -0.02, -0.41 and -0.46 respectively.
Of these, only V31A had been reported in laboratory isolates. Two other mutants of this site have also been described in laboratory isolates and are listed in the work of Salverda but none have been found in clinical isolates [19].

D38
The amino acid of site 38 is an aspartic acid that is located in the first alpha-helix of TEM-1. The distance to 251 site is 29 Å and accessibility to solvent is 57. We find 3 mutants selected: D38R, D38N, and D38K with selection coefficient of 15, 2.2 and 19 and predicted ΔΔG of -0.23, -0,02 and -0.35 kcal/mol respectively. Of note, the 2 most selected mutations are predicted to be more stable.
Of these, only D38N had been described in clinical isolates. However, this site is frequently subject to natural evolution because D38A, D38E, D38G, D38V, and D38Y have already been described among natural isolates.

Q39
It is also located in the first alpha-helix of TEM-1. The distance to site 251 is 28Å and accessibility to solvent is 57. We found only one selected mutant: Q39K. This substitution is found in multiple clinical isolates and has been shown to increase the MIC for ceftazidime and aztreonam[33]. This mutation is also found on its own in TEM-2. The predicted ΔΔG is -0.11 kcal/mol and its frequency increased 6.5 times before and after selection. Q39K results in a charge alteration of the surrounding region of the protein and that may affect β-lactamase translocation and thus the effective concentration of enzyme in the periplasmic space[19].

A42
The substitution A42G is known to play a role in cefotaxime resistance and was frequently identified in laboratory experiments[34]. This position is 29Å far from the 251 site and accessibility is 13, which means that the site is buried. The selection coefficient was 2.5, which is quite low compared to other sites. It is not predicted to be stabilizing as ΔΔG is positive (1.23 kcal/mol). The fact that it is found in a compensatory mutation suggests that it may have an influence on the overall stability of the protein but this could be limited explaining the low selection coefficient found.

I47V, E48M, and L49M

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

These 3 sites are in beta-sheet number 2 of TEM-1. Their accessibility is very low (around 0) and their distance to 251 site is between 12 and 15.5 Å. Selection coefficients are 9.5, 3.5, and 30 respectively. The prediction of stability for these sites is 1.29, -0.56, and 1.15 kcal/mol respectively. Only E48M is predicted to increase stability in TEM-1. I47V has already been described in laboratory isolates and L49M has been scarcely found in clinical isolates[19]. On the opposite, E48 has no described mutants.

L51

We found 2 substitutions on L51: L51M and L51T. This site is near the 251 site (10 Å) and accessibility is low (8.5). They were selected 3.5 and 8.2 times respectively and their ΔΔG prediction is of 1.29 and 2.10 kcal/mol respectively which means that they are not supposed to be stabilizing mutations. Furthermore, all substitutions on this site have predicted stability greater than 0. This is not consistent with the fact that 3 substitutions on this site L51I, L51F, and L51P are found in clinical isolates and that these substitutions play a role both in the activity (increase activity towards cephalosporins) and the stability of the enzyme [22,23]. It is probably its association with other mutations that increase the stability of the protein.

N52

Only N52E was selected in our experiment (selection of 3.5 times) whereas N52H, N52K, N52D, N52D, N52T, and N52Y have been found in multiple laboratory isolates. Accessibility to solvent is very high (88.5) and distance to 251 is small (12 Å). Prediction stability is -0.46 suggesting a role in stability as it is corroborated by experimental studies [24,25].

F60 and P62

F60 is in the beta-sheet number 1 of TEM and P62 is not involved in a secondary structure. Their distance to site 251 is between 21 and 22 Å and their accessibility to solvent is 20 and 25 respectively. Substitutions F60C, F60Y, F60W, P62G, and P62A were selected at respectively 2.1, 14.5, 17.5, 5.6, and 6.9 times and their prediction of stability is positive. None of these substitutions have been described elsewhere suggesting a particular interaction with site 251.

E63

Five different substitutions were selected in our experiment: E63N, E63W, E63F, E63H and E63Y with selection coefficient of 5.5, 7, 8.5, 8.5, and 32 respectively. Accessibility to solvent is high (78.5) and this site is distant to 251 (25 Å). All substitutions on site 63 are predicted to increase the stability of the protein. There is no known function for this site but several amino acid substitutions have been found at a high frequency in laboratory isolates [19].

S124

Both S124E and S124D were selected with an increased frequency of more than 5 times. This site is in the alpha-helix 4 and is located at 23Å to 251 site. Accessibility is quite high (32.5). These substitutions are not predicted to increase stability. They haven't been found in isolates elsewhere but S124G, S124N, and S124R have been found in clinical isolates and are to modulate the resistance phenotype [19].

E147

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

Amino acid substitutions E147A, E147G, and E147K have been found in laboratory isolates in several studies. We found only E147G in our selection experiment with an increase in the frequency of 4.6 times. Distance to 251 is of 33 Å, accessibility to solvent of 54, and prediction of ΔΔG is 1.41 kcal/mol. This substitution even if distant to the active site is probably a compensatory mutation because it has been found in compensation in several destabilized mutants laboratory[26].

H153

The substitution H153R has been found in our selection experiment (frequency has increased 10 times) in multiple clinical isolates such as other substitutions on the same site (H153D, H153L, H153Q, H153S, and H153Y). It is located in alpha-helix 6 at 30 Å to site 251 with an accessibility of 49 which is quite high. The frequent identification of this substitution in laboratory experiments suggests adaptive effects. It has been demonstrated to increase cefotaxime resistance to enzyme stability [25,26].

N175

N175 is located in the omega loop of TEM-1 at 35 Å to 251 and with very high accessibility to solvent (of 100). The substitution N175G was found 6.5 times more frequently after selection than before suggesting compensation of the destabilized site 251. An increase in protein stability is predicted (ΔΔG is -0.81 kcal/mol). Other substitutions on the same site have already been described in clinical and laboratory isolates (N175D, N175I, N175S, and N175Y) with impact on cefotaxime and ceftazidime resistance. Its localization in the omega-loop allows modifying the hydrolysis spectrum of beta-lactams [19,28,35] .

M182

M182T is the most well-known substitution in TEM backgrounds. It increases the thermodynamic stability of the enzyme and acts as a global suppressor of missense mutations that reduce enzyme stability [8,13,26,27]. Another explanation for the mode of action of M182T could be that it suppresses aggregation and misfolding induced by other mutations [36]. We found this substitution with an increase in the frequency of 340 times making her the most compensatory mutation. However, we also found other substitutions at variable frequencies such as M182S, M182C, M182F, M182I, M182L, and M182A with selection coefficient of 144, 14, 10, 2.5, and 2.3 respectively. The large compensation with the M182S mutation is also remarkable and has not yet been described in natural isolates. This site is located at 23.2 Å to site 251 and has medium accessibility of 19. It is not predicted to be stabilizing regarding the popMusic prediction.

A184

Site 184 is at 20 Å to site 251 and has low accessibility of 12. Three substitutions A184V, A184P, and A184L have been selected at frequencies of 3.5, 2.8, and 66 times in our experiments showing a strong compensation of this site. No increase in stability is predicted. Two other substitutions A184T and A184V have been found in laboratory isolates but not in our experiment.

E212

We observe a strong selection on site 212 which is located near site 251 (12.5Å) with accessibility of 46. It is in the alpha-helix 9. This site has never

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

been described as being of particular interest on stability, which is corroborated by prediction stability (ΔΔG>0). Its proximity to site 251 suggests local epistasis and particularly with the substitution E212A which is strongly selected (X46.5) but also with E212M and E212S (both frequencies were multiplied by 7).

G218

The substitution G218D was selected 3.5 times after the selection experiment. It is located at 17.3 Å of site 251 and is not implicated in a particular secondary structure. Prediction stability is 0.05 kcal/mol and thus, the mutation seems to be quite neutral. Only the substitution G218E has been described in clinical isolates with no known effect.

A224

This site is located in the alpha-helix 10 near site 251, at 12.6 Å with accessibility of 22. The substitution A224V was strongly selected with a selection coefficient of 34.5. This substitution found in clinical isolates is known to increase enzyme stability [25,26]. It is also located in proximity to site 251 which makes possible a local interaction in our protein.

F230

Located in beta-sheet number 5 of TEM, F230 is one of the closest amino acids of site 251 (located at 5.5 Å) with small accessibility to solvent (of 11). The selection of the substitution F230K (X 5.4 times) is very probably due to a very strong local epistasis. No increase in stability is predicted using popMusic and this site had never been described as playing a role in the stability of phenotype of resistance.

R275

The R275 is located in the 12th alpha-helix of TEM-1 with very low accessibility to solvent and at 24 Å to site 251. It is known to be involved in increasing the resistance to inactivation by clavulanate, a beta-lactamase inhibitor, but also in the resistance of cefotaxime. Furthermore, its effect on stability has been well documented and particularly for substitutions R275L and R275Q that were both found in studies selecting for stabilizing mutations [23,26,37]. Substitutions at residue 275 may be assimilated to a global suppressor mutation[13]. We found 3 substitutions selected: R275I was strongly selected with a selection coefficient of 54 but also R275T and R275V which were selected about 2.5 times. Only R275I has a negative ΔΔG (-0.14 kcal/mol) corroborating its potential impact on stability.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

Supplementary Table 1: list of compensating mutations in G251W TEM variants. The colors from white to red illustrate the increase in allelic frequency after amoxicillin selection.

Selected mutation (TEM-1 G251W)	Increase allele frequency under amoxicillin selection
V31A	8.91
V31K	30.75
V31M	16.81
V31Q	11.06
V31R	17.21
D38K	18.91
D38N	2.22
D38R	15.03
Q39K	6.25
A42G	2.28
I47V	9.56
E48M	3.55
L49M	30.19
L51M	3.66
L51T	8.24
N52E	3.35
F60C	2.10
F60W	17.37
F60Y	14.53
P62A	16.91
P62G	9.90
E63F	8.44
E63H	8.46
E63N	5.60
E63W	6.85
E63Y	32.10
S124E	5.18
S124D	5.30
E147G	4.61
H153R	10.14
N175G	6.61
M182A	2.31
M182C	13.99

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

M182F	10.19
M182I	7.07
M182L	2.47
M182S	143.79
M182T	339.62
A184L	66.24
A184P	2.85
A184V	3.62
E212A	46.59
E212M	7.03
E212S	7.09
G218D	3.16
A224A	2.93
A224V	34.64
F230K	5.39
R275I	53.76
R275T	2.61
R275V	2.28

Increase in allele frequency	1	5	10	20	30	50	75	100

Supplementary Table 2: Informations on the 3 amino acids that are found exclusively in beta-lactamases that harbor G251: amino acid, ambler numbering positions, 3D coordinates (x, y and z using PDB reference 1BTL) and distance to G251 in Ångströms.

Amino acid	Ambler position	pos_x (Calpha)	pos_y (Calpha)	pos_z (Calpha)	distance to 251 site (Å)
GLU	48	2.614	-0.909	18.168	12.2
GLY	251	-7.683	5.219	20.315	0.0
ARG	259	-3.89	-0.063	20.442	6.5
TRP	290	-3.544	5.42	11.889	9.4

Supplementary Table 3: list of primers used in the study

oligo name	5'-3' sequence
TEM-mut-G251W	ATCATTGCAGCACTGTGGCCAGATGGTAAGCC
TEM-pSKUNK-DIM-F	CTGTTGACAATTAATCATCGG
TEM-pSKUNK-DIM-R	GAACGCCAGCAAGACGTAG
NGS-TEM-part1-1-F	5’ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATTTCACACAGGAGGAAGGA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

NGS-TEM-part1-2-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNAATTTCACACAGGAGGAAGGA
NGS-TEM-part1-3-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNAATTTCACACAGGAGGAAGGA
NGS-TEM-part1-4-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNAATTTCACACAGGAGGAAGGA
NGS-TEM-part1-1-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTTGCTCTTGCCCGGCGT
NGS-TEM-part1-2-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNAGTTGCTCTTGCCCGGCGT
NGS-TEM-part1-3-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNAGTTGCTCTTGCCCGGCGT
NGS-TEM-part1-4-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNAGTTGCTCTTGCCCGGCGT
NGS-TEM-part2-1-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATGATGAGCACTTTTAAAGTTCT
NGS-TEM-part2-2-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNAATGATGAGCACTTTTAAAGTTCT
NGS-TEM-part2-3-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNAATGATGAGCACTTTTAAAGTTCT
NGS-TEM-part2-4-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNAATGATGAGCACTTTTAAAGTTCT
NGS-TEM-part2-1-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACGATCAAGGCGAGTTACAT
NGS-TEM-part2-2-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNAACGATCAAGGCGAGTTACAT
NGS-TEM-part2-3-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNAACGATCAAGGCGAGTTACAT
NGS-TEM-part2-4-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNAACGATCAAGGCGAGTTACAT
NGS-TEM-part3-1-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAACGATCGGAGGACCGAA
NGS-TEM-part3-2-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNACAACGATCGGAGGACCGAA
NGS-TEM-part3-3-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNACAACGATCGGAGGACCGAA
NGS-TEM-part3-4-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNACAACGATCGGAGGACCGAA
NGS-TEM-part3-1-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCAGATTTATCAGCAATAAACCA
NGS-TEM-part3-2-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNTCCAGATTTATCAGCAATAAACCA
NGS-TEM-part3-3-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNTCCAGATTTATCAGCAATAAACCA
NGS-TEM-part3-4-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNTCCAGATTTATCAGCAATAAACCA
NGS-TEM-part4-1-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGCGGATAAAGTTGCAGGA
NGS-TEM-part4-2-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNAGGCGGATAAAGTTGCAGGA
NGS-TEM-part4-3-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNAGGCGGATAAAGTTGCAGGA
NGS-TEM-part4-4-F	5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNAGGCGGATAAAGTTGCAGGA
NGS-TEM-part4-1-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGGTCGAATGCATAAGCTTACT
NGS-TEM-part4-2-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNCGGTCGAATGCATAAGCTTACT
NGS-TEM-part4-3-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNCGGTCGAATGCATAAGCTTACT

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 March 2022 doi:10.20944/preprints202203.0246.v1

NGS-TEM-part4-4-R	5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNCGGTCGAATGCATAAGCTTAC T