FEBS Letters 586 (2012) 3330–3335

FEBS Letters

journal homepage: www.FEBSLetters.org

Sequence–function–stability relationships in proteins from datasets of functionally annotated variants: The case of TEM β-lactamases

Luciano A. Abriata [a,1,*], Merijn L. M. Salverda [b], Pablo E. Tomatis [a,2]

[a] Instituto de Biología Molecular y Celular de Rosario, Rosario, Argentina
[b] Laboratory of Genetics, Wageningen University, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 2 April 2012
Revised 2 July 2012
Accepted 4 July 2012
Available online 28 July 2012

Edited by Takashi Gojobori

Keywords:
TEM
Lactamase
Structure–function relationship
Function-stability tradeoff
Antibiotic resistance
Protein evolution

ABSTRACT

A dataset of TEM lactamase variants with different substrate and inhibition profiles was compiled and analyzed. Trends show that loops are the main evolvable regions in these enzymes, gradually accumulating mutations to generate increasingly complex functions. Notably, many mutations present in evolved enzymes are also found in simpler variants, probably originating functional promiscuity. Following a function-stability tradeoff, the increase in functional complexity driven by accumulation of mutations fosters the incorporation of other stability-restoring substitutions, although our analysis suggests they might not be as "global" as generally accepted and seem instead specific to different networks of protein sites. Finally, we show how this dataset can be used to model functional changes in TEMs based on the physicochemical properties of the amino acids.

1. Introduction

Understanding structure–function–stability relationships in proteins is a key issue in fundamental and applied science [1–3]. Since the early stages of structural biology, structure–function relationships have been dissected through functional assays on natural and artificial mutants of probably every protein studied to date. These efforts have been key instruments to answer questions about the role of specific amino acids on the function and stability of proteins. However, they are typically low-throughput so they can hardly provide other more general information, which might instead be available as trends inside larger datasets. Even experiments involving mutagenesis and selection usually fail to provide long records of mutated variants with annotated functional changes, for this would require extensive sequencing and bench work to measure function that might be expensive, tedious and not always rewarding.

Fortunately, the work of several groups has accumulated over decades for some interesting and well-behaved proteins, adding up to datasets with tens to a few hundred entries of different variants with annotated functional changes relative to a reference. Among them, one of the most studied proteins are the β-lactamases, which stand as the main mechanism by which opportunistic and pathogenic bacteria become resistant to β-lactam antibiotics [4,5]. On top of their value for fundamental science, evolution of these enzymes represents a constant challenge for the clinical treatment of infections and for the design of new antibiotics and inhibitors [6,7]. As a consequence, β-lactamases have been the subject of systematic studies on several natural and laboratory-evolved variants [8–18]. Taking advantage of this, we have herein compiled a dataset of the most studied group of β-lactamases, the TEM β-lactamases, and analyzed it to extract structure–function–stability relationships.

TEM-1 was the first β-lactamase reported of its kind and has become the most common β-lactamase found in *Enterobacteriaceae* [19]. It confers clinically relevant levels of resistance to penicillins and early cephalosporins, while its activity against later-generation cephalosporins such as cefotaxime is marginal. However, several TEM mutants with <10 substitutions have naturally evolved over time the ability to hydrolyze several new cephalosporins and/or overcome the effect of clinical inhibitors [20]. In this work, sequences and functional information for TEM-1-like and functionally enhanced variants arising from the clinic and from laboratory evolution experiments were compiled and analyzed. We assessed (i) how substitutions are distributed along the protein sequence

* Corresponding author. Address: Suipacha 590 (2000) Rosario, Argentina.
E-mail address: luciano.abriata@epfl.ch (L.A. Abriata).
[1] Current address: Laboratory of Biomolecular Modeling, Institute of Bioengineering, School of Life Sciences, Swiss Federal Institute of Technology, EPFL Lausanne, Switzerland.
[2] Current address: Biochemisches Institut, Universität Zürich, Zürich, Switzerland.

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2012.07.010

and structure, (ii) how mutations are acquired to produce new functions, (iii) how they relate with stability-restoring mutations and (iv) how this kind of data can be used to model and predict functional changes upon mutation in enzymes.

2. Methods

2.1. Dataset of TEM lactamases

The dataset of TEM variants gathers the sequences and functional profiles for TEM-1 and 181 different mutants of this protein (Table S1). It was built by merging entries from Lahey Clinic's β-Lactamase Database [21] with data available from directed evolution experiments in [8,10–12,14–15,17–18,20,22–24] and references therein. When a variant was present in clinical records and in directed evolution experiments, the information was merged into one entry. Deletion and insertion mutants were not included in the dataset.

Functional profiles are defined according to the phenotype each variant can confer to bacteria, namely the capacity to grow in the presence of the third-generation β-lactam cefotaxime and the capacity to overcome inhibition of lactamase activity by clavulanate. Four phenotypic groups are thus defined as in Lahey Clinic's database [21,25–26]: 2b variants are not able to confer bacteria resistance against cefotaxime and are inhibited by clavulanate, 2be variants confer resistance to cefotaxime but are sensitive to clavulanate, 2br variants are resistant to inhibition by clavulanate but do not confer resistance against cefotaxime, and 2ber variants can confer cefotaxime resistance being insensitive to clavulanate. For clinical variants, the phenotype reported in Lahey Clinic's database was adopted and information on their minimum inhibitor concentration (MIC) values was added to the dataset when available. Laboratory-evolved variants conferring MICs larger than 1 μg/mL against cefotaxime when expressed in *Escherichia coli* were classified as 2be and the remaining as 2b (i.e. applying the MIC cutoff used in Lahey Clinic's database).

The MICs conferred by 2be variants span three orders of magnitude (histogram in Fig. S1). Splitting this group at a threshold MIC of 8 μg/mL led to two groups, termed 2be and 2bee, with MIC values which differ by at least one order of magnitude (2bee being the enzymes producing MIC values >8 μg/mL).

2.2. Plots and calculations on structures

Information on secondary and tertiary structure of TEM-1 was extracted from PDB ID 1BTL. The energetic frustration index was calculated with the Protein Frustratometer Server [27] (http://bioinf.qb.fcen.uba.ar/frustra/index.php).

2.3. Phenotype prediction

The model developed in this work for sequence-based phenotype prediction of TEM lactamases is described under Results. It was implemented as in-house scripts. The DNA sequence for TEM-1 used as starting point in the simulations corresponds to GenBank entry ACT97477.1. Output ASCII tables were processed with standard spreadsheet programs. Principal Components Analysis for the calculation of reduced properties was performed using the program SimplePCAs.

3. Results and discussion

A dataset of 182 unique TEM variants with known sequences and functional profiles was compiled as described under Methods, by merging entries from Lahey Clinic's β-Lactamase Database with data collected from directed evolution experiments. The functional profiles are annotated in the form of five groups based on the phenotype they confer to bacteria as defined by Lahey Clinic's database [21,25–26] and as described in detail in Section 2. In brief, these groups are termed 2b (variants that cannot confer resistance against the third-generation β-lactam cefotaxime and are inhibited by clavulanate, such as TEM-1), 2br (variants which cannot confer resistance to cefotaxime either, but are resistant to clavulanate), 2be (variants which do confer resistance to cefotaxime but are sensitive to clavulanate), 2bee (2be variants which confer high resistance levels to cefotaxime) and 2ber (variants which confer resistance against cefotaxime and are not inhibited by clavulanate).

The database includes a total of 567 amino acid substitutions spread along the entire sequence (289 amino acids long including the signal peptide) as shown in Fig. 1A. Table S1 displays all the entries in the dataset including substitutions relative to TEM-1 and assigned phenotypic groups. The final composition of the database and other statistics are shown in Fig. 1. Sequence numbering utilized throughout the text follows Ambler's scheme as used in current literature [28].

3.1. Loops as evolvable hotspots

The first remarkable observation is that 85% of the substitutions in the dataset map to only 17 positions in the sequence, and that these positions are located either at loops or at the termini of secondary structure elements, except for position 276 which is in the middle of the C-terminal helix (Fig. 1A). Given that the large majority of the substituted sites show up already in the 2b group (Fig. 2) we can state that the dataset is dominated by adaptive substitutions. This is in line with the fact that it is very difficult to find neutral mutations in TEMs even under laboratory evolution conditions.

In the 3D structure of the protein, the substituted positions lay on the protein surface around the catalytic triad (formed by Ser70, Lys73 and Glu166, which are never substituted themselves); except for Gln39, Ala42, Gly92 and Met182 which are 15 Å or farther away from the active site (Fig. 1B). This reveals a high tolerance of surface loops to mutations and suggests that these elements are the main evolvable regions in proteins; moreover, we might speculate that loops have evolved to be evolvable. In structural terms, the evolvability of loops could be due to (1) their lower detrimental impact on stability compared to substitutions in the protein core and/or (2) the fact that they are flexible and involved in ligand binding, thus being more prone to be exploited by evolution for tuning enzymatic activity.

Interestingly, not all loops around the active site bear new-function substitutions. Calculation of the energetic mutational frustration for TEM-1 using the Protein Frustratometer Server [27] points to a cluster of highly frustrated interactions that overlaps with the highly substituted loops (red contact lines in Fig. S2). This match suggests that the identification of frustrated regions around active sites in proteins can be used as an indicator of possible evolvable hotspots, which represents a novel application of the concept of frustration in protein structure.

3.2. How accumulation of new-function mutations drives TEM evolution

The average number of substitutions relative to TEM-1 in each phenotypic group shows that the development of more complex activities (such as enhanced activity against cefotaxime, 2bee, or combining activity against cefotaxime with resistance to inhibition by clavulanic acid, 2ber) requires, on average, more mutations to accumulate (Fig. 1C and Table S2). Analysis of the distributions of substitutions within groups (Fig. 2) shows that several mutations leading to groups 2be, 2bee, 2br and 2ber already exist in the sim-

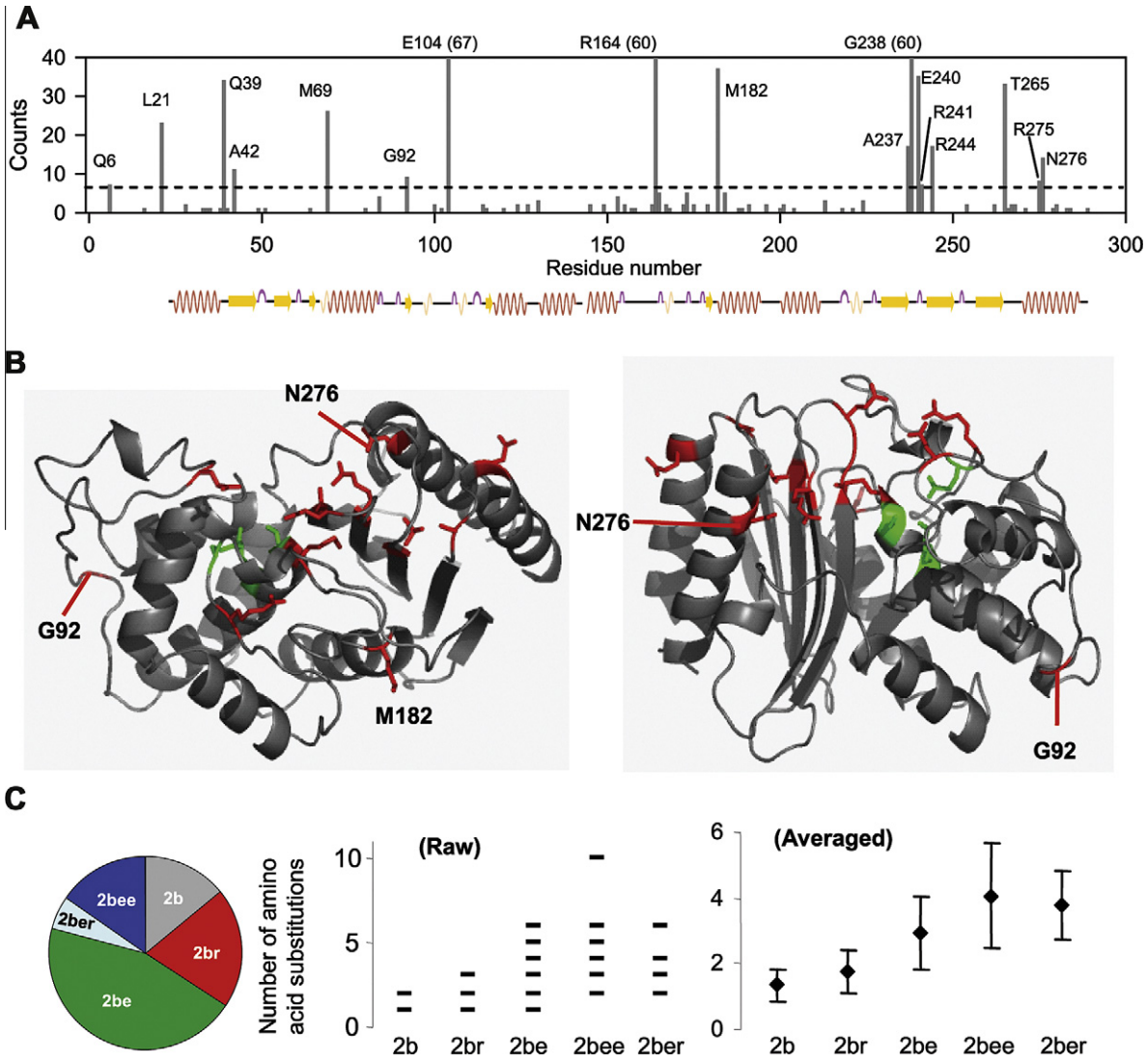

Fig. 1. Distribution of mutations in TEM β-lactamases. (A): Number of substitutions plotted against the sequence of TEM-1 (some bars are trimmed off for the sake of clarity, in these cases the number of substitutions is given in parentheses). Ambler's numbering scheme (in which residues 1, 2, 239 and 253 do not exist) is adopted in this figure and throughout the whole text. Structural features are shown under the amino acid numbers, where the red lines represent α-helices and yellow lines represent β-sheets. (B): The catalytic triad (Ser70, Lys73 and Glu166) and residues accounting for 85% of the substitutions in the dataset (taken above the line for 6 counts in panel A) are mapped in green and red, respectively, on the structure of TEM-1 (PDB ID 1BTL). (C): Composition of the database and number of amino acid replacements observed for each phenotypic group (middle: raw counts, right: averaged per group ± standard deviation).

plest group 2b and that they seem to have combined to create the evolved phenotypes (i.e. they are adaptive). If the five groups of enzymes are regarded as static averages of an evolving enzyme, we can attribute our observation to the effect of neutral drift, which creates potential new-function mutations that are then selected for and further merged to create new phenotypes. Moreover, the distributions of substitutions within each group (Fig. 2) indicate that, overall, group 2ber can be regarded approximately as the combination of groups 2be/2bee and 2br (except for position 182 which was not found in any 2ber enzyme, see below) implying some degree of modularity that can be exploited by evolution.

3.3. Function-stability tradeoffs operating within different modular networks

A number of experimental works have shown that removal of key catalytic residues usually results in stabilization of the protein while new-function substitutions usually produce a decrease in stability [3,11,29–31]. This function-stability tradeoff is a very important principle, since it implies that there will be an upper limit for the increase in functional complexity on a given protein structure, set by the stability threshold. Such limitation can be overcome by the incorporation of compensatory mutations that restore stability, allowing the protein to allocate new substitutions that are functionally beneficial but destabilizing and could not confer any fitness in the previous genetic background [3,32].

Residue Met182 of the TEM family is interesting in this regard since its substitution by Thr results in stabilization of the protein fold by 2–3 kcal/mol [11,33].[3] According to our dataset, the M182T substitution occurs in only 7.7% of the 2b variants (Fig. 2) while this percentage rises to 22.2% and 46.4% for 2be and 2bee variants, respectively. This behavior is in line with the fact that new-function mutations are usually destabilizing and thus prompt the addition of the stabilizing mutation. But surprisingly, the M182T

[3] The other highly substituted residues distant from the active site are Gln39, Ala42 and Gly92, which have not been studied in detail but whose substitutions have been assigned stabilizing roles as well [20].

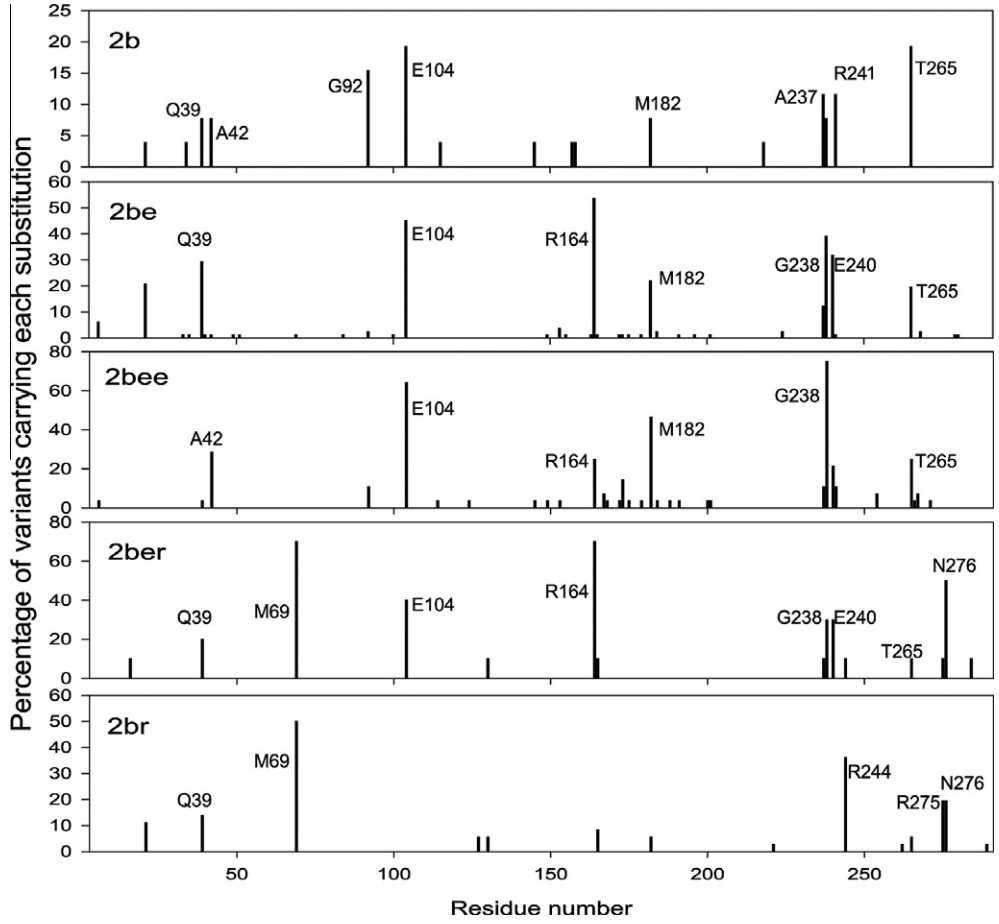

Fig. 2. Distribution of mutations in each phenotypic group. Fraction of counts for each mutation relative to all mutations (%) within each phenotypic group.

substitution shows up in none of the 2ber variants, even though they are expected to be the least stable since they are highly evolved and have a large number of substitutions. Given that the number of M182T substitutions is also lower than expected in 2br variants (5.6% against 7.7% for 2b and 22.2% for 2be), we suggest that it either cannot restore stability losses due to acquisition of mutations that give resistance to clavulanate and/or it interferes with those mutations blocking their capacity to confer such resistance. We thus suggest that M182T is not a truly "global" suppressor as generally regarded, and propose that evolution of 2ber enzymes has required and will require the introduction of other novel stability-restoring mutations. Indeed, directed evolution experiments have found substitutions other than M182T that could act as partial and "global" suppressors [34–35], although little is known about their interactions with new-function substitutions.

Substitution of Glu240 by Lys has also been suggested to rescue stability [20,36], and accordingly, it is more abundant in 2bee and 2ber (~30% in each group) than in 2b enzymes (0%). Furthermore, over 40% of the TEM variants with substitutions at Arg164, which broaden substrate spectrum and are destabilizing [11,36], are also substituted at Glu240. Notably, this stability-restoring substitution is much less common in variants with substitutions of either Glu104 or Gly238 (8% and 17%, respectively). On the contrary, the M182T substitution is common in variants with substitutions at Glu104 and Gly238 (33% and 30%, respectively, destabilizing TEM-1 by almost 2 kcal/mol [11]) but it is only half as common in variants with substitutions at Arg164 (17%). These observations suggest two approximately independent networks of sites in TEM proteins, which turn out to be similar to the two modular networks proposed with an independent approach [37]. Moreover, we can now suggest that each module consists of its (nearly) exclusive set of new-function and stabilizing substitutions. One such network involves Glu104 and Gly238, for which destabilizing substitutions are mainly rescued by the M182T mutation; and the other involves Arg164 with a stability-restoring role for the E240K substitution.

3.4. Predicting the activity of a TEM protein from its sequence

The problem of predicting protein function from sequence is one of the long-term goals in bioinformatics, with straightforward implications in understanding protein evolution and in facilitating protein and drug design. Predicting functional changes upon substitution within a family of enzymes is a less challenging goal that can serve as an intermediate step. Indeed, similar dataset-based approaches have recently been reported for three systems with a relatively large number of well-studied variants [38–40].

Here, the dataset of TEM variants with annotated functional profiles was employed to build a model relating a TEM protein's sequence with the phenotype it would confer to bacteria. We devised a simple scheme in which the query sequence is compared against all known sequences from the dataset, and its phenotype is predicted as that of the closest match in terms of the Euclidean distance in a 2-dimensional space made up by two "reduced properties" summed along all positions. The two "reduced properties" are the two principal components derived from a principal component analysis of 11 uncorrelated physicochemical properties of the amino acids (Table S3 and S4). Similar versions of "reduced" properties have been successfully utilized for property-based classification of proteins [41] and even to analyze relationships be-

tween the structure of the genetic code and the physicochemical properties of the amino acids [42].

The method was tested by splitting the dataset into 1000 different training and test subsets with sizes varying from 0.02: 0.98 to 0.98: 0.0 (fractions of entries included in the training: test subsets). Predictions were carried out for all entries in the test subsets based on each corresponding training subset, and the frequency of correctly assigned phenotypes was averaged around the corresponding fraction of variants used for training. The results (% of correct predictions against % of entries used in the training subset) are shown in Fig. 3, and are compared in Fig. S3 against the outcome of randomly assigning the phenotypic groups (black vs. red).

We found that our approach is capable of predicting phenotypes and, as expected, increasing the fraction of entries in the training set results in better predictions. Remarkably, the use of 40% of the entries for training already produces correct predictions for over 65% of the remaining sequences. In the leave-one-out limit (i.e. predicting the phenotype of each of the 182 variants in the dataset using the other 181 entries as training set) the global prediction accuracy is 76.4%, which is very close to that of other attempts in systems such as T4 lysozyme, HIV-1 protease, etc. [38–40] although our method is conceptually simpler and more transparent about the underlying physical basis for classification.

We also found that the prediction accuracies vary between phenotypic groups. In the leave–one–out scheme, the method has an outstanding performance for the three best represented groups: 2b, 2be and 2br (84.5, 79.3 and 94.4% correct predictions respectively). Instead, the prediction capability is moderate for group 2bee (57.1%) and poor for group 2ber (20%). The two latter groups are the smallest in the dataset and confer the most complex phenotypes; thus it is clear that a larger dataset is required to improve the prediction capability of the model. However, prediction of 2ber phenotypes might be improved by employing a radically different approach, based on the finding that these enzymes combine substitutions present in groups 2be/2bee and 2br as discussed above: the presence of essential features from groups 2be/2bee and 2br could be predicted independently, and the phenotype 2ber assigned if both kinds of features were identified.

Finally, it is important to bear in mind that the database was built from wild type-like and new-function variants, so only the neutral and positive effects of mutations are accounted for. Instead, the negative effects of substitutions at positions other than the catalytic triad are not modeled at all. Other neutral mutations and functionally deleterious mutations could be included in the dataset if available in the future and processed with the same methods, while the effect of mutations compromising stability could be assessed with programs for the prediction of stability changes upon mutation.

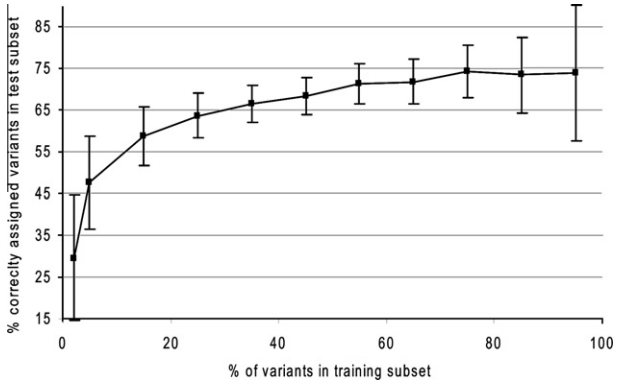

Fig. 3. Prediction of TEM phenotypes from sequence. Plot showing the percentage accuracy in the prediction of phenotype from sequence when the database of TEM variants is randomly split into training and test subsets of different sizes, as described in the text. More information in Fig. S3.

4. Conclusions

In this work we have assembled and analyzed a dataset of TEM-1-like and new-function TEM variants with <10 amino acid substitutions relative to the primordial TEM-1. Our approach allowed us to derive general and TEM-specific relationships between sequence, structure, function and stability which are not evident from isolated results on small numbers of variants. We could also show how such datasets can be used to build knowledge-based models to predict functional changes upon mutation in proteins. We forecast that similar analyses on other datasets will become more common in the future, contributing in novel ways to our understanding of proteins.

Acknowledgements

CONICET, EMBO and the Marie Curie Actions are acknowledged for postdoctoral fellowships to LAA. MLMS received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement 225167 (e-FLUX). PET is a staff member from IBR-CONICET. Prof. Alejandro Vila (IBR-CONICET, Argentina) is acknowledged for helpful discussions of the results and during manuscript preparation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2012.07.010.

References

[1] Pleiss, J. (2011) Protein design in metabolic engineering and synthetic biology. Curr. Opin. Biotechnol. 22, 611–617.
[2] Bloom, J.D., Labthavikul, S.T., Otey, C.R. and Arnold, F.H. (2006) Protein stability promotes evolvability. Proc. Nat. Acad. Sci. USA 103, 5869–5874.
[3] Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D.S. (2008) How Protein Stability and New Functions Trade Off. PLoS Comput. Biol. 4, e1000002.
[4] Medeiros, A.A. (1997) Beta-Lactamases: quality and resistance. Clin. Microbiol. Infect. 3 (Suppl 4), S2–S9.
[5] Matagne, A., Lamotte-Brasseur, J. and FrŠre, J.M. (1998) Catalytic properties of class A beta-lactamases: efficiency and diversity. Biochem. J. 330, 581–598.
[6] Bush, K. (2010) Alarming β-lactamase-mediated resistance in multidrug-resistant Enterobacteriaceae. Curr. Opin. Microbiol. 13, 558–564.
[7] Paterson, D.L. and Bonomo, R.A. (2005) Extended-spectrum beta-lactamases: a clinical update. Clin. Microbiol. Rev. 18, 657–686.
[8] Weinreich, D.M., Delaney, N.F., Depristo, M.A. and Hartl, D.L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312, 111–114.
[9] Tomatis, P.E., Fabiane, S.M., Simona, F., Carloni, P., Sutton, B.J. and Vila, A.J. (2008) Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility 1. Proc. Natl. Acad. Sci. USA 105, 20605–20610.
[10] Long-McGie, J., Liu, A.D. and Schellenberger, V. (2000) Rapid in vivo evolution of a beta-lactamase using phagemids. Biotechnol. Bioeng. 68, 121–125.
[11] Wang, X., Minasov, G. and Shoichet, B.K. (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J. Mol. Biol. 320, 85–95.
[12] Orencia, M.C., Yoon, J.S., Ness, J.E., Stemmer, W.P. and Stevens, R.C. (2001) Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nat. Struct. Biol. 8, 238–242.
[13] Abriata, L.A., Gonzalez, L.J., Llarrull, L.I., Tomatis, P.E., Myers, W.K., Costello, A.L., Tierney, D.L. and Vila, A.J. (2008) Engineered mononuclear variants in Bacillus cereus metallo-beta-lactamase BcII are inactive. Biochemistry 47, 8590–8599.
[14] Stemmer, W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370, 389–391.
[15] Partridge, S.R. and Hall, R.M. (2005) Evolution of transposons containing blaTEM genes. Antimicrob. Agents Chemother. 49, 1267–1268.
[16] Tomatis, P.E., Rasia, R.M., Segovia, L. and Vila, A.J. (2005) Mimicking natural evolution in metallo-beta-lactamases through second-shell ligand mutations. Proc. Natl. Acad. Sci. USA 102, 13761–13766.
[17] Barlow, M. and Hall, B.G. (2003) Experimental prediction of the natural evolution of antibiotic resistance. Genetics 163, 1237–1241.

[18] Baraniak, A., Fiett, J., Mrowka, A., Walory, J., Hryniewicz, W. and Gniadkowski, M. (2005) Evolution of TEM-type extended-spectrum beta-lactamases in clinical Enterobacteriaceae strains in Poland. Antimicrob. Agents Chemother. 49, 1872–1880.
[19] Datta, N. and Kontomichalou, P. (1965) Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. Nature 208, 239–241.
[20] Salverda, M.L.M., Arjan, G.M., de Visser, J. and Barlow, M. (2010) Natural evolution of TEM-1 beta-lactamase: experimental reconstruction and clinical relevance. FEMS Microbiol. Lett., 1015–1036.
[21] Thai, Q.K., Bos, F. and Pleiss, J. (2009) The Lactamase Engineering Database: a critical survey of TEM sequences in public databases. BMC Genomics 10, 390.
[22] Barlow, M. and Hall, B.G. (2002) Predicting Evolutionary Potential. In vitro evolution accurately reproduces natural evolution of the tem beta-lactamase. Genetics 161, 1355B–1355.
[23] Salverda, M.L.M., Dellus, E., Gorter, F.A., Debets, A.J.M., van der Oost, J., Hoekstra, R.F., Tawfik, D.S. and de Visser, J. (2011) Initial mutations direct alternative pathways of protein evolution. PLoS Genetics 7, e1001321.
[24] Hall, B.G. (2002) Predicting evolution by in vitro evolution requires determining evolutionary pathways. Antimicrob. Agents Chemother. 46, 3035–3038.
[25] Bush, K., Jacoby, G.A. and Medeiros, A.A. (1995) A functional classification scheme for beta-lactamases and its correlation with molecular structure. Antimicrob. Agents Chemother. 39, 1211–1233.
[26] Bush, K. and Jacoby, G.A. (2010) Updated functional classification of *f*-lactamases. Antimicrob. Agents Chemother. 54, 969–976.
[27] Ferreiro, D.U., Hegler, J.A., Komives, E.A. and Wollyness, P.G. (2011) On the role of frustration in the energy landscapes of allosteric proteins, Proc. Natl. Acad. Sci. USA. Epub ahead of print.
[28] Ambler, R.P. et al. (2011) A standard numbering scheme for the Class Abeta-lactamases. Biochem. J. 276, 269–270.
[29] Romero, P.A. and Arnold, F.H. (2009) Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. 10, 866–876.
[30] Shoichet, B.K., Baase, W.A., Kuroki, R. and Matthews, B.W. (1995) A relationship between protein stability and protein function. Proc. Natl. Acad. Sci. USA 92, 452–456.
[31] Beadle, B.M. and Shoichet, B.K. (2002) Structural bases of stability-function tradeoffs in enzymes. J. Mol. Biol. 321, 285–296.
[32] Bloom, J.D. and Arnold, F.H. (2009) In the light of directed evolution: pathways of adaptive protein evolution. Proc. Natl. Acad. Sci. USA, 9995–10000.
[33] Huang, W. and Palzkill, T. (1997) A natural polymorphism in b-lactamase is a global suppressor. Proc. Nat. Acad. Sci. USA 94, 8801–8806.
[34] Marciano, D.C., Pennington, J.M., Wang, X., Wang, J., Chen, Y., Thomas, V.L., Shoichet, B.K. and Palzkill, T. (2008) Genetic and structural characterization of an L201P global suppressor substitution in TEM-1 beta-lactamase. J. Mol. Biol. 384, 151–164.
[35] Brown, N.G., Pennington, J.M., Huang, W., Ayvaz, T. and Palzkill, T. (2010) Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM *f*-lactamases. J. Mol. Biol. 404, 832–846.
[36] Raquet, X., Vanhove, M., Lamotte-Brasseur, J., Goussard, S., Courvalin, P. and FrŠre, J.M. (1995) Stability of TEM beta-lactamase mutants hydrolyzing third generation cephalosporins. Proteins 23, 63–72.
[37] Guthrie, V.B., Allen, J., Camps, M. and Karchin, R. (2011) Network Models of TEM β-Lactamase Mutations Coevolving under Antibiotic Selection Show Modular Structure and Anticipate Evolutionary Trajectories. PLoS Comp. Biol. 7, e1002184.
[38] Masso, M. and Waisman, I.I. (2007) Accurate prediction of enzyme mutant activity based on a multibody statistical potential. Bioinformatics 23, 3155–3161.
[39] Chasman, D. and Adams, R.M. (2001) Predicting the functional consequences of non-synonymous single nucleotide polymorphisms. J. Mol. Biol. 307, 683–706.
[40] Dobson, P.D. and Doig, A.J. (2005) Predicting enzyme class from protein structure without alignments. J. Mol. Biol. 345, 187–199.
[41] Lapinsh, M., Gutcaits, A., Prusis, P., Post, C., Torbjörn, L. and Wikberg, J.E.S. (2002) Classification of G-protein coupled receptors by alignment-independent extraction of principal chemical properties of primary amino acid sequences. Protein Sci. 11, 795–805.
[42] Sjöstrom, M. and Wold, S. (1985) A multivariate study of the relationship between the genetic code and the physical-chemical properties of amino acids. J. Mol. Evol. 22, 272–277.