doi:10.1016/j.jmb.2006.01.095 J. Mol. Biol. (2006) 358, 280–288

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Mapping the Distribution of Conformational Information Throughout a Protein Sequence

Leopoldo G. Gebhard[1,2], Valeria A. Risso[1,2], Javier Santos[1,2] Raul G. Ferreyra[1,2], Martín E. Noguera[1] and Mario R. Ermácora[1,2]*

[1]*Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Sáenz Peña 180, (1876) Bernal Buenos Aires, Argentina*

[2]*Consejo Nacional de Investigaciones Científicas y Técnicas, Rivadavia 1917 (1033) Ciudad Autónoma de Buenos Aires, Argentina*

**Corresponding author*

The three-dimensional structure of protein is encoded in the sequence, but many amino acid residues carry no essential conformational information, and the identity of those that are structure-determining is elusive. By circular permutation and terminal deletion, we produced and purified 25 *Bacillus licheniformis* β-lactamase (ESBL) variants that lack 5–21 contiguous residues each, and collectively have 82% of the sequence and 92% of the non-local atom–atom contacts eliminated. Circular dichroism and size-exclusion chromatography showed that most of the variants form conformationally heterogeneous mixtures, but by measuring catalytic constants, we found that all populate, to a greater or lesser extent, conformations with the essential features of the native fold. This suggests that no segment of the ESBL sequence is essential to the structure as a whole, which is congruent with the notion that local information and modular organization can impart most of the tertiary fold specificity and cooperativity.

Keywords: protein folding; sequence patterns; conformational information; folding code; folding units

Introduction

The conformation of a protein is encoded in the sequence,[1] and unraveling the folding code has become one of the greatest challenges to contemporary science. A major difficulty in solving this problem is the lack of obvious correspondence between sequence and three-dimensional structure: proteins tolerate a surprisingly large number of mutations;[2–8] there are several examples of non-homologous proteins belonging to the same structural class;[9] proteins with highly similar sequence but different fold have been created;[10] and identical sequences of 8–11 residues adopt different structures in different proteins.[11,12]

To elucidate the logic of folding it is necessary to establish if there are structure-determining sequence positions. However, even though sequence homology is the most powerful predictor of tertiary structure, exhaustive examination of natural and experimentally mutated proteins has so far failed to identify prototypic sequence signatures characteristic of each kind of protein fold. With respect to this, the evidence suggests that the amino acid alphabet may be simplified,[8] and even simple patterns of hydrophobic and hydrophilic residues may promote native-like folding and efficient core packing.[13] Furthermore, many residues may carry no global conformational information, as can be concluded from statistical analyses,[14] and the observation that some protein fragments do achieve native-like folds.[15–25]

With the above considerations in mind, we devised an experimental procedure for the identification of sequence segments that may carry essential conformational information. The procedure, which is a refinement and extension of our previous work,[25,26] involves terminal truncation of circularly permuted proteins. The approach allows sliding a deletion window along the chain, and the resulting abridged variants reveal the structural consequences of lacking specific parts of the sequence. Unlike conventional site-specific mutagenesis, this segmental deletion can be used to switch off interactions involving main chain atoms.

As a proof of principle, we performed segmental deletion on the two-domain, 264-residue protein *Bacillus licheniformis* β-lactamase[27] (ESBL; Figure 1). The first domain, which is made of middle-chain residues, exhibits a central α-helix surrounded by α

Abbreviations used: BP, benzyl penicillin; cp, circularly permuted; ESBL, *B. licheniformis* exo small β-lactamase; NC, nitrocefin; SEC, size-exclusion chromatography.

E-mail address of the corresponding author: ermacora@mail.unq.edu.ar

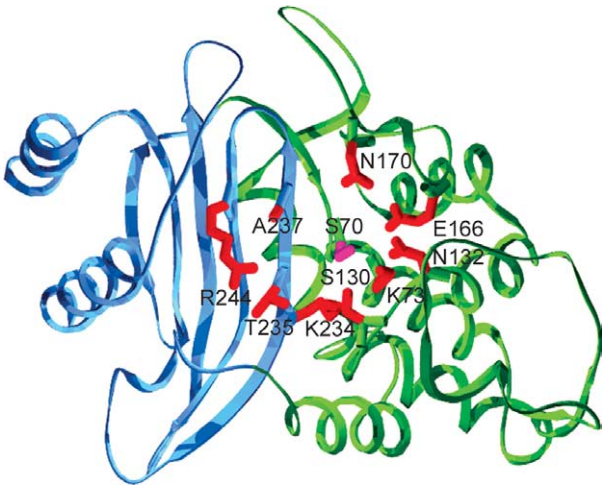

Figure 1. Molecular model of ESBL. The α and α+β domains are shown in green and blue, respectively. The residues that participate in catalysis are indicated. The Figure was prepared using Swiss-PDBViewer (http://www.expasy.org/spdbv/) and POV-Ray™ (http://www.povray.org/).

and 3_{10} helices. The second domain contains a five-stranded, antiparallel β-sheet plus the N and C-terminal α-helices. The general folding properties of ESBL have been characterized.[25,28]

A total of 25 ESBL variants carrying deletions that cover almost the entire sequence were prepared and purified. *In vitro* enzymic activity was found to be a convenient probe for the detection of native-like structure in the presence of vast amounts of misfolded protein. Circular dichroism (CD) spectroscopy and hydrodynamic techniques were used as complements to further assess the structural consequences of the deletions. Contrary to expectation, the results suggest that no part of ESBL sequence carries essential conformational information. This finding is best explained by a perspective of folding in which the amino acid positions are only locally coupled and the protein matrix is built by the assembly of self-organizing modules.

Results

Expression and global properties of ESBL deletion variants

Combining circular permutation and terminal deletion, we produced ESBL variants lacking fragments of 5–21 residues along the sequence up to a formal coverage of 82% of the residues (Figure 2; the nomenclature of ESBL variants is explained in Table 1). To represent the structural consequences of the deletions, all contacts between residues separated in sequence by four or more positions were calculated from the crystallographic model (PDB entry 4blm). With a 4.5 Å cutoff, there are 2897 atom–atom contacts involving 536 residue–residue interactions. The distribution of contacts removed by each truncation is shown in Figure 3. Typically, in the network of tertiary contacts, each mutant lacks one node that ramifies to three to five elements of distant tertiary structure and establishes more than 100 atom–atom contacts. Summing over all the variants, 92% of the non-local atom–atom contacts were removed.

The ESBL variants could be expressed with good yields (10–300 mg/l of culture) and purified to homogeneity. In roughly half of the cases, significant amounts of insoluble product accumulated. Interestingly, the variants with deletions involving the first 37 or the last 41 amino acid positions were the most insoluble.

In the far-UV region, nearly all ESBL variants exhibit CD spectra with native-like shape and moderate variations in signal intensity likely to arise from the intrinsic contribution of the deleted residues to the native signal (Figure 4). Cp276\282, the only exception to this trend, shows a spectrum suggestive of a large change in secondary structure. In the near-UV, the changes between mutants are more conspicuous: cp64, cp255, and cp267\276, have CD spectra with native-like shape and intensity, several have native-like shapes but decreased intensities; and nearly half show spectra with little or no structure.

The optical results show that the average secondary structure content of ESBL is remarkably resistant to segmental deletion. In contrast, the average tertiary structure of most ESBL variants was disrupted significantly, which is consistent with the observed tendency of this protein to populate partially folded states.[25,28]

Molecular size and aggregation state analysis by size-exclusion chromatography (SEC) also revealed dissimilar behaviors: six variants are exclusively monomeric and as compact as the wild-type protein (cp217\226, cp163\178, cp227\231, cp267\276, cp64, cp255); four yield compact monomeric species along with different amounts of aggregates and expanded monomers (cp78\86, cp96\116,

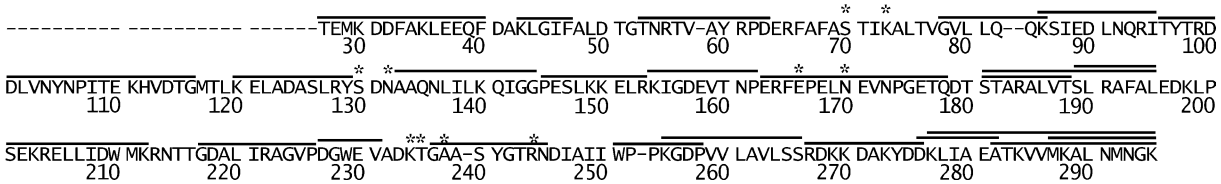

Figure 2. Sequence coverage. Bars encompass the residues deleted in each ESBL variant. Numbering is that proposed by Ambler.[56] Asterisks show catalytically important residues.

Table 1. Catalytic parameters of ESBL variants

Variant	Deletion (residues)	k_{cat}[a] (s^{-1})	K_m (M)	k_{cat}/K_m ($s^{-1}\,M^{-1}$)
cp43\47[b]	5	5.9×10^{0}	2.5×10^{-4}	2.4×10^{4}
cp53\63	10	1.9×10^{0}	4.2×10^{-4}	4.5×10^{3}
cp78\86	7	1.2×10^{1}	3.4×10^{-4}	3.5×10^{4}
cp86\95	10	9.0×10^{-2}	1.4×10^{-4}	6.4×10^{2}
cp96\116	21	1.9×10^{2}	2.6×10^{-4}	7.3×10^{5}
cp120\129	10	2.8×10^{0}	3.0×10^{-4}	9.3×10^{3}
cp133\144	12	9.0×10^{-2}	2.8×10^{-4}	3.2×10^{2}
cp145\153	9	2.2×10^{-2}	1.2×10^{-4}	1.8×10^{2}
cp154\162	9	4.0×10^{-2}	5.9×10^{-4}	6.8×10^{1}
cp163\178	16	1.7×10^{-1}	2.0×10^{-4}	8.5×10^{2}
cp181\188	8	1.8×10^{0}	4.8×10^{-4}	3.8×10^{3}
cp181\195	15	5.0×10^{-2}	8.0×10^{-5}	6.2×10^{2}
cp189\195	7	2.5×10^{-1}	2.3×10^{-4}	1.1×10^{3}
cp201\212	12	6.6×10^{0}	2.6×10^{-4}	2.5×10^{4}
cp217\226	10	1.1×10^{1}	1.9×10^{-3}	5.8×10^{3}
cp227\231	5	4.0×10^{1}	4.0×10^{-4}	1.0×10^{5}
cp236\245	9	3.0×10^{-2}	3.2×10^{-4}	9.4×10^{1}
cp251\258	7	7.0×10^{0}	1.6×10^{-4}	4.4×10^{4}
cp255\266	12	7.8×10^{0}	2.2×10^{-4}	3.5×10^{4}
cp267\276	10	3.7×10^{2}	1.2×10^{-4}	3.1×10^{6}
cp276\282	7	1.2×10^{0}	1.9×10^{-4}	6.3×10^{3}
Δ27\40	14	2.8×10^{2}	1.8×10^{-4}	1.6×10^{6}
Δ287\295	9	5.0×10^{2}	1.6×10^{-4}	3.1×10^{6}
Δ282\295	14	1.5×10^{-1}	3.0×10^{-4}	5.0×10^{2}
Δ277\295	19	9.5×10^{-2}	2.5×10^{-4}	3.8×10^{2}
cp64	0	1.4×10^{3}	9.0×10^{-5}	1.6×10^{7}
cp255	0	2.5×10^{3}	1.2×10^{-4}	2.1×10^{7}
Wild-type	0	2.4×10^{3}	1.3×10^{-4}	1.8×10^{7}

[a] Averages of two, three or four independent measurements are reported. The standard deviation of k_{cat} and K_m was $<20\%$.
[b] ESBL variants names are as follows: the prefix cp indicates circular permutation; Δ is used for abridged variants with wild-type connectivity; two numbers separated by\identify the N and C terminus of the removed sequences; single numbers are used to identify the new N terminus of full-length, circularly permuted variants; amino acid numbering is the consensus for class A β-lactamases.[56]

cp236\245, Δ287\295), and the remaining populate dimeric and aggregated states only (not shown).

The specific effects of the pentaglycine connection between residues 27 and 295 on ESBL conformation were assessed by examining cp64 and cp255. Judging from the SEC and CD data, the conformation of these two full-length, circularly permuted variants is identical with that of the wild-type protein. This was confirmed by the catalytic parameters (Table 1). Thus, the common pentaglycine connection *per se* does not alter the fold of ESBL significantly and can be considered neutral as regards to conformational information.

Optical and hydrodynamic data concurrently show that the abridged variants tend to form complex mixtures in which altered conformations predominate. Therefore, to assess whether they retain the capacity to fold to native-like states, we resorted to a more sensitive and specific probe.

Catalysis as a probe for native folding

Catalysis depends strongly on conformational details,[29] and inactivation precedes or accompanies early conformational changes during protein unfolding.[30] Moreover, catalytic activity can monitor folding with very high levels of sensitivity and specificity because the signal proceeds only from the native conformation. This makes feasible the detection of traces of native fold in preparations containing large amounts of unfolded and/or misfolded protein.

In ESBL, residues involved directly in catalysis and/or substrate binding[31,32] are particularly well suited to monitor the relative spatial arrangement and conformational integrity of the two protein domains (Figure 1). In the α domain, the O^{γ} atom of S70 is activated for nucleophilic attack on the lactam carbonyl group to yield an acyl enzyme intermediate, and E166 participates as a general base in the subsequent deacylation step. The role of K73 is in facilitating the protonation of the lactam leaving group in acylation and of the O^{γ} leaving group in deacylation. This is most likely achieved through electrostatic effects on other active site residues, like S130, and by the transfer of a proton to E166 *via* a fixed water molecule. N132 and N170 contribute to substrate binding by making two hydrogen bonds to the carbonyl amide group on lactam C10.

On the other hand, several α + β-domain residues participate in the catalytic mechanism. Main-chain atoms of A237 establish two hydrogen bonds to C6 and C7 lactam substituents and, along with the three hydrogen bonds joining R244 and T235 to the carboxylate at C2, ensure that the substrate covers the Ser70 O^{γ} with the target carbonyl group placed properly on the enzyme oxyanion hole. There, nitrogen atoms from A237 and S70 polarize the carbon–oxygen bond involved in the formation of the tetrahedral intermediate. In addition, K234 makes a contribution to the active site hydrogen

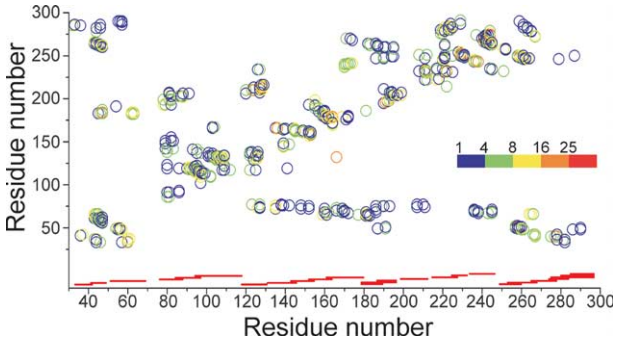

Figure 3. Pairs of residues that contact each other in the native structure of ESBL. Units are residue number. Residues separated by four or more positions in sequence are considered a pair if they make contact through atoms separated by less than 4.5 Å. Colors indicate how many atom–atom contacts each pair of residues establishes. Red bars indicate the segment of sequence removed in each truncated variant.

bond network and, along with S130 and T235, fixes the substrate, through lacta N4 and the carboxylate group at C3, in the right position for catalysis.

Furthermore, the precise three-dimensional register of active site residues depends on the integrity of several elements of secondary and tertiary structure from both domains: (a) S70 and K73 monitor α-helix 71–83, the central, buried "master column" of the α-domain; (b) K234, T235, A237, and R244 survey the central, five-stranded β-sheet in the α+β domain; (c) E166 and N170 report on the conformation of the Ω loop; and (d) S130 and N132 sense the loop connecting helices 120–128 and 132–140.

The entangled nature of the active site and the number and the intricacy of its geometric constraints indicate that the conformational integrity of most of the two domains is required to sustain catalysis, and that ESBL can be enzymically active only if folded in a native-like fashion. Further evidence confirms that catalytic activity can be used to monitor the native fold of ESBL as a whole: (a) spectroscopic, hydrodynamic, and chemical modification experiments demonstrated that the ESBL domains do not unfold independently during urea-induced unfolding experiments at equilibrium (J.S. *et al.*, unpublished results); and (b) isolated recombinant ESBL fragments corresponding to the α and α+β domains have a strong tendency to aggregate when refolded from urea solutions and are enzymically inactive (V.A.R. & M.R.E., unpublished results).

We considered how to distinguish true enzymic activity from uncatalyzed hydrolysis or from the reaction with isolated nucleophiles. Fortunately, an enlightening guide to enzymic activity thresholds could be found in previous works.[3,33] At micromolar concentrations, compared with uncatalyzed reactions, typical enzymes increase reaction rates by 2–17 orders of magnitude.

In good agreement with previous reports,[34] we determined that the uncatalyzed first-order rate for benzyl penicillin (BP) hydrolysis in buffer A (100 mM sodium phosphate, (pH 7.0 at 25 °C)) is $2.6\times10^{-7}\ s^{-1}$. Also, we found that the rate in the presence of lysozyme or bovine serum albumin (BSA) is similar to that in buffer alone (Table 2). These two proteins are large and diverse enough to represent typical protein surfaces.

We did not use active site mutants of ESBL as negative controls of enzymic activity because mutants lacking "essential" catalytic residue may still perform catalysis through alternative mechanisms and surrogate groups.[35,36] For instance, an S70A

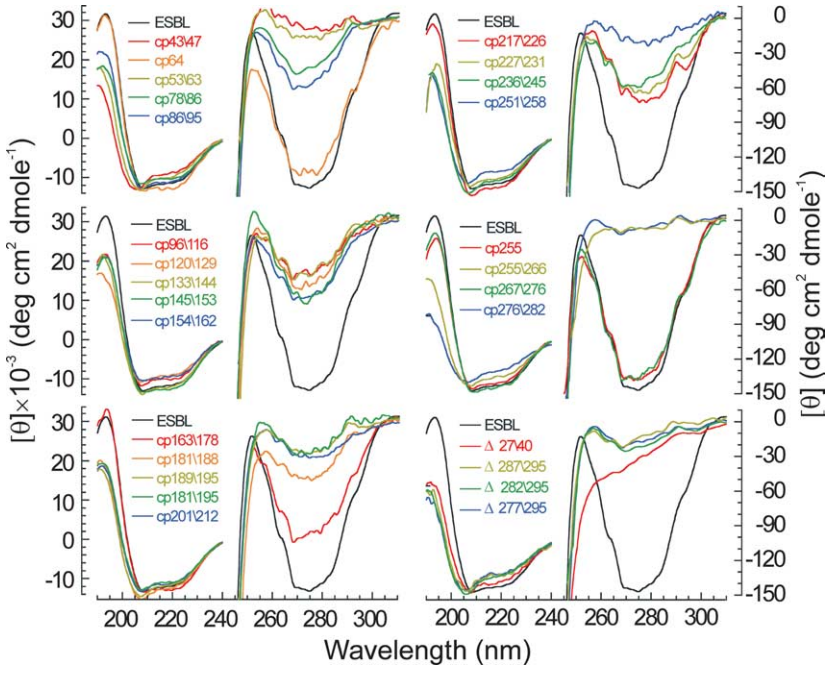

Figure 4. Far-UV and near-UV spectra. The variants were refolded by dialysis against buffer A before analysis (see Methods).

Table 2. Reference values for β-lactamase activity

Reagent	Rate ($M^{-1} s^{-1}$)	Typical velocity[a] ($M^{-1} s^{-1}$)	Ratio[b]
Hydroxide[c]	1.5×10^{-1}	3.1×10^{-10}	5.8×10^{-1}
Reaction buffer[d]	–	5.2×10^{-10}	1.0×10^{0}
Lysozyme[e]	2.2×10^{-1}	4.4×10^{-10}	8.5×10^{-1}
BSA[e]	1.5×10^{-1}	3.0×10^{-10}	5.7×10^{-1}
ESBL(Cya_2)[e]	4.4×10^{-1}	8.8×10^{-10}	1.7×10^{0}
cp154\162	6.8×10^{1}	1.4×10^{-7}	2.6×10^{2}
Wild-type ESBL	1.8×10^{7}	3.6×10^{-2}	6.9×10^{7}
S70A SAGBL[f]	2.0×10^{2}	4.3×10^{-7}	8.4×10^{2}
Wild-type SAGBL[f]	2.8×10^{6}	5.6×10^{-3}	1.1×10^{7}

[a] Calculated as $v=[\text{Reagent}][\text{BP}]k$, where $[\text{Reagent}]=1\times10^{-6}$ M, $[\text{BP}]=2\times10^{-3}$ M and k is the second-order rate; or, for reaction buffer, as $v=[\text{BP}]k$, where k is the apparent first-order rate.
[b] Quotient of reactive group velocities and reaction buffer velocity.
[c] In water, at 30 °C.[34]
[d] This work.
[e] Calculated by measuring hydrolysis rate as a function of protein concentration.
[f] *Streptomyces albus* G β-lactamase.[38]

variant of ESBL, which lacks the nucleophile that form the intermediate acyl-enzyme complex, showed an increased rate of substrate hydrolysis compared with the reaction in buffer alone.[37] Similarly, S70A *Streptomyces albus G* β-lactamase exhibited substantial activity (Table 2).[38] Instead of active site mutants, we assayed ESBL($Cya)_2$, a variant of ESBL with sulfonic acid at positions 126 and 265, which is intended to be permanently unfolded (J.S. *et al.*, unpublished results). As expected, BP hydrolysis in the presence of ESBL($Cya)_2$ showed no Michaelian behavior and yielded a linear plot up to 2 mM substrate (Figure 5). From the slope of the plot, an apparent second-order rate constant of $0.44\ M^{-1} s^{-1}$ was calculated (Table 2).

Since the rates of BP hydrolysis by ESBL($Cya)_2$, lysozyme, and BSA are similar, it can be concluded safely that the rate of uncatalyzed hydrolysis, either by solvent or by protein residues in unspecific arrays, is not greater than $0.44\ M^{-1} s^{-1}$. On the other extreme, we found that k_{cat}/K_m for wild-type ESBL is $1.8\times10^{7}\ M^{-1} s^{-1}$. These values set a natural scale for BP hydrolysis spanning eight orders of magnitude.

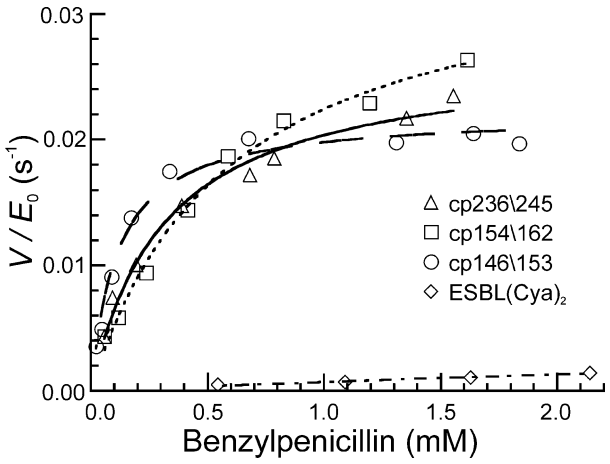

Figure 5. Examples of reaction rate against substrate concentration curves. The three circularly permuted and truncated ESBL variants with the lowest k_{cat}/K_m values are shown (see Table 1) along with ESBL($Cya)_2$, the unfolded negative control for enzymic catalysis (see the text).

Lastly, the catalytic parameters for all the abridged β-lactamases were determined from Michaelis–Menten curves (Table 1; Figure 5). To judge the results properly, the following considerations should be taken into account. (a) Reported k_{cat}/K_m values are lower limits because most of the lactamase preparations contain significant amounts of inactive protein. (b) Although only results for BP hydrolysis are shown, specific activity toward BP and nitrocefin (NC) was comparable. (c) Cross-contamination between ESBL variants was avoided by using new chromatographic matrix and buffers in each purification. (d) A dummy chromatography of a control homogenate of cells expressing an unrelated protein demonstrated that no enzymically active contaminants co-purify with ESBL (not shown). (e) The rate of BP hydrolysis by a raw control bacterial homogenate was similar to that promoted by buffer A alone (not shown).

Considering both the activity of existing enzymes and the activity scale established above, it seems reasonable to assume that true lactamase-catalyzed reactions will have second-order rates at least two orders of magnitude larger than those from the control proteins. By this criterion, we found that all the truncated variants prepared for this work display genuine enzymic activity (Figure 6). Since the observed enzymic activity is unlikely to be the result of a novel ESBL fold induced by the truncations, and given the relationship between catalytic efficiency and tertiary structure discussed above, we must conclude that all the preparations of truncated ESBL variants contain, to a greater or lesser extent, molecules with the essential features of the ESBL native fold.

Discussion

The ESBL folding code is robust

Most of the circularly permuted and abridged ESBL variants prepared for this work exhibit

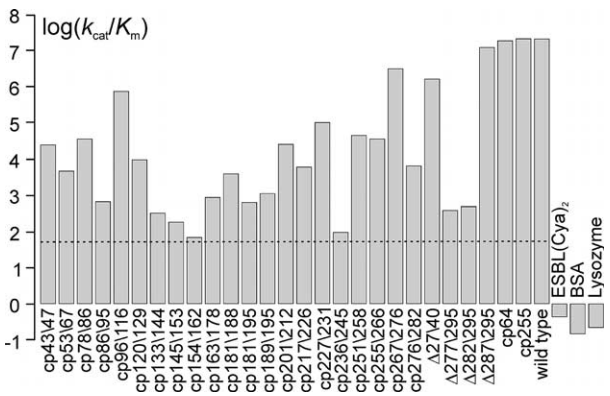

Figure 6. Catalytic activity. The dotted line is the assumed threshold for genuine enzymic activity (see the text and Tables 1 and 2).

conformational heterogeneity, a pronounced tendency to form aggregates, and low refolding yield, but all of them retain the ability to fold into a native-like, enzymically active conformation. Since these variants lack segments of sequence, significant amount of specific tertiary contacts, and wild-type backbone extremes, this implies a conformational code that is robust against elimination of information, changes in chain connectivity, and alteration in the relative location of information units. Moreover, since the applied deletion scheme perturbs the entire network of long-range, native contacts, it follows that the instructions for ESBL structure cannot be in the form of a unique pattern of essential interactions distributed along the sequence. Likewise, since atom–atom contacts are a consequence of the type of residue at each sequence position, the three-dimensional specificity must be conveyed independently of any particular detail of the sequence.

It would not be possible to reach the above conclusions from residue substitution experiments or evolutionary variability analysis alone because: (a) mutations modify but do not eliminate the conformational information at a given sequence position; (b) sequence positions that are highly tolerant to amino acid substitution still may carry significant non-local conformational information; (c) invariant or highly conserved residues may be non-essential for the global conformation and reflect only local structural constraints; and (d) the combinatory nature of compensating mutations precludes an exhaustive experimental exploration of all possible sequences related to the same fold.

The observed robustness of ESBL folding code imposes restrictions on the arrangement of conformational information over the length of the sequence. If robustness were achieved by redundancy, there would be sets of complete folding instructions stored in different sites along the chain. However, using Radar,[39] we found no internal homology in ESBL and, thus, the existence of redundant blocks of information in it seems unlikely.

A more parsimonious way to explain ESBL robustness is modular construction. In the hierarchic model of folding,[40] the structure determinants are local in sequence, long-range contacts only consolidate pre-existing local structure, and non-local, three-dimensional specificity results from the binding of independently ordered modules. By definition, hierarchic models are tolerant to segmental deletion, because the associated conformational disruption can be circumscribed to particular modules without precluding the folding and assembly of the others.

The robustness of ESBL code is more difficult to justify by assuming models of folding that propose the growth of tertiary structure by concerted consolidation of secondary and tertiary interactions from a nucleation site.[41] Segment deletion eliminates simultaneously hundred of long-range contacts, including those established by main chain atoms (Figure 3). If the folding of ESBL were a process dependent on the specific and concerted formation of a significant number of tertiary contacts, the deletion scheme would have identified several variants incapable of folding. Yet, formally, the possibility exists that multiple alternative routes for folding may allow these mutants to compensate for the loss of any particular set of tertiary contacts. This, however, would be possible only if ESBL contained a large number of potential nucleation sites.

To our best knowledge, the results hereby reported for ESBL constitute the first verification over an entire sequence of the robustness of the folding code. Since ESBL is representative of medium-sized and relatively complex protein architecture, it remains to be seen if the folding code is also robust in small, single-domain proteins with a simpler design. Nevertheless, the fact that diverse abridged proteins showing native-like folding could be produced suggests that ESBL is not an exception.[15–24] Admittedly, there are many reports of folding impairment as a consequence of circular permutation, deletions, and even point mutations. From systematic circular permutation, Iwakura *et al*.[42] have concluded that DHFR contains ten sequential elements essential for folding, which is seemingly at odds with the picture that emerges from the experiments with ESBL. However, as discussed above, to ponder on the evidence, it is important to separate conceptually the sequence changes that may cause instability, aggregation, affect kinetics, or otherwise reduce the yield of native structure, from those that may eliminate altogether the possibility of native fold.[43] If what is under examination is the specificity of folding, then negative results should be considered inconclusive unless the possibility of native-like folding, even at trace levels, is explicitly discarded. Failure to do so may still allow learning about the structural determinants of folding stability and efficiency, but will not clarify the logic of the folding code.

Concerning this, we believe that many of the studies reporting folding incapacity have not been properly designed to detect the presence of small quantities of natively folded molecules among a large excess of misfolded products.

A revealing example further highlights the need to distinguish between fold specificity and stability. The 1–103 fragment of the 149-residue protein staphylococcal nuclease appears unstructured by NMR analysis. However, if stabilized by mutations V66L and G88V, this fragment adopts a stable tertiary structure that is a close variation of the OB-fold found in the X-ray structure of the parent protein.[44] Thus, it cannot be assumed that the fragment is unable to fold in the absence of the 46 C-terminal residues.

The modular design of protein structure

The ability of the permutation–truncation approach to switch off elements of tertiary structure is perhaps its most interesting feature. Circularly permuted, truncated proteins can be considered as designed folding intermediates, in which specific elements of tertiary structure are unfolded by elimination of sequence. In this regard, as probes for folding reach residue resolution and more detailed analysis are performed, the hidden multistate nature of protein molecules becomes more apparent. The picture that emerges from equilibrium studies by NMR spectroscopy in conjunction with hydrogen exchange shows elements of native structure whose transient unfolding does not depend on global unfolding.[45,46] Thus, our results and the latest direct structural information concur to challenge established ideas on long-range cooperativity of protein structure.

Switching off structural elements by sequence deletion is also conceptually close to the computational approach to macromolecular equilibrium advocated by Hilser and colleagues.[47] Those authors proposed that the native state is a statistical ensemble of conformations that originate from the existence of local unfolding transitions throughout most of the protein molecule. Their calculations, based on segmental unfolding, achieved remarkable success in predicting experimental hydrogen-exchange protection factors. From the informational standpoint, the experiments with ESBL and the independent evidence mentioned above support the contention that no essential subset of tertiary interaction encodes the native structure, and that the native protein matrix is an intrinsically modular construction that is determined locally.

Since its introduction,[48,49] the concept of modular protein design has been the inspiration for many investigations. In particular, several authors have undertaken the characterization of autonomous folding units by a variety of approaches.[50–53] By definition, folding units are made of contiguous residues and contain elements of secondary and local tertiary structure. In solution, the predominant conformation of these units would be similar to that acquired in the native protein; but it is possible that the mutually stabilizing association between units may select alternative, less-well populated conformations. Thus, there would be no formal contradiction between modular assembly and conformational ambiguity. In agreement with this idea, prediction methods in which structure sampled by local sequence is approximated by the distribution of prototype segments structures in data banks achieved impressive success.[54]

In sum, the present results demonstrate that the ESBL fold is resistant to sequence deletion to an extent that precludes the existence of discrete patterns of amino acid positions essential for specifying the overall three-dimensional structure. Thus, no particular network of inter-atomic interactions is required to define the fold. This property of ESBL is congruent with experimental evidence and theoretical considerations suggesting that local information and modular organization can impart most of the tertiary fold specificity. If this is a property of proteins in general, theoretical and empirical studies on protein folding will be bolstered by a huge reduction in the need for sequence space exploration.

Methods

General details

BP and NC were from Sigma (St. Louis, Missouri) and Calbiochem (La Jolla, CA), respectively. Protein purity was assessed by SDS-PAGE. CD and SEC were carried out as described.[25]

Preparation of ESBL variants

Site-specific and abridged mutants of ESBL with wild-type chain connectivity were prepared essentially as described.[25] The general strategy for circular permutation was that applied to DHRF by Iwakura *et al*.[42] Briefly, a repeat of five Gly codons was ligated between two copies of the DNA coding for ESBL, and the construction was the template for PCR amplification of the segmentally deleted variants. Primers included restriction sites for cloning into pET9a (Novagen, Madison, USA). The absence of unwanted mutations was confirmed by DNA sequencing (IIB, UNSAM, Argentina). To prepare ESBL(Cya)$_2$, S126C–S265C ESBL, which has 80% of the specific activity of wild-type ESBL and one inaccessible cysteine residue per protein domain (J.S. *et al*., unpublished results), was oxidized with performic acid,[55] dialyzed against buffer A, and freed from aggregates by filtration through 0.1 μm pore size filters.

Escherichia coli BL21 (DE3) cells transformed with pET9a carrying the modified ESBL genes were cultured in Luria-Bertani medium at 37 °C to $A_{600nm} \sim 1.0$, then induced with 1 mM IPTG (3 h), and finally harvested by centrifugation. ESBL, S126C–S265C ESBL, and circularly permuted 255 ESBL (cp255) were purified as described.[28] The variants that accumulated with good yield in inclusion bodies were purified as described.[25] The rest, which were soluble but unstable, conformationally altered, or prone to aggregation, could not be purified by conventional protocols. For these variants, we applied the following purification procedure. First, cell homogenates were subjected to a short incubation (10 min, 20 °C)

with 5 M urea, 5 mM glycine, 25 mM phosphoric acid (pH 3.5). We discovered serendipitously that this treatment causes bulk precipitation of most bacterial proteins and leaves in solution the unstably folded ESBL variants close to purity. Second, the proteins so isolated were subjected to ionic exchange chromatography under denaturing conditions.[25,28] In all cases, protein refolding was done at 4 °C by dialysis against 50 mM sodium phosphate (pH 7.0), and particulate material was eliminated by centrifugation.

Enzyme kinetics

Substrate hydrolysis was monitored by UV absorption at 240 nm (BP; $\Delta\varepsilon = 570\ M^{-1}\ cm^{-1}$) or 486 nm (NC; $\Delta\varepsilon = 20000\ M^{-1}\ cm^{-1}$) with a Shimadzu UV-160A spectrophotometer (Shimadzu, Japan). All the assays were performed at 25 °C in buffer A. Reactions were initiated by adding 50 µl of 0.01–25 µM β-lactamase to 450 µl of substrate (2.0 mM BP or 0.1 mM NC). Initial rates were calculated from the change in absorbance that ensued from the consumption of no more than 10% of the substrate. k_{cat} and K_m were calculated by fitting the Michaelis–Menten equation to velocities at various concentrations of BP. Proteins unrelated to β-lactamase (0.05–5.0 mM) were assayed. Non-enzymic hydrolysis of BP was determined by incubating the antibiotic in buffer A (3 days; at 25 °C in the dark). Two, three or four independent measurements were performed for each protein and controls.

Acknowledgements

We thank Professor Anthony Fink for the lactamase gene and many suggestions on this protein. This work was supported by grants from CONICET, UNQ, and ANPCyT.

References

1. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.
2. Palzkill, T. & Botstein, D. (1992). Probing beta-lactamase structure and function using random replacement mutagenesis. *Proteins: Struct. Funct. Genet.* **14**, 29–44.
3. Axe, D. D., Foster, N. W. & Fersht, A. R. (1998). A search for single substitutions that eliminate enzymatic function in a bacterial ribonuclease. *Biochemistry*, **37**, 7157–7166.
4. Wen, J., Chen, X. & Bowie, J. U. (1996). Exploring the allowed sequence space of a membrane protein. *Nature Struct. Biol.* **3**, 141–148.
5. Rennell, D., Bouvier, S. E., Hardy, L. W. & Poteete, A. R. (1991). Systematic mutation of bacteriophage T4 lysozyme. *J. Mol. Biol.* **222**, 67–88.
6. Matthews, B. W. (1993). Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **62**, 139–160.
7. Gassner, N. C., Baase, W. A. & Matthews, B. W. (1996). A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl Acad. Sci. USA*, **93**, 12155–12158.
8. Miller, O. J. & Dalby, P. A. (2004). Exposing relationships using directed evolution. *Trends Biotechnol.* **22**, 203–205.
9. Orengo, C. A., Jones, D. T. & Thornton, J. M. (1994). Protein superfamilies and domain superfolds. *Nature*, **372**, 631–634.
10. Alexander, P. A., Rozak, D. A., Orban, J. & Bryan, P. N. (2005). Directed evolution of highly homologous proteins with different folds by phage display: implications for the protein folding code. *Biochemistry*, **44**, 14045–14054.
11. Sudarsanam, S. (1998). Structural diversity of sequentially identical subsequences of proteins: identical octapeptides can have different conformations. *Proteins: Struct. Funct. Genet.* **30**, 228–231.
12. Minor, D. L., Jr & Kim, P. S. (1996). Context dependent secondary structure formation of a designed protein sequence. *Nature*, **380**, 730–734.
13. Moffet, D. A. & Hecht, M. H. (2001). *De novo* proteins from combinatorial libraries. *Chem. Rev.* **101**, 3191–3203.
14. Rackovsky, S. (1993). On the nature of the protein folding code. *Proc. Natl Acad. Sci. USA*, **90**, 644–648.
15. Rico, M., Jimenez, M. A., Gonzalez, C., De Filippis, V. & Fontana, A. (1994). NMR solution structure of the C terminal fragment 255 316 of thermolysin: a dimer formed by subunits having the native structure. *Biochemistry*, **33**, 14834–14847.
16. Patrick, W. M. & Blackburn, J. M. (2005). *In vitro* selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. *FEBS J.* **272**, 3684–3697.
17. Trevino, R. J., Gliubich, F., Berni, R., Cianci, M., Chirgwin, J. M., Zanotti, G. & Horowitz, P. M. (1999). NH_2-terminal sequence truncation decreases the stability of bovine rhodanese, minimally perturbs its crystal structure, and enhances interaction with GroEL under native conditions. *J. Biol. Chem.* **274**, 13938–13947.
18. Takahashi, K., Noguti, T., Hojo, H., Ohkubo, T. & Go, M. (2001). Conformational characterization of designed minibarnase. *Biopolymers*, **58**, 260–267.
19. De Sanctis, G., Falcioni, G., Polizio, F., Desideri, A., Giardina, B., Ascoli, F. & Brunori, M. (1994). Mini myoglobin: native like folding of the NO derivative. *Biochim. Biophys. Acta*, **1204**, 28–32.
20. Kim, K., Cistola, D. P. & Frieden, C. (1996). Intestinal fatty acid-binding protein: the structure and stability of a helix-less variant. *Biochemistry*, **35**, 7553–7558.
21. Rose, T., Brune, M., Wittinghofer, A., Le Blay, K., Surewicz, W. K., Mantsch, H. H. *et al.* (1991). Structural and catalytic properties of a deletion derivative (delta 133 157) of *Escherichia coli* adenylate kinase. *J. Biol. Chem.* **266**, 10781–10786.
22. Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Sci.* **8**, 2251–2257.
23. Searle, M. S., Zerella, R., Williams, D. H. & Packman, L. C. (1996). Native-like beta-hairpin structure in an isolated fragment from ferredoxin: NMR and CD studies of solvent effects on the N-terminal 20 residues. *Protein Eng.* **9**, 559–565.
24. Tasayco, M. L. & Carey, J. (1992). Ordered self-assembly of polypeptide fragments to form nativelike dimeric trp repressor. *Science*, **255**, 594–597.
25. Santos, J., Gebhard, L. G., Risso, V. A., Ferreyra, R. G., Rossi, J. P. & Ermácora, M. R. (2004). Folding of an abridged beta-lactamase. *Biochemistry*, **43**, 1715–1723.

26. Clerico, E. M., Peisajovich, S. G., Ceolin, M., Ghiringhelli, P. D. & Ermácora, M. R. (2000). Engineering a compact non-native state of intestinal fatty acid-binding protein. *Biochim. Biophys. Acta*, **1476**, 203–218.
27. Moews, P. C., Knox, J. R., Dideberg, O., Charlier, P. & Frère, J. M. (1990). Beta-lactamase of *Bacillus licheniformis* 749/C at 2 Å resolution. *Proteins: Struct. Funct. Genet.* **7**, 156–171.
28. Frate, M. C., Lietz, E. J., Santos, J., Rossi, J. P., Fink, A. L. & Ermácora, M. R. (2000). Export and folding of signal-sequenceless Bacillus licheniformis beta-lactamase in *Escherichia coli*. *Eur. J. Biochem.* **267**, 3836–3847.
29. Fersht, A. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Freeman, New York.
30. Tsou, C. L. (1995). Inactivation precedes overall molecular conformation changes during enzyme denaturation. *Biochim. Biophys. Acta*, **1253**, 151–162.
31. Fonze, E., Vanhove, M., Dive, G., Sauvage, E., Frere, J. M. & Charlier, P. (2002). Crystal structures of the *Bacillus licheniformis* BS3 class A beta-lactamase and of the acyl-enzyme adduct formed with cefoxitin. *Biochemistry*, **41**, 1877–1885.
32. Lietz, E. J., Truher, H., Kahn, D., Hokenson, M. J. & Fink, A. L. (2000). Lysine-73 is involved in the acylation and deacylation of beta-lactamase. *Biochemistry*, **39**, 4971–4981.
33. Radzicka, A. & Wolfenden, R. (1995). A proficient enzyme. *Science*, **267**, 90–93.
34. Llinás, A., Vilanova, B., Frau, J., Muñoz, F., Donoso, J. & Page, M. I. (1998). Chemical reactivity of penicillins and cephalosporins. Intramolecular involvement of the acyl-amido side chain. *J. Org. Chem.* **63**, 9052–9060.
35. Peracchi, A. (2001). Enzyme catalysis: removing chemically 'essential' residues by site-directed mutagenesis. *Trends Biochem. Sci.* **26**, 497–503.
36. Carter, P. & Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature*, **332**, 564–568.
37. Hokenson, M. J., Cope, G. A., Lewis, E. R., Oberg, K. A. & Fink, A. L. (2000). Enzyme-induced strain/distortion in the ground-state ES complex in beta-lactamase catalysis revealed by FTIR. *Biochemistry*, **39**, 6538–6545.
38. Jacob, F., Joris, B. & Frere, J. M. (1991). Active-site serine mutants of the *Streptomyces albus* G beta-lactamase. *Biochem. J.* **277**, 647–652.
39. Heger, A. & Holm, L. (2000). Rapid automatic detection and alignment of repeats in protein sequences. *Proteins: Struct. Funct. Genet.* **41**, 224–237.
40. Fitzkee, N. C., Fleming, P. J., Gong, H., Panasik, N., Jr, Street, T. O. & Rose, G. D. (2005). Are proteins made from a limited parts list? *Trends Biochem. Sci.* **30**, 73–80.
41. Daggett, V. & Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18–25.
42. Iwakura, M., Nakamura, T., Yamane, C. & Maki, K. (2000). Systematic circular permutation of an entire protein reveals essential folding elements. *Nature Struct. Biol.* **7**, 580–585.
43. Lattman, E. E. & Rose, G. D. (1993). Protein folding—what's the question? *Proc. Natl Acad. Sci. USA*, **90**, 439–441.
44. Alexandrescu, A. T., Gittis, A. G., Abeygunawardana, C. & Shortle, D. (1995). NMR structure of a stable "OB fold" sub domain isolated from staphylococcal nuclease. *J. Mol. Biol.* **250**, 134–143.
45. Maity, H., Maity, M., Krishna, M. M., Mayne, L. & Englander, S. W. (2005). Protein folding: the stepwise assembly of foldon units. *Proc. Natl Acad. Sci. USA*, **102**, 4741–4746.
46. Bhutani, N. & Udgaonkar, J. B. (2003). Folding subdomains of thioredoxin characterized by native-state hydrogen exchange. *Protein Sci.* **12**, 1719–1731.
47. Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. (1998). The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl. Acad. Sci. USA*, **95**, 9903–9908.
48. Wetlaufer, D. B. (1981). Folding of protein fragments. *Advan. Prot. Chem.* **34**, 61–92.
49. Lesk, A. M. & Rose, G. D. (1981). Folding units in globular proteins. *Proc. Natl Acad. Sci. USA*, **78**, 4304–4308.
50. Panchenko, A. R., Luthey_Schulten, Z., Cole, R. & Wolynes, P. G. (1997). The foldon universe: a survey of structural similarity and self-recognition of independently folding units. *J. Mol. Biol.* **272**, 95–105.
51. Tsai, C. J. & Nussinov, R. (1997). Hydrophobic folding units derived from dissimilar monomer structures and their interactions. *Protein Sci.* **6**, 24–42.
52. Fischer, K. F. & Marqusee, S. (2000). A rapid test for identification of autonomous folding units in proteins. *J. Mol. Biol.* **302**, 701–712.
53. Peng, Z. Y. & Wu, L. C. (2000). Autonomous protein folding units. *Advan. Protein Chem.* **53**, 1–47.
54. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93.
55. Aitken, A. & Learmonth, M. (1996). Performic acid oxidation. In *The Protein Protocols Handbook* (Walker, J. M., ed.), pp. 341–342, Humana Press Inc., Totowa, NJ.
56. Ambler, R. P., Coulson, F. W., Frere, J. M., Ghuysen, J. M., Joris, B., Forsman, M. *et al*. (1991). A standard numbering scheme for the class A β-lactamases. *Biochem. J.* **276**, 269–270.

Edited by C. R. Matthews

(Received 12 December 2005; received in revised form 26 January 2006; accepted 27 January 2006)
Available online 9 February 2006