PNAS RESEARCH ARTICLE | BIOCHEMISTRY

Network of epistatic interactions in an enzyme active site revealed by large-scale deep mutational scanning

Allison Judge[a], Banumathi Sankaran[b], Liya Hu[a], Murugesan Palaniappan[c], André Birgy[a,d], B. V. Venkataram Prasad[a], and Timothy Palzkill[a,1]

Edited by Marc Ostermeier, Johns Hopkins University, Baltimore, MD; received August 6, 2023; accepted February 14, 2024 by Editorial Board Member Stephen J. Benkovic

Cooperative interactions between amino acids are critical for protein function. A genetic reflection of cooperativity is epistasis, which is when a change in the amino acid at one position changes the sequence requirements at another position. To assess epistasis within an enzyme active site, we utilized CTX-M β-lactamase as a model system. CTX-M hydrolyzes β-lactam antibiotics to provide antibiotic resistance, allowing a simple functional selection for rapid sorting of modified enzymes. We created all pairwise mutations across 17 active site positions in the β-lactamase enzyme and quantitated the function of variants against two β-lactam antibiotics using next-generation sequencing. Context-dependent sequence requirements were determined by comparing the antibiotic resistance function of double mutations across the CTX-M active site to their predicted function based on the constituent single mutations, revealing both positive epistasis (synergistic interactions) and negative epistasis (antagonistic interactions) between amino acid substitutions. The resulting trends demonstrate that positive epistasis is present throughout the active site, that epistasis between residues is mediated through substrate interactions, and that residues more tolerant to substitutions serve as generic compensators which are responsible for many cases of positive epistasis. Additionally, we show that a key catalytic residue (Glu166) is amenable to compensatory mutations, and we characterize one such double mutant (E166Y/N170G) that acts by an altered catalytic mechanism. These findings shed light on the unique biochemical factors that drive epistasis within an enzyme active site and will inform enzyme engineering efforts by bridging the gap between amino acid sequence and catalytic function.

epistasis | cooperativity | enzyme mechanism | enzyme evolution | fitness

Significance

Epistasis between amino acids is an indicator of critical residue interactions that drive protein function and evolution. Previously, comprehensive studies have surveyed epistasis within nonenzymatic proteins. Unlike nonenzymatic proteins, however, enzymes rely on the making and breaking of covalent bonds and the stabilization of high-energy transition states. Here, we offer a systematic study of epistasis within an enzyme active site, revealing that positive epistasis is common, mediated by substrate interactions, and concentrated at residue positions tolerant to substitutions. This will inform both enzyme engineering efforts and our basic understanding of the biochemical drivers of amino acid interactions that contribute to enzyme catalysis.

The structure, function, and evolution of proteins stem from both the primary amino acid sequence and complex interactions between amino acid residues, leading to epistasis. To understand enzyme function and to engineer enzymes for use in chemical production, biomedical research, and potential therapeutics, we must understand the basic mechanisms that drive enzyme function through epistasis between residues. Our ability to efficiently engineer an enzyme for a given function depends on understanding how amino acid sequence relates to enzyme function, which is driven by residue interactions (1, 2). These interactions are exceedingly complicated, with highly interconnected networks of residues that serve different roles, leading to the use of machine learning as a prominent tool for enzyme engineering (3, 4).

Deep mutational scanning (DMS) for individual codons is a widely used tool for testing the role of individual amino acids within a protein (5) and has been implemented in several enzyme systems, revealing the contributions of individual residues to enzyme function (6–8). This involves randomizing a single amino acid residue, creating a library containing all 20 amino acids at that given position, and testing for protein function. Increasingly, thanks to advances in NGS (next-generation sequencing) technology and cost efficiency, DMS for multiple residue positions, rather than a single position, has been used to characterize epistasis between amino acids within proteins (9–11), within a protein–peptide interaction interface (12), in a protein–protein interaction interface (13), and in sequential residues across an enzyme (14). Previous work in directed enzyme evolution (15–17), natural enzyme evolution (18–20), and enzyme evolvability (21, 22) informs the current understanding of epistasis and the contributions of amino acid interactions to enzyme function. To add to this literature, and to further advance enzyme engineering, we empirically surveyed amino acid interactions within an enzyme active site using extensive pairwise DMS as previously implemented in nonenzymatic protein systems (9–14). This provides both a basic scientific understanding of how sequence dictates protein function and a valuable dataset to inform machine learning algorithms for protein engineering.

Compared to nonenzymatic proteins, which typically do not rely on the breaking and reforming of covalent bonds for function, an enzyme active site requires precise chemistry and the stabilization of high-energy transition states. We, therefore, hypothesized that surveying

Author contributions: A.J., M.P., B.V.V.P., and T.P. designed research; A.J., B.S., L.H., M.P., and A.B. performed research; A.J., B.S., L.H., M.P., A.B., B.V.V.P., and T.P. analyzed data; and A.J., B.V.V.P., and T.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. M.O. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: timothyp@bcm.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2313513121/-/DCSupplemental.

Published March 14, 2024.

an enzyme active site would provide relevant information specific to enzyme function versus that of nonenzymatic proteins. Here, we test this hypotheis by comprehensively characterizing pairwise epistasis within the active site of CTX-M (cefotaximase from Munich).

CTX-M-14, a class A β-lactamase antibiotic resistance enzyme, provides a well-studied model system with a straightforward means of functional characterization—growth rate of *Escherichia coli* in media containing β-lactam antibiotic. Further, β-lactamases are a representative model system in that they act as serine hydrolases, which catalyze hydrolysis of an amide bond. Serine hydrolases play a role in a number of crucial physiological processes in both prokaryotic and eukaryotic organisms and represent important therapeutic targets (23, 24).

Epistasis is defined here as nonadditive fitness effects when two mutations are present within the active site compared to the fitness of the constituting single mutations. Since enzyme function is highly dependent on the substrate, we tested two β-lactam substrates, which produced different epistasis profiles. Notably, the CTX-M-14 enzyme is an outstanding catalyst for the substrates tested, with a catalytic efficiency (k_{cat}/K_M) for cefotaxime hydrolysis of approximately ~1 to 3 × 10^6 M^{-1} s^{-1} (25–27). Mutations to CTX-M-14 are largely detrimental to its β-lactamase function, which we previously demonstrated in a DMS experiment of individual codons throughout the active site (27). Therefore, we anticipated a mutational fitness landscape disparate from previous studies targeting the evolution of a novel catalytic function (15, 17, 18) since the vast majority of double mutations are expected to produce a nonfunctional enzyme. As a result, many instances of positive epistasis found in this study can be attributed to masking, a classic form of positive epistasis between genes where two nonfunctional genes in the same pathway produce a nonadditive result because the function level cannot go lower in the double than the nonfunctional level observed for the single mutants (28). Alternatively, we observe instances of positive epistasis that are not attributed to masking and interrogate one such example (E166Y/N170G) that utilizes an altered catalytic mechanism that resembles a class C β-lactamase enzyme.

Through wide-scale DMS of the CTX-M-14 active site, we show that positive epistasis is common, that epistatic interactions are mediated through the substrate, that residues more tolerant of amino acid substitutions serve as hotspots of positive epistasis, and that an active site residue that performs key chemical steps (Glu166) is surprisingly amenable to compensation, suggesting a path to the evolution of altered enzyme mechanisms.

Results

DMS Reveals the Relative Fitness of CTX-M-14 Single and Double Mutants. Based on X-ray crystal structures of CTX-M enzymes, 17 active site residues were targeted for mutagenesis and functional selection as shown in Fig. 1*A*. In previous work by our group, DMS of single codons was performed to elucidate the individual functions of the 17 selected amino acids within the CTX-M-14 active site (27). Here, in order to understand their interacting functions, codons were randomized in pairs so that each residue was randomized with 16 partners, giving 136 double-position libraries (16 + 15…+ 2 + 1 = 136) containing 49,096 double mutants (19^2 × 136) in total. To assess the amino acid sequence requirements for function, each library was introduced into *E. coli*, and the resulting clones were selected for function by growth in media containing either cefotaxime or ampicillin, both of which are β-lactam antibiotics that CTX-M-14 readily hydrolyzes. The distribution of amino acid pairs in the population of functional clones after the selection was assessed by NGS to determine the frequency of amino acids relative to the wild-type amino acid. For comparison, we also determined the frequency of each amino acid in the naive library pools before β-lactam antibiotic selection.

The frequency of each possible amino acid substitution was compared to the frequency of the wild-type amino acid and was used to calculate the relative fitness of each amino acid substitution conferred to *E. coli* after antibiotic selection (Eq. **1**) (17).

$$F = \log_{10}\left[\frac{N_{mut}^{sel}}{N_{mut}^{naive}}\right] - \log_{10}\left[\frac{N_{WT}^{sel}}{N_{WT}^{naive}}\right]. \quad [1]$$

The first term of Eq. **1** accounts for the frequency of a given amino acid (N_{mut}^{sel}) within an antibiotic-selected library and its frequency in the naïve library (N_{mut}^{naive}). The second term accounts for the frequency of the wild-type amino acid before (N_{WT}^{naive}) and after selection (N_{WT}^{sel}). Within each library, the wild-type enzyme has a relative fitness value $F = 0$, while a deleterious amino acid substitution has a fitness $F < 0$, and a beneficial substitution has a fitness $F > 0$. This calculation provides a quantitative estimate of the relative levels of function supplied by each amino acid substitution at a given active site position for a given antibiotic selection. The expected relative fitness of a double mutant, based on the performance of its two constituting single mutations within the same library, was compared to the observed fitness of the double mutant to determine epistasis according to the DMS2 model developed by Schmiedel and Lehner (29).

In the DMS2 model (Fig. 1*B*), a Loess regression is used to determine epistasis between mutations. The fitness of each double mutation (z-axis) is compared to the additive fitness of the individual single mutants based on a nonlinear model. Double mutations for which the fitness was significantly greater than expected based on the constituting single mutations (above the 95th percentile) were designated as positively epistatic, whereas double mutations where the fitness was significantly less than expected based on the constituting single mutations (below the 5th percentile) were characterized as negatively epistatic. Error for each fitness measurement was estimated based on sequencing coverage (number of counts), as previously described by Schmiedel and Lehner (29). Additionally, the standard deviation STD between fitness values of synonymous codons was implemented as an additional error estimate, as described previously by Firnberg et al. (30). As depicted in *SI Appendix*, Fig. S1, the median error value estimated based on sequencing coverage by the DMS2 program was $\sigma = 0.27$ for cefotaxime and $\sigma = 0.28$ for ampicillin selection. The median error estimated based on synonymous codon fitness values was $STD = 0.23$ and $STD = 0.25$ for cefotaxime and ampicillin, respectively. Furthermore, we found that error between synonymous codons generally decreased with higher postantibiotic selection count numbers (*SI Appendix*, Figs. S2*F* and S3*F*), which is accounted for in the DMS2 structure program σ value estimate. Further description of error estimation can be found in the *Materials and Methods* and *SI Appendix, Supplementary Materials*.

Negative epistasis detection was limited since many double mutants fell below the limit of detection described in the *Materials and Methods* section due to low frequency. Enzyme variants with fitness values below this lower limit of detection were deemed ineligible for negative epistasis evaluation. Only 2,396 double mutants were eligible to be assessed for negative epistasis for cefotaxime and 6,354 were eligible for ampicillin. This was expected given that we previously established that even single mutations within the CTX-M-14 are generally quite deleterious to function (27). Additionally, compensatory mutations (instances of positive

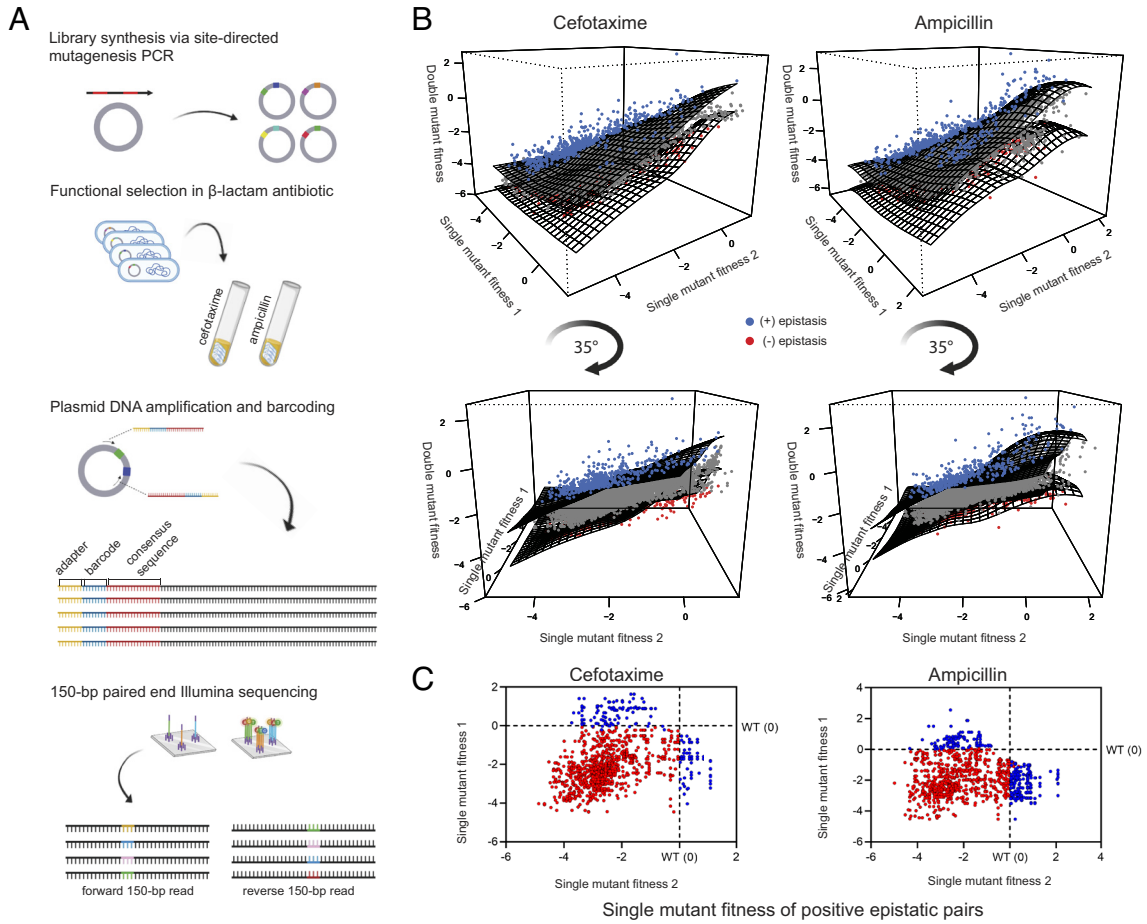

Fig. 1. (*A*) Randomized libraries were created, selected for function, and sequenced to determine epistasis between amino acid residues. (*B*) The DMS2 method established by Schmiedel and Lehner was used to determine epistasis by plotting the observed double-mutant fitness (*z*-axis) versus the fitness of the constituting single mutants (*x*- and *y*-axes). The black mesh represents the median running surface of the Loess regression to determine positive epistasis so that double mutants above the 95th percentile surface exhibit positive epistasis (blue), and those below the 5th percentile surface exhibit negative epistasis (red). Double mutants that did not exhibit epistasis lie between the 5th and 95th percentile surfaces (gray). A sample of 10,000 mutants is shown here to limit the pdf size. (*C*) Among residues exhibiting positive epistasis, the single-mutant fitness of the constituting single mutations was largely deleterious (red) or in some cases combines a deleterious mutation with a neutral or positive fitness mutation (blue).

epistasis) that improve enzyme fitness in a given mutation background are of greater interest for the purposes of protein engineering and will be the primary result discussed here. Among the 42,218 double mutation pairs eligible for positive epistasis determination for cefotaxime, 1,360 (3.2%) exhibited positive epistasis. For ampicillin, 1,103 (2.7%) of 40,663 eligible double mutants exhibited positive epistasis. Eligibility requirements were determined by abundance of sequencing counts, as described in the *Materials and Methods* section.

Positive epistasis results were disparate between substrates, evidenced by the differential connectivity in Fig. 2 *A* and *B*. Cefotaxime and ampicillin share several hotspots of positive epistasis for both substrates (*SI Appendix*, Fig. S4), including 166 and 104, but differ at other positions. Most starkly, position 235 is mutated in 16% of positively epistatic pairs for cefotaxime (224/1,360) but is mutated in only 7.8% of positively epistatic pairs for ampicillin (86/1,103). More specifically, of the double mutations that exhibit positive epistasis for cefotaxime and ampicillin, 268 pairs exhibit epistasis for both, making up 20% (268/1,360) and 24% (268/1,103), respectively. While hydrolysis of both substrates by CTX-M-14 occurs by the same catalytic mechanism, these differential results suggest that compensation to rescue partial enzyme function can occur through different paths, depending on the substrate.

Mutants that displayed positive epistasis can be divided into two categories depending on whether the effects of individual mutations had a neutral (or slightly positive) fitness effect or a deleterious fitness effect. For cefotaxime resistance, most of the positive epistasis arose from the combination of two mutations (88%), each of which was deleterious to function. In this case, mutations combined to create a double mutant that was more fit than expected based on the two constituting mutations (Fig. 1*C*, red points). It was less common that one of the two mutations was deleterious, while the other was neutral or positive (12%) (Fig. 1*C*, blue points). In this case, the negative effect of one mutation was offset more than expected by the positive or neutral effect of the other mutation. Interestingly, there are no double mutants observed where two neutral or positive mutations combined to produce an enzyme that was more active than expected. This could be related to the fact that cefotaxime is an excellent substrate for the wild-type enzyme and improving it further is difficult. Positively epistatic double mutants for ampicillin were more likely to have one neutral or positive mutation, with 75% of positively epistatic pairs including two deleterious mutations and 25% including one neutral or positive mutation. This is likely because there is a wider range of CTX-M-14 mutations that have no effect on ampicillin fitness compared to cefotaxime (27).

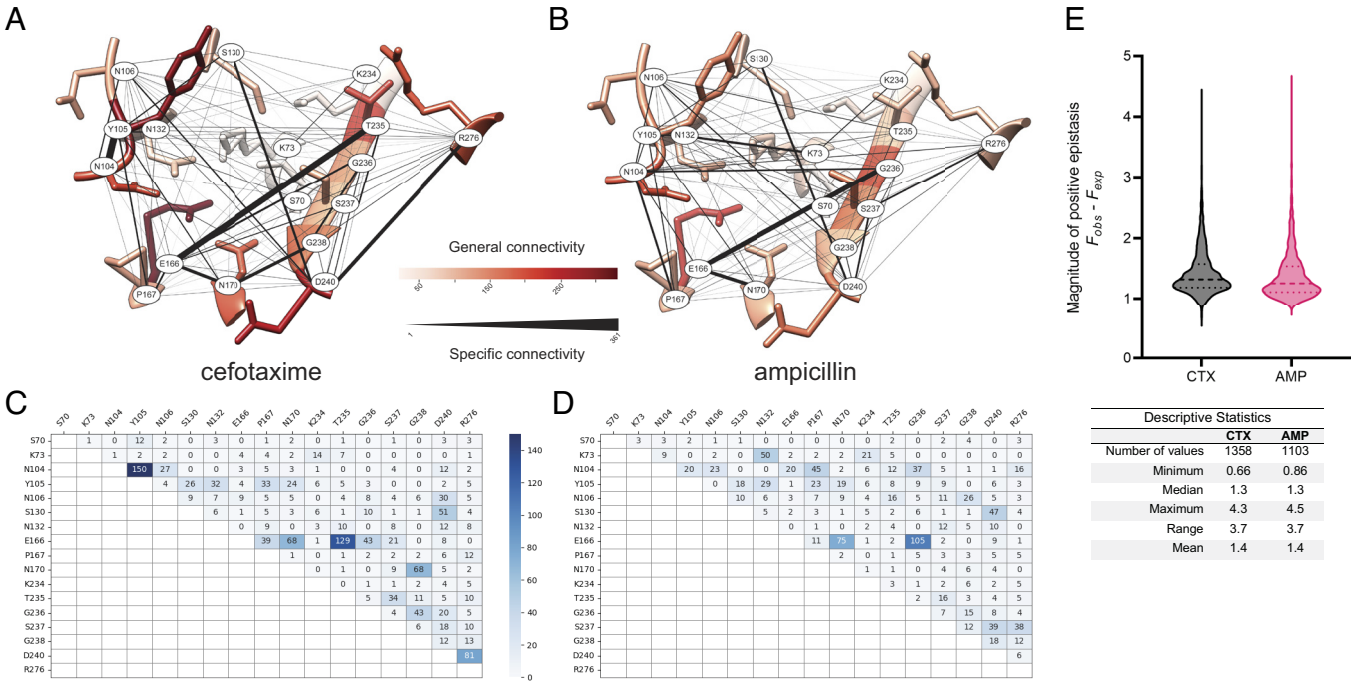

Fig. 2. The positive epistasis network for the 17 CTX-M-14 active site residues tested in this study for (*A*) cefotaxime and (*B*) ampicillin. The red color gradient represents the general connectivity of each residue or the number of positive epistatic pairs each residue participates in, regardless of the second residue in a given interaction. The thickness of each black line represents the specific connectivity between two pairs of residues or the number of positive epistatic pairs for each combination of residues. Specific positive epistasis is quantified in the blue heatmaps below for (*C*) cefotaxime and (*D*) ampicillin. (*E*) The relative level or magnitude of positive epistasis was calculated by subtracting the expected fitness (median of the Loess regression for a given point in 3-D space shown in Fig. 1*B*) from the observed fitness for each double mutant that exhibits positive epistasis. The plot displays the level of deviation from expected fitness for each double mutant on a $\log_{10}$ scale. The median deviation from expected fitness was $F = 1.3$ or 20× fitter than expected on a $\log_{10}$ scale for both substrates tested.

Epistasis Exhibits a Highly Connected Network, with Tolerant Residue Positions Highly Represented. Fig. 2 shows the network of amino acid connections that lead to positive epistasis for cefotaxime (Fig. 2*A*) and ampicillin (Fig. 2*B*) among CTX-M-14 variants. The red color scale represents the general connectivity of a given position so that more darkly colored residues are involved in a higher number of double position mutations that lead to positive epistasis, regardless of the interacting partner. This scale is quantified in *SI Appendix*, Fig. S4. The lines connecting positions represent specific connectivity between them or the number of double mutants that lead to positive epistasis for a given pair of positions. Specific connectivity is quantified in the heatmap below, for cefotaxime (Fig. 2*C*) and ampicillin (Fig. 2*D*). Positive epistasis is generally observed across active site residues, with each of the 17 active site residues participating in at least 21 positive epistatic mutations.

Mutations exhibiting positive epistasis do so at variable levels. The magnitude of epistasis for each positive epistasis pair was calculated by subtracting the expected fitness value (the median Loess regression fit) from the observed fitness value (Fig. 2*E*). Synergistic mutations are centered around $F = +1.3$ greater than expected, which is a 20-fold increase in the $\log_{10}$-based fitness value. As shown in Fig. 1*C*, many positively epistatic mutant pairs are composed of two deleterious mutations. They combine to produce a double-mutant enzyme that performs better than expected, fitting the definition of positive epistasis, but the majority of positively epistatic double mutants are still significantly less fit than the wild-type enzyme—the average double-mutant fitness among positively epistatic pairs is $F = -2.1$ for cefotaxime and $F = -1.7$ for ampicillin (*SI Appendix*, Fig. S5). This is an example of masking epistasis and is analogous to genetic systems where two nonfunctional genes in the same pathway do not have an additive effect (28). Within CTX-M-14, as an example, mutations S70P and D240C exhibit positive epistasis. Positive epistasis between S70P and D240C is not because the combination of mutations produces a highly functional enzyme (this double mutant has a fitness $F = -1.7$); instead, positive epistasis is detected because S70P effectively kills the enzyme, making the second D240C mutation inconsequential and thereby masking its effect in a way that is not specific to the amino acid identity at position 240. In future engineering efforts where the goal is improved catalytic activity, it will be important to differentiate between positive epistasis due to a nonspecific masking effect such as this, versus due to improved catalytic activity.

Residue Mutation Tolerance Plays a Role in Positive Epistasis. It has been established that neutral or slightly beneficial mutations can act as generic stabilizers within proteins, opening the door to a host of mutations that are deleterious in isolation (22, 31, 32). It follows, therefore, that if a residue position can survey a wide number of mutations with only a modest fitness cost, it could participate in a wider range of compensatory (positively epistatic) interactions with surrounding residues. We tested this hypothesis by comparing each residue's tolerance for mutation to its likelihood to participate in positive epistasis. The mutation tolerance for each residue, or the effective number of substitutions (k^*), was determined in a previous study (27). k^* (calculated by Equation 2, *Materials and Methods*) reflects the diversity of amino acids that occur at a position following selection and ranges from 1, where only a single amino acid is found, to 20, where each amino acid type is present at equal frequency, i.e., maximal diversity (33, 34). We found no clear correlation between k^* values and positive epistasis for cefotaxime (Fig. 3*A*), which is limited by the stringent sequence requirements for cefotaxime in the CTX-M active site.

In contrast, sequence requirements for ampicillin are far less strict, with k^* values as high as 12.3 and four out of 17 residues having values above five. As seen in Fig. 3*B*, we found that higher mutation tolerance generally leads to more participation in positive epistatic pairs for ampicillin. Increases in mutation tolerance at low k^* values are loosely correlated with an increase in compensatory (positive) epistasis. As mutational tolerance increases past $k^* = 2$ to 3, however, there are diminishing returns on increasing compensability. This suggests that a residue does not need to be capable of surveying a broad range of amino acids but that some wiggle room in fitness space can lead to compensatory mutational pairs and opens evolutionary paths for a protein.

Generic Compensators Are Common among Positive Epistasis Interactions. The 10 mutations that most often participate in positive epistasis, which we have termed "generic compensators" since they are not allele specific, are shown in Fig. 3*C*. Generic compensators identified within CTX-M-14 stabilize the enzyme or provide a neutral/positive fitness effect that compensates for a wide range of mutations elsewhere in the active site. Among the generic compensators identified, 4/10 for cefotaxime and 3/10 for ampicillin have a neutral or positive fitness effect (blue bars) in isolation. In the entire population, 20% of single mutations had a neutral or positive fitness effect for ampicillin, while only 9.5% had a neutral or positive fitness effect for cefotaxime. This indicates that individually beneficial mutations are overrepresented among generic compensators.

Generic compensators were assessed for stability changes to CTX-M-14 to determine whether compensation was due to an improvement in catalysis or due to stabilization of the enzyme. We found that 8/10 (cefotaxime) and 9/10 (ampicillin) generic compensators were predicted to enhance the stability of the enzyme based on melting temperature determination and/or FoldX modeling of ΔΔG. Stabilizing mutations that allow mutations at other positions have been thoroughly described in another class A β-lactamase, TEM-1 (35, 36). The most common mutation among pairs of mutations exhibiting epistasis against cefotaxime, D240P, was experimentally tested for protein stability. While proline mutations often stabilize protein structure, due to a more rigid peptide bond, D240P results in a lower melting temperature and, therefore, destabilizes CTX-M-14 (*SI Appendix*, Fig. S6). The remaining generic compensators were examined for stability, quantified as ΔΔG, using the FoldX program (37). The most common mutation in ampicillin epistatic pairs, Y105W, was also found to destabilize the enzyme, based on FoldX predictions based on its crystal structure (PDB: 7UON). This indicated that the D240P and Y105W mutations, rather than stabilizing the enzyme, provided a catalytic advantage that led to positive epistasis with many partner mutations, which was sufficient to overcome any destabilizing effect these mutations have on the protein (*SI Appendix*, Figs. S6 and S7). The remaining generic compensators [aside from mutations at G236, which likely destabilize the enzyme by introduction of a side chain that disrupts the β3 strand (27)] are stabilizing as shown in *SI Appendix*, Fig. S7 and indicated by an asterisk in Fig. 3*C*. Glutamate 166 mutations that act as generic compensators were universally found to be stabilizing (*SI Appendix*, Fig. S7). This is unsurprising, as it has been established in class A β-lactamases (38), and other enzymes (22), that catalytic residues are often destabilizing; therefore, mutation of these residues has a stabilizing effect. This leads to a tradeoff between stability and

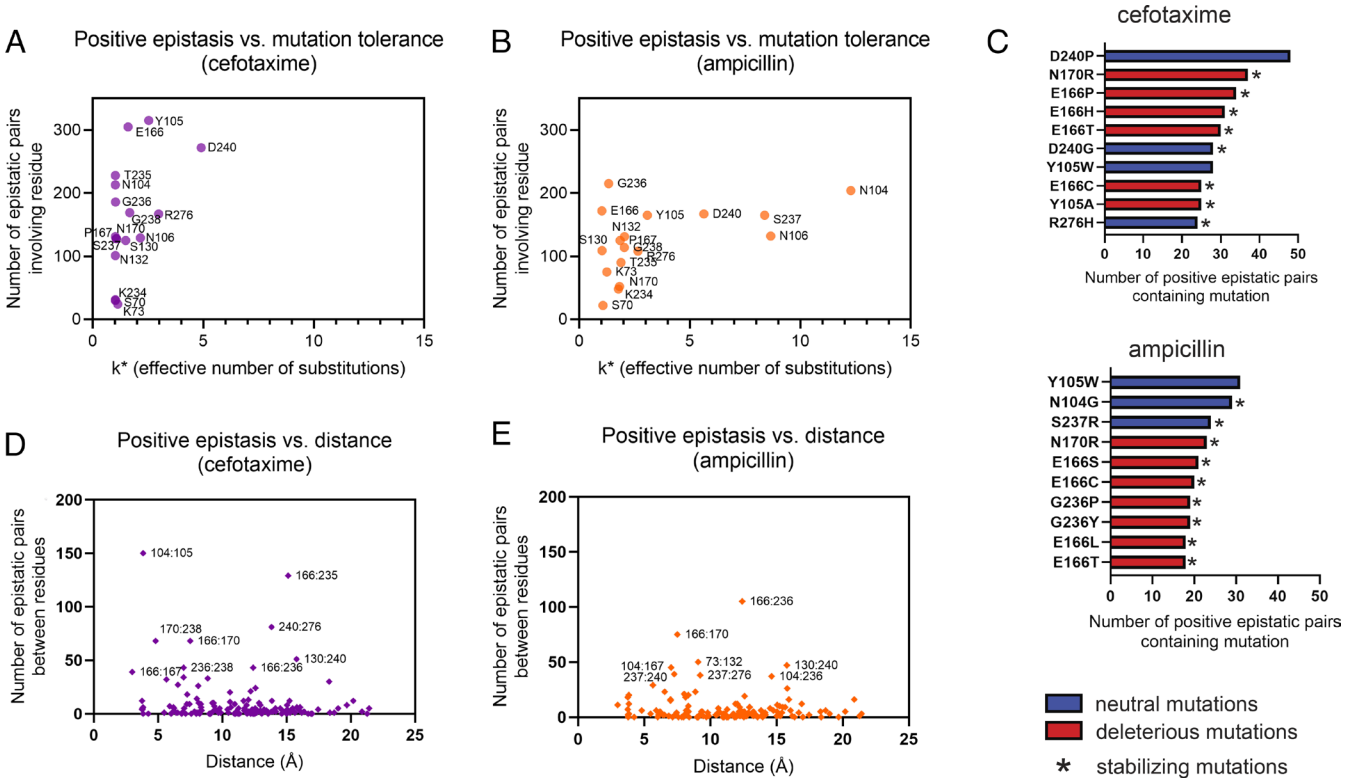

Fig. 3. The relationship between positive epistasis and other factors is plotted to understand the drivers of positive epistasis within the active site. Effective number of substitutions, or k^* (x-axis), plotted versus the number of positive epistasis pairs (y-axis) for a given position when selected against cefotaxime (*A*) or ampicillin (*B*). The top 10 most common mutations found among positive epistatic mutation pairs for cefotaxime or ampicillin (*C*). Stabilizing mutations are indicated by an asterisk, and their estimated ΔΔG values for their effects on protein stability are shown in *SI Appendix*, Fig. S7. The blue and red bars indicate that the mutation had a positive or neutral (blue) or negative (red) effect on enzyme fitness relative to the wild type. The Ångstrom distance between alpha carbons (x-axis) plotted versus the number of positive epistasis pairs (y-axis) for each pair of positions when selected against cefotaxime (*C*) or ampicillin (*D*).

catalytic activity, described as the stability–function hypothesis (39).

Substrate Proximity Allows Interaction from Physically Distant Residues. A primary principle of coevolution-guided structural determination is that residues in close proximity are more likely to be epistatic to one another and, therefore, to coevolve (40, 41). This principle has also been demonstrated in DMS experiments of nonenzymatic proteins and between proteins in protein–protein interactions (9). We find in the CTX-M-14 active site, however, that the distance between residues (up to 21 Å between alpha carbons in this study) is not a determinant of positive epistasis between them. Since the catalytic selection used here requires the presence of a substrate/transition state in the active site, the substrate can bridge the distance between residues, limiting the effects of distance on amino acid cooperativity (Fig. 3 *D* and *E*). The substrate as a determining factor is also evidenced by the variation in results between cefotaxime and ampicillin.

Glutamate 166 and threonine 235 are one such example of distant residues exhibiting high levels of epistasis for cefotaxime, but not ampicillin, hydrolysis. Of the 361 (19^2) double-mutant pairs in the Glu166 and Thr235 library, 129/361 mutation pairs (35%) exhibit positive epistasis. Both Glu166 and Thr235 are conserved among CTX-M enzymes, and alternative amino acids have poor fitness against cefotaxime—with the exception of E166Y, which has partial function (27). The residue alpha carbons sit 15.3 Å apart, with their closest atoms (Glu166 OE1 and Thr235 OG1) 10.7 Å apart in both the apo structure (PDB: 1YLT) and the substrate-bound structure (PDB: 4PM5). This distance places the residues outside of the direct contact range of 4 to 8 Å (29, 42). Both residues do, however, interact with cefotaxime as shown in PDB 4PM5 (*SI Appendix*, Fig. S8). In the wild-type enzyme (and likely in those double mutants exhibiting positive epistasis), epistatic interactions between positions 166 and 235 are likely mediated through the cefotaxime substrate. The identity of Thr235 in a wild-type background is key to positioning cefotaxime appropriately so that Glu166 is correctly oriented to deprotonate a water molecule during acyl-enzyme hydrolysis (*SI Appendix*, Fig. S9*A*). Through amino acid changes at 166 and 235, this precise orientation is changed, allowing for covariation at the two residue positions mediated by substrate interactions.

Glu166 Is Highly Amenable to Compensation Compared to Other Key Catalytic Residues. Ser70 serves as the nucleophile in the serine β-lactamase mechanism (*SI Appendix*, Fig. S9*A*). Previous studies have shown that orienting the substrate, particularly cephalosporin substrates, within the active site to achieve an optimized Bürgi–Dunitz angle of nucleophilic attack on the β-lactam ring is highly precise (43). Lys73 serves as part of the proton shuttle that facilitates both acylation of the substrate and deacylation of the covalent intermediate to form the hydrolyzed product (44). Despite Glu166 performing active site chemistry that requires precise reactivity and molecular geometry, we found that Glu166 is much more amenable to compensation than Ser70 and Lys73. For all active site residues tested, 166 is the most prevalent position among paired mutations that exhibit positive epistasis for both substrates, with 330 and 234 pairs for cefotaxime and ampicillin, respectively (*SI Appendix*, Fig. S4). Conversely, positions 70 and 73 are the residues least involved in epistatic pairs for cefotaxime and among the least for ampicillin (*SI Appendix*, Fig. S4).

The low incidence of compensation for positions 70 and 73 and the high incidence of compensation for position 166 indicate that the role of Glu166 is much more "compensate-able" compared to those of Ser70 and Lys73. This may be due to the intricate proton shuttle involving Ser70 and Lys73, which necessitates residues with the proper pK_a picking up and losing a proton at precisely the right time (*SI Appendix*, Fig. S9*A*). Even changes to the environment surrounding an amino acid can lead to changes in pK_a that alter catalysis (45), so it is likely that CTX-M enzymes have evolved with little room for error in this finely tuned mechanism. The pairwise mutations tested here are likely insufficient to create an altered proton shuttle mechanism that results in efficient hydrolysis. The role of Glu166, however, can be (very poorly) substituted for by bulk water in the active site. This may be the catalytic mechanism that prevails among stabilizing Glu166 substitutions (Fig. 3*C* and *SI Appendix*, Fig. S7), suggesting that the tradeoff between stability and catalytic function plays a larger role at position 166 compared to positions 70 and 73.

While Glu166 is highly conserved over evolution, we previously demonstrated that 1) Tyr166 can be substituted and retain partial CTX-M fitness and 2) low levels of CTX-M activity against ceftazidime, a cephalosporin antibiotic, are not at all dependent on the identity of Glu166 but still rely on other active site residues, such as Ser70 and Lys73 (27). These two findings indicate that the fitness landscape is more forgiving for mutations at Glu166 compared to Ser70 and Lys73. While this fitness landscape is not forgiving enough to be evident in β-lactamase evolution, which has fitness constraints beyond those evaluated in this controlled experiment, it is sufficient to move toward an alternative enzyme mechanism with the addition of one compensating mutation. We next tested one of these: E166Y paired with N170G.

Volume and Physicochemical Compensation Allow for an Altered Catalytic Mechanism. Positions 166 and 170 have high levels of epistatic interaction, including 75 mutation pairs with positive epistasis for ampicillin and 68 pairs for cefotaxime. Among these, E166Y/N170G was selected for further validation since it has the largest deviation from expected fitness. The double mutant's relative fitness, $F = 0.66$, is +3.3 greater than the median Loess regression ($F = -2.67$), indicating that the enzyme is approximately 2,000-fold more fit than expected. The E166Y/N170G double mutant was tested for stability to rule out the possibility that its increased fitness is solely due to increased protein stability. We demonstrated that E166Y/N170G has a decreased melting temperature compared to wild-type CTX-M-14 (*SI Appendix*, Fig. S6), suggesting that its fitness was instead due to catalytic function.

The wild-type, the E166Y and N170G single mutants, and the E166Y/N170G double-mutant enzymes were purified and tested for activity by determining steady-state kinetic parameters (*SI Appendix*, Fig. S10). In the class A serine β-lactamase mechanism, $k_{cat}(s^{-1})$ represents the substrate turnover rate once substrate is bound. K_M (μM) is dependent on the substrate binding affinity (K_D) and the acylation (k_2), and deacylation (k_3) rates. k_{cat}/K_M ($\mu M^{-1}\ s^{-1}$) is often referred to as the catalytic efficiency, representing the rate of substrate binding (k_1 and k_{-1}) and acylation (k_2) (46).

The mutation of Glu166 to tyrosine led to a steep drop-off in catalytic activity for cefotaxime (Table 1): The turnover rate, k_{cat}, decreased 30,000-fold, and the catalytic efficiency, k_{cat}/K_M, decreased 300-fold, ameliorated only by an extremely low K_M, which in this case was unlikely to be due to tighter substrate binding, but rather due to a greatly decreased deacylation rate, or k_3. The N170G substitution also lowered catalytic activity for cefotaxime, although not to the extent of E166Y. The k_{cat} of the N170G enzyme decreased twofold while K_M increased eightfold, resulting in a 15-fold decrease in k_{cat}/K_M compared to the wild-type enzyme (Table 1). Similar to the E166Y mutant, the E166Y/

Table 1. Steady-state kinetic parameters of CTX-M-14 hydrolysis of ampicillin and cefotaxime

	Cefotaxime			Ampicillin		
CTX-M-14	k_{cat}, s^{-1}	K_M, μM	k_{cat}/K_M, $\mu M^{-1}\ s^{-1}$	k_{cat}, s^{-1}	K_M, μM	k_{cat}/K_M, $\mu M^{-1}\ s^{-1}$
WT*	76 ± 3	105 ± 11	0.72 ± 0.08†	55 ± 2‡	33 ± 4	1.7 ± 0.2
E166Y*	0.0024 ± 5 × 10^{-5}	7.0 ± 0.7	0.00035 ± 4 x 10^{-5}	0.0015 ± 1.6 × 10^{-4}	8.4 ± 4	0.00018 ± 9 x 10^{-5}
N170G	40 ± 3	850 ± 120	0.047 ± 0.008	1.8 ± 0.03	10 ± 1	0.18 ± 0.02
E166Y/N170G	0.15 ± 0.003	2.5 ± 0.4	0.06 ± 0.01	0.0026 ± 0.0002	2.1 ± 1	0.0012 ± 6 x 10^{-5}

*Previously published data (26).
†The SE for k_{cat}/K_M was calculated by propagating the error of $k_{cat}$and K_M, respectively, using Equation 2 (*Materials and Methods*).
‡The ± symbol represents the SE for each value.

N170G double mutant exhibited a low K_M of 2 μM, but the turnover rate was greatly increased, 62× higher than that for the E166Y single mutant. While the E166Y/N170G double mutant was not as effective of an enzyme as wild-type CTX-M-14 (k_{cat}/K_M is 12-fold lower than that of the wild type), introduction of N170G, a mutation which is also detrimental in the wild-type background, showed a marked improvement over the E166Y single mutant (exhibiting a 170-fold increase in k_{cat}/K_M), indicative of epistasis. We suspect that while less fit when tested at increasing cefotaxime concentrations in steady-state kinetics experiments, the E166Y/N170G enzyme reached a fitness threshold that allowed it to survive at a similar frequency as the wild-type enzyme in survival selection (F = 0.66). It should also be noted that steady-state kinetics experiments were performed at 25 °C, while fitness selection was performed at 37 °C.

To quantitate the above observations on nonadditivity of the E166Y and N170G substitutions, we performed thermodynamic cycle analysis using Equation 3 (47):

$$\Delta\Delta G_{E166Y/N170G} = \Delta\Delta G_{E166Y} + \Delta\Delta G_{N170G} + \Delta G_I.$$

The change in free energy due to the mutations (ΔΔG) was calculated for the k_{cat}/K_M values for the single mutants and E166Y/N170G double mutant (Table 2), according to Equation 4 (*Materials and Methods*). The ΔΔG for the E166Y and N170G enzymes was 4.53 and 1.62 kcal/mol, respectively, which sums to 6.15 kcal/mol. However, the ΔΔG for the E166Y/N170G double mutant is 1.47 kcal/mol, which results in a very large interaction energy (ΔG_I) of −4.68 kcal/mol (Table 2). Thus, the double mutant is −4.68 kcal/mol more active than expected based on simple additivity of the effects of the single mutants. In other words, the k_{cat}/K_M value for the E166Y/N170G double mutant is 2,600-fold higher than expected based on the values of the single mutants. Thus, the thermodynamic cycle results agree with the predictions from deep sequencing that E166Y and N170G act nonadditively in cefotaxime hydrolysis. Note that the E166Y/N170G mutant showed an interaction energy (ΔG_I) of −2.42 kcal/mol with ampicillin as substrate; however, the large error on the ΔG_I value (5 kcal/mol) precludes any conclusion with regard to synergy and epistasis (Table 2). Fitness results, similarly, did not predict epistasis for E166Y/N170G under ampicillin selection, with a double-mutant fitness value of F = −2.3. This substrate-specific result may be due to the unique orientation of the Ω-loop (residues 165 to 170) required for cefotaxime hydrolysis, but not for ampicillin hydrolysis. This is consistent with our previously published results where Ω-loop mutations P167E and N170G have limited effects on ampicillin fitness for CTX-M-14 (27).

To further understand the interaction between the E166Y and N170G mutations, the double-mutant protein was crystallized, and the X-ray crystal structure in comparison to the wild-type enzyme (PDB: 1YLT) can be seen in Fig. 4 with statistics reported in Table 3 (PDB: 8SJ3) (Table 3). We found that N170G compensates for the E166Y mutation, which introduces a bulky tyrosine residue, through volume compensation. This relationship is quantified in *SI Appendix*, Fig. S11 based on the equation of Dutheil and Galtier (48). The E166Y/N170G crystal structure also showed that the catalytic water molecule coordinated between Glu166, Asn170, and Ser70 in the wild-type enzyme was no longer present in the double mutant (Fig. 4*C*). Instead, Tyr166 interacted directly with Ser70, suggesting that the catalytic mechanism differs from the wild type since Glu166 acts by deprotonating the catalytic water, which then carries out the deacylation reaction (*SI Appendix*, Fig. S9*A*). The mechanism cannot be defined based only on this crystal structure but may resemble that of class C β-lactamase enzymes, which utilize a tyrosine (Tyr150) in place of the glutamate found in class A β-lactamases (49, 50). Note that there is a new water molecule in the E166Y/N170G structure that is coordinated by the hydroxyl of Tyr166, the hydroxyl of Ser70, and the carbonyl oxygens of Leu169 and Ser237 (Fig. 4*C* and *SI Appendix*, Fig. S12). One possibility is that Tyr166 acts as a base to abstract the proton from this water that then attacks the acyl-enzyme ester for the deacylation reaction (*SI Appendix*, Fig. S9*B*). However, the positioning of this water is not optimal for deacylation, which would be consistent with the lower catalytic activity relative to the wild type (*SI Appendix*, Fig. S9*B*). Finally, the N170G substitution creates space for Tyr166 not only due to removal of the Asn side chain but also due to a conformational change of the omega loop on which residues 166 and 170 reside (Fig. 4*D*). The loop is twisted and shifted away from the active site to create more space for Tyr166 and substrate. Coincident with the conformation change, the phi/psi angles of residue 170 change from −178.7°, 12.1° in the wild type to −78.6°, 115.2° for the E166Y/N170G enzyme. Note that the −78.6°, 115.2° phi/psi angles are not favorable for asparagine. Therefore, the glycine substitution compensates for the presence of Tyr166 by removal of the side chain as well as promoting a conformational change that creates additional extra space by sampling angles that are not possible with asparagine at position 170.

Table 2. Free energy values and additivity relationships for single and double mutants for k_{cat}/K_M

		E166Y	N170G	E166Y/N170G	ΔG_I
ΔΔG k_{cat}/K_M (kcal/mol)	Ampicillin	5.40 ± 2.75	1.31 ± 0.21	4.29 ± 2.19	−2.42 ± 5.15
	Cefotaxime	4.53 ± 0.69	1.62 ± 0.52	1.47 ± 0.29	−4.68 ± 1.50

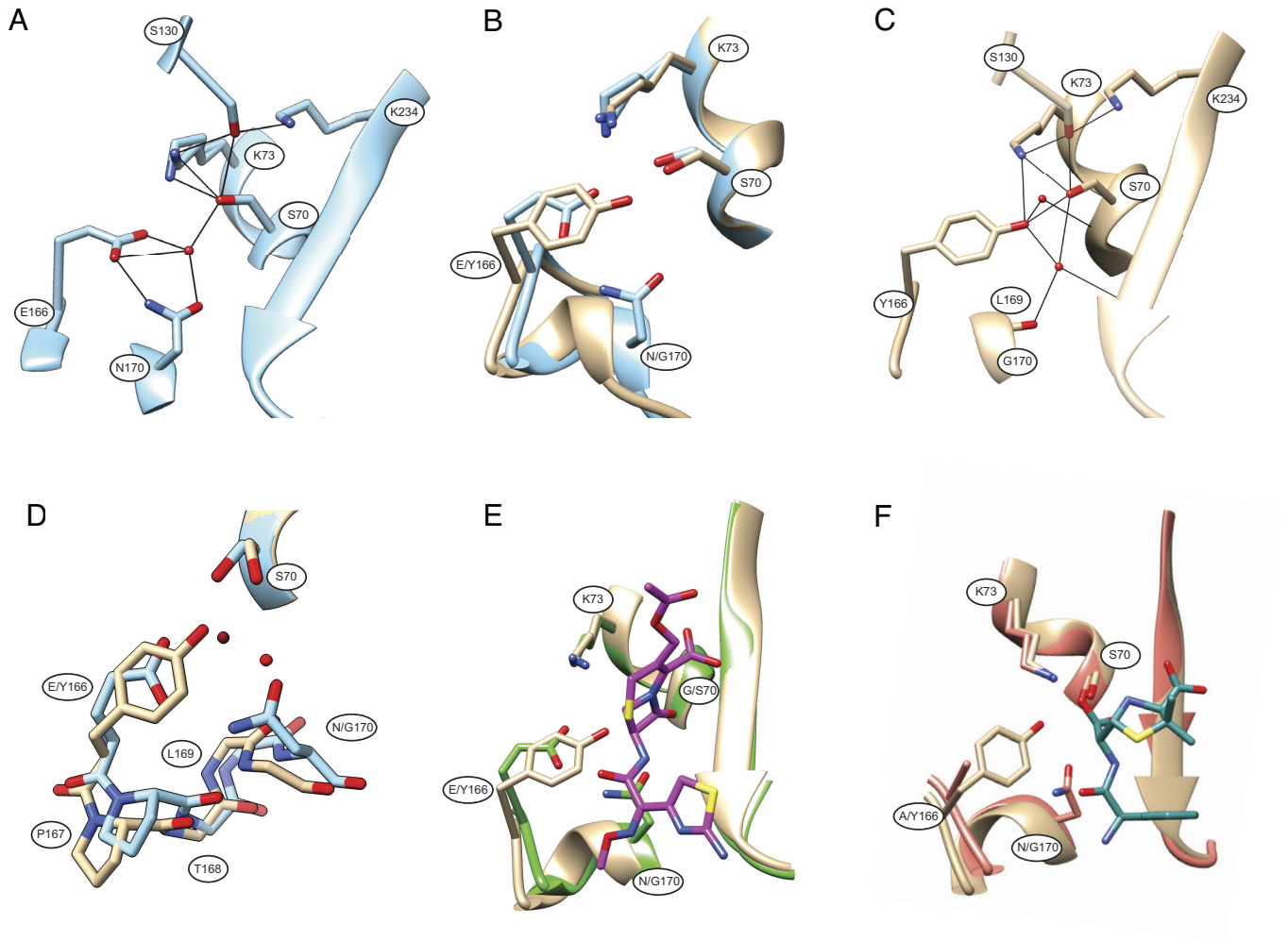

Fig. 4. Crystal structure of the CTX-M-14 E166Y/N170G mutant (PDB: 8SJ3, tan) compared to CTX-M-14^{WT} (light blue). Oxygen is shown in red and nitrogen in dark blue. Black lines represent potential hydrogen bonds, determined based on the proximity of oxygen and hydrogen atoms by the Chimera FindHBond tool. (*A*) In the wild-type enzyme, Ser70 is primed for nucleophilic attack through a catalytic water molecule, which is activated by Glu166. Here, you see the water molecule coordinated between Glu166 and Ser70. (*B*) The bulkier tyrosine base of the E166Y/N170G mutant enzyme (tan) is compensated for by a glycine mutation in the same loop, which opens up space in the bottom of the active site. (*C*) In the mutant enzyme structure, the catalytic water molecule is no longer coordinated between residue 166 and Ser70, suggesting that CTX-M-14 E166Y/N170G acts by an altered catalytic mechanism, made possible by volume and physicochemical compensation, or epistasis, between the two mutations. The omit map displaying electron density for the two active site waters can be found in *SI Appendix*, Fig. S8. (*D*) In a more detailed stick model representation of the active site that shows the entire omega loop, the catalytic water for the wild type is shown as a dark red sphere, while the new water in E166Y/N170G is in red. Note the twist and shift of the omega loop residues in the double mutant away from Ser70 to expand the active site. The side chains of Thr168 and Leu169 are not shown for clarity. (*E*) The cefotaxime-bound structure of the S70G mutant (PDB: 4PM5, green) is overlapped with the E166Y/N170G mutant structure to demonstrate how the mutations may alter orientation of the enzyme in proximity to cefotaxime. (*F*) The acylated ampicillin enzyme–substrate complex of the E166A mutant enzyme (PDB: 8B2W, salmon) is overlapped with the E166Y/N170G mutant structure to display the orientation of Tyr166 toward the acylated substrate.

Discussion

This survey of a serine hydrolase active site demonstrates basic principles of compensation between residues, some of which deviate from previously established principles of nonenzymatic proteins. We found that positive epistasis is prevalent throughout the CTX-M active site. Not only is positive epistasis substrate-dependent, but physically distant residues affect one another through the substrate. Additionally, the residues that most frequently participate in positive epistasis are more amenable to mutation.

Positive epistasis was prevalent throughout the CTX-M active site, with each of the 17 residues surveyed participating in at least one positive epistatic interaction (*SI Appendix*, Fig. S4) and approximately 3% of mutation pairs exhibiting positive epistasis (3.2% for cefotaxime and 2.7% for ampicillin). Previous studies of epistasis in the RNA binding domain of the *Saccharomyces cerevisiae* Poly(A) protein (RRM) (9) and the IgG-binding domain of protein G (GB1) (10) describe a low incidence of positive epistasis that is limited to a handful of residues. In another study of a class A β-lactamase by Gonzalez and Ostermeier, all sequential pairs of mutations throughout TEM-1 were created. Similarly to this survey of the CTX-M active site, they found a high incidence of positive epistasis (6.8%) (14), although the survey was limited to sequential pairs and did not focus on the β-lactamase active site. Taken together, this indicates that an enzyme active site contains residues that are more highly dependent on one another for function, compared to the RNA or protein binding interface of a protein.

We found epistatic residue pairs within the CTX-M active site that interact from a distance through the substrate. This is reflected in the substrate-specific epistasis between mutations at Glu166 and Thr235 and the E166Y/N170G double-mutant enzyme. High levels of epistasis between Glu 166 and Thr235 were evident under cefotaxime selection, but not ampicillin selection (Fig. 2 *A*–*D*), indicating that positioning of cefotaxime for catalysis is responsible

for their functional interaction. The E166Y and N170G mutations were found to be highly positively epistatic for cefotaxime (Table 1) but were not significantly epistatic for ampicillin according to both fitness results (*SI Appendix*) and steady-state kinetics (Table 1). Bridging through the substrate is likely why close proximity did not increase the likelihood of epistasis between residues. Previous studies found that positive epistasis was most enriched between residues close together in physical space for an RNA binding domain (9) and in the IgG-binding domain of protein G (GB1) (10). Substrate selectivity is key to our basic understanding of enzyme function, and understanding its drivers is of great interest to the field of enzymology.

Although epistasis was not entirely limited to specific hotspot residues as described in other, nonenzymatic proteins, we found that it was enriched in generic compensators. Many of these generic compensators were destabilizing, indicating that instead of stabilizing the protein to balance out additional mutations, they provide a background with a catalytic advantage for the enzyme. There was a general association between a position's tolerance to mutation (k^*) and participation in positive epistasis, indicating that mutability opens doors to positive epistasis, leading some neutral mutations, such as D240P and Y105W, to act as generic compensators. Generic compensating mutations that are neutral in a wild-type background open doors to additional mutations through epistatic interactions. This supports previous findings that neutral mutations earlier in evolution can cement a given evolutionary pathway as complementary mutations accumulate alongside them (51, 52).

Table 3. Data collection and refinement statistics (molecular replacement)

	CTX-M-14 E166Y/N170G (8SJ3)
Data collection	
Space group	P 41
Cell dimensions	
a, *b*, *c* (Å)	42.37, 42.37, 262.4
α, β, γ (°)	90.00, 90.00, 90.00
$CC_{1/2}$	0.82 (0.49)*
Resolution (Å)	32.97 – 1.50 (1.55 – 1.5)
$I/\sigma I$	3.26 (2.18)
Completeness (%)	96.0 (99.5)
Redundancy	4.9 (5.7)
Refinement	
Resolution (Å)	32.97 – 1.5
No. reflections	70,681 (7283)
R_{work}/R_{free}	0.1424/0.1976
No. atoms	4,575
Protein	4,069
Ligand/ion	-
Water	506
B-factors (Å^2)	18.0
Protein	16.74
Ligand/ion	-
Water	28.73
rmsd	
Bond lengths (Å)	0.006
Bond angles (°)	0.91

Statistics belong to one crystal.
*Values in parentheses are for highest-resolution shell.

This ability to survey a range of mutations that stabilize the protein or restructure the active site was confirmed to be pivotal for positive epistatic interactions and the development of an altered catalytic mechanism in the case of E166Y/N170G.

Due to high sequence stringency for function, the pervasive epistatic interactions found within the CTX-M active site in this study are not readily observable over natural evolution. A primary focus of the current literature on amino acid interaction within enzymes focuses on evolutionary analysis and/or mutations that are important in evolution (18–20). Evolution largely excludes detrimental mutations, preventing the accumulation of multiple complementary mutations and, therefore, an unbiased view of amino acid interactions that can be observed through DMS. By providing this unbiased empirical survey of amino acid interactions, we were able to interrogate trends in amino acid cooperation that evolution-based studies exclude and which may be important for future enzyme engineering efforts.

Generally, wild-type CTX-M-14 function was so comparatively high that few mutants survived selection in significant quantities. The resulting positive epistasis, therefore, largely represents compensatory epistasis that partially rescues a highly detrimental single mutation, including for the subsequently studied E166Y/N170G enzyme. Empirical studies on enzyme evolution generally aim to modify an enzyme to bind to a new substrate and catalyze a novel chemical reaction that was previously impossible or inefficient for that enzyme (15–17). Instead, we observed compensation in a different context, where the wild-type CTX-M-14 enzyme is exceedingly fit for the tested function, and the surrounding fitness landscape has steep fitness drop-offs for most mutations in the active site. Similar to evolutionary analyses, directed evolution focuses on a handful of mutations that provide a fitness advantage for the desired function. While this is a much more direct way of obtaining a desired function, we instead provide a broad view of active site amino acid cooperation and its effect on CTX-M-14 function. Furthermore, while the mechanistic scope of this study was narrowed to a select few mutations, the larger dataset will provide value in elucidating trends in amino acid cooperation (via machine learning or otherwise) that are not evident in previous studies.

The data provided here give insight into the drivers of epistasis between amino acids, which include highly specific enzyme–substrate interactions and neutral mutations providing catalytically favorable backgrounds for residue cooperation. The resulting dataset is also a valuable tool in training protein engineering algorithms since recent studies have shown that supervised algorithms outperform unsupervised algorithms (53, 54). By providing empirical datasets that test amino acid interaction within enzyme active sites, such as this one, we can bridge the gap in understanding how amino acid sequence translates to protein function.

Materials and Methods

Library Construction. Codon randomization libraries were constructed by oligonucleotide-directed mutagenesis of plasmid pTP123 containing the wild-type CTX-M-14 sequence. Positions close in the primary sequence (<18-bp) were randomized using partially overlapping primer pairs in site-directed mutagenesis PCR. Positions far apart in the primary sequence (>18-bp) were randomized by generating codon-randomized megaprimers, which were subsequently used for site-directed mutagenesis PCR. A detailed description of library construction is included in *SI Appendix*.

Fitness Selection. Codon randomization libraries were transformed into *E. coli* (XL1-Blue), and the resulting enzymes were selected for the ability to confer β-lactam antibiotic resistance in media containing ampicillin or cefotaxime. Plasmid DNA was isolated from surviving bacteria, amplified and barcoded via PCR, and underwent 150-bp paired-end Illumina sequencing to determine

the codons present at the positions of interest before and after β-lactam selection. Detailed descriptions of the *Bacterial Strain Used*, *Determination of the Appropriate β-lactam Antibiotic Concentrations*, *Functional Library Selection*, *NGS Preparation*, and *Deep Sequencing* are included in *SI Appendix*.

Computational Processing. Following deep sequencing, codon counts were determined using a custom Python 3.0 script. The number of times each amino acid occurred at the position of interest was counted to give totals used to calculate relative fitness using Equation 1 by Stiffler et al. (17). Epistasis classification was performed using the DMS2 method described by Schmiedel and Lehner (28). A detailed description of computational processing can be found in *SI Appendix*, including sequence coverage per library, read number requirements, error estimates, and determination of the lower limit of detection.

Biochemical Characterization. Select mutants were further characterized using steady-state enzyme kinetics, X-ray crystallography, differential scanning fluorimetry (DSF), and FoldX modeling. Detailed protocols for bacterial strains and plasmids used for protein purification, enzyme kinetics, X-ray crystallography, DSF, and FoldX modeling can be found in *SI Appendix*.

Statistics and Reproducibility. Error in fitness based on NGS reads following selection was based on the number of reads for each mutant, as described by Schmiedel and Lehner (28) and by comparison of synonymous codons, as described by Firnberg et al. (29). Detailed descriptions of the SE shown in Tables 1 and 2 and *SI Appendix*, Fig. S6.

Data, Materials, and Software Availability. The CTX-M-14 E166Y/N170G structure information has been deposited in the Protein Data Bank (https://www.rcsb.org/structure/8SJ3) (55) and assigned as PDB ID 8SJ3. Sequencing counts, calculated residue fitness, and Loess regression values used to determine epistasis, and all custom scripts for end pairing, variant counts, and fitness calculation are available on GitHub (https://github.com/Palzkill-Lab/CTXM_epistasis) (56). All other data shown in the main text or supplementary figures can be found in supporting information. CTX-M-14 protein expression constructs are available from Addgene. Paired-end sequencing reads were merged and processed using custom scripts written and executed in Python 3. Following fitness calculations, Schmiedel and Lehner's DMS2 program was executed in R to determine epistasis (29). Heatmaps were created using the Seaborn package as part of Matplotlib, executed in Python. All custom scripts for end pairing, variant counts, and fitness calculation are available on GitHub (https://github.com/Palzkill-Lab/CTXM_epistasis).

ACKNOWLEDGMENTS. This work was funded by NIH grant AI32956 to T.P. and Welch Foundation grant Q1279 to B.V.V.P. A.J. was funded by NIH training grant T32 GM120011. M.P. is supported by NIH grant (R03 CA259664) and Cancer Prevention Research Institute of Texas (CPRIT) grant (RP220524). The ALS-ENABLE beamlines used to collect X-ray diffraction data are supported in part by the NIH, National Institute of General Medical Sciences, grant P30 GM124169-01. The Advanced Light Source is a Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231. We thank Dr. H.F. Gilbert for providing feedback while we prepared this manuscript.

Author affiliations: [a]Verna and Marrs McLean Department of Biochemistry and Molecular Pharmacology, Baylor College of Medicine, Houston, TX 77030; [b]Department of Molecular Biophysics and Integrated Bioimaging, Berkeley Center for Structural Biology Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [c]Department of Pathology and Immunology, Center for Drug Discovery, Baylor College of Medicine, Houston, TX 77030; and [d]Infections, Antimicrobials, Modelling, Evolution, UMR 1137, French Insitute for Medical Research (INSERM), Faculty of Health, Université Paris Cité, Paris 75006, France

1. N. Ferruz *et al.*, From sequence to function through structure: Deep learning for protein design. *Comput. Struct. Biotechnol. J.* **21**, 238–250 (2023).
2. M. J. Harms, J. W. Thornton, Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366 (2010).
3. K. K. Yang, Z. Wu, F. H. Arnold, Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
4. M. Defresne, S. Barbe, T. Schiex, Protein design with deep learning. *Int. J. Mol. Sci.* **22**, 11741 (2021).
5. D. M. Fowler, S. Fields, Deep mutational scanning: A new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
6. P. A. Romero, T. M. Tran, A. R. Abate, Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7159–7164 (2015).
7. J. R. Klesmith, J.-P. Bacik, E. E. Wrenbeck, R. Michalczyk, T. A. Whitehead, Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2265–2270 (2017).
8. Z. Sun, L. Hu, B. Sankaran, B. V. V. Prasad, T. Palzkill, Differential active site requirements for NDM-1 β-lactamase hydrolysis of carbapenem versus penicillin and cephalosporin antibiotics. *Nat. Commun.* **9**, 4524 (2018).
9. D. Melamed, D. L. Young, C. E. Gamble, C. R. Miller, S. Fields, Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *RNA* **19**, 1537–1551 (2013).
10. C. A. Olson, N. C. Wu, R. Sun, A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol. CB* **24**, 2643–2651 (2014).
11. Y. Chen *et al.*, Deep mutational scanning of an oxygen-independent fluorescent protein creiLOV for comprehensive profiling of mutational and epistatic effects. *ACS Synth. Biol.* **12**, 1461–1473 (2023), 10.1021/acssynbio.2c00662.
12. C. L. Araya *et al.*, A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16858–16863 (2012).
13. G. Diss, B. Lehner, The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).
14. C. E. Gonzalez, M. Ostermeier, Pervasive pairwise intragenic epistasis among sequential mutations in TEM-1 β-lactamase. *J. Mol. Biol.* **431**, 1981–1992 (2019).
15. M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chem. Int. Ed.* **44**, 4192–4196 (2005).
16. J. Wang, G. Li, M. T. Reetz, Enzymatic site-selectivity enabled by structure-guided directed evolution. *Chem. Commun.* **53**, 3916–3928 (2017).
17. M. A. Stiffler, D. R. Hekstra, R. Ranganathan, Evolvability as a function of purifying selection in TEM-1 β-lactamase. *Cell* **160**, 882–892 (2015).
18. A. Rauwerdink *et al.*, Evolution of a catalytic mechanism. *Mol. Biol. Evol.* **33**, 971–979 (2016).
19. J. Zhang, H. Yang, M. Long, L. Li, A. M. Dean, Evolution of enzymatic activities of testis-specific short-chain dehydrogenase/reductase in Drosophila. *J. Mol. Evol.* **71**, 241–249 (2010).
20. B. Steinberg, M. Ostermeier, Shifting fitness and epistatic landscapes reflect trade-offs along an evolutionary pathway. *J. Mol. Biol.* **428**, 2730–2743 (2016).
21. K. Buda, C. M. Miton, N. Tokuriki, Pervasive epistasis exposes intramolecular networks in adaptive enzyme evolution. *Nat. Commun.* **14**, 8508 (2023).
22. N. Tokuriki, F. Stricher, L. Serrano, D. S. Tawfik, How protein stability and new functions trade off. *PLOS Comput. Biol.* **4**, e1000002 (2008).
23. D. A. Bachovchin, B. F. Cravatt, The pharmacological landscape and therapeutic potential of serine hydrolases. *Nat. Rev. Drug Discov.* **11**, 52–68 (2012).
24. J. Z. Long, B. F. Cravatt, The metabolic serine hydrolases and their functions in mammalian physiology and disease. *Chem. Rev.* **111**, 6022–6063 (2011).
25. Y. Chen, J. Delmas, J. Sirot, B. Shoichet, R. Bonnet, Atomic resolution structures of CTX-M β-lactamases: Extended spectrum activities from increased mobility and decreased stability. *J. Mol. Biol.* **348**, 349–362 (2005).
26. C. J. Adamski *et al.*, Molecular basis for the catalytic specificity of the CTX-M extended-spectrum β-lactamases. *Biochemistry* **54**, 447–457 (2015).
27. A. Judge *et al.*, Mapping the determinants of catalysis and substrate specificity of the antibiotic resistance enzyme CTX-M β-lactamase. *Commun. Biol.* **6**, 1–11 (2023).
28. W. Bateson, G. Mendel, *Mendel's Principles of Heredity* (Courier Corporation, 2013).
29. J. M. Schmiedel, B. Lehner, Determining protein structures using deep mutagenesis. *Nat. Genet.* **51**, 1177–1186 (2019).
30. E. Firnberg, J. W. Labonte, J. J. Gray, M. Ostermeier, A. Comprehensive, High-resolution map of a gene's fitness landscape. *Mol. Biol. Evol.* **31**, 1581–1592 (2014).
31. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, Protein stability promotes evolvability. *Proc. Natl. Acad. Sci.* **103**, 5869–5874 (2006).
32. L. I. Gong, M. A. Suchard, J. D. Bloom, Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2**, e00631 (2013).
33. P. S. Shenkin, B. Erman, L. D. Mastrandrea, Information-theoretical entropy as a measure of sequence variability. *Proteins Struct. Funct. Bioinforma.* **11**, 297–313 (1991).
34. Z. Deng *et al.*, Deep sequencing of systematic combinatorial libraries reveals β-lactamase sequence constraints at high resolution. *J. Mol. Biol.* **424**, 150–167 (2012).
35. W. Huang, T. Palzkill, A natural polymorphism in β-lactamase is a global suppressor. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8801–8806 (1997).
36. D. C. Marciano *et al.*, Genetic and structural characterization of an L201P global suppressor substitution in TEM-1 β-lactamase. *J. Mol. Biol.* **384**, 151–164 (2008).
37. J. Schymkowitz *et al.*, The FoldX web server: An online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
38. V. Stojanoski *et al.*, Removal of the side chain at the active-site serine by a glycine substitution increases the stability of a wide range of serine β-lactamases by relieving steric strain. *Biochemistry* **55**, 2479–2490 (2016).
39. B. K. Shoichet, W. A. Baase, R. Kuroki, B. W. Matthews, A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 452–456 (1995).
40. T. A. Hopf *et al.*, Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430 (2014).
41. D. de Juan, F. Pazos, A. Valencia, Emerging methods in protein co-evolution. *Nat. Rev. Genet.* **14**, 249–261 (2013).
42. M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, T. Hwa, Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 67–72 (2009).
43. Y. He, J. Lei, X. Pan, X. Huang, Y. Zhao, The hydrolytic water molecule of Class A β-lactamase relies on the acyl-enzyme intermediate ES* for proper coordination and catalysis. *Sci. Rep.* **10**, 10205 (2020).
44. V. Soeung *et al.*, A drug-resistant β-lactamase variant changes the conformation of its active site proton shuttle to alter substrate specificity and inhibitor potency. *J. Biol. Chem.* **295**, 18239–18255 (2020).
45. E. Edwardes Moore *et al.*, Understanding the local chemical environment of bioelectrocatalysis. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2114097119 (2022).
46. M. Vanhove, X. Raquet, T. Palzkill, R. H. Pain, J.-M. Frère, The rate-limiting step in the folding of the cis-Pro167Thr mutant of TEM-1 β-lactamase is the trans to cis isomerization of a non-proline peptide bond. *Proteins Struct. Funct. Bioinforma.* **25**, 104–111 (1996).

47. J. A. Wells, Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509–8517 (1990).
48. J. Dutheil, N. Galtier, Detecting groups of coevolving positions in a molecule: A clustering approach. *BMC Evol. Biol.* **7**, 242 (2007).
49. R. Tripathi, N. N. Nair, Mechanism of acyl-enzyme complex formation from the Henry–Michaelis complex of class C β-lactamases with β-lactam antibiotics. *J. Am. Chem. Soc.* **135**, 14679–14690 (2013).
50. R. Tripathi, N. N. Nair, Deacylation mechanism and kinetics of acyl-enzyme complex of class C β-lactamase and cephalothin. *J. Phys. Chem. B* **120**, 2681–2690 (2016).
51. C. Bank, R. T. Hietpas, J. D. Jensen, D. N. A. Bolon, A systematic survey of an intragenic epistatic landscape. *Mol. Biol. Evol.* **32**, 229–238 (2015).
52. I. Kim *et al.*, Energy landscape reshaped by strain-specific mutations underlies epistasis in NS1 evolution of influenza A virus. *Nat. Commun.* **13**, 5775 (2022).
53. Y. Xu *et al.*, Deep dive into machine learning models for protein engineering. *J. Chem. Inf. Model.* **60**, 2773–2790 (2020).
54. Y. Luo *et al.*, ECNet is an evolutionary context-integrated deep learning framework for protein engineering. *Nat. Commun.* **12**, 5743 (2021).
55. A. Judge, T. Palzkill, B. Sankaran, B. V. V. Prasad, L. Hu, Research Collaboratory for Structural Bioinformatics. Protein Data Bank. https://www.rcsb.org/structure/8SJ3. Deposited 17 April 2023.
56. A. Judge, CTXM_epistasis. Github. https://github.com/Palzkill-Lab/CTXM_epistasis. Deposited 10 April 2023.