doi:10.1016/j.jmb.2011.05.027 J. Mol. Biol. (2011) 411, 143–157

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Exploring the Structure–Function Loop Adaptability of a $(\beta/\alpha)_8$-Barrel Enzyme through Loop Swapping and Hinge Variability

Adrián Ochoa-Leyva[1], Francisco Barona-Gómez[2], Gloria Saab-Rincón[1], Karina Verdel-Aranda[2], Filiberto Sánchez[1] and Xavier Soberón[1,3]*

[1]*Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Cuernavaca, C.P. 62210, México*
[2]*Evolution of Metabolic Diversity Laboratory, Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), CINVESTAV-IPN, Km 9.6 Libramiento Norte, Carretera Irapuato-León, Irapuato, C.P. 36822, México*
[3]*Instituto Nacional de Medicina Genómica (INMEGEN), Secretaría de Salud, México City, Distrito Federal, C.P. 01900, Mexico*

Received 13 December 2010; received in revised form 31 March 2011; accepted 18 May 2011
Available online 26 May 2011

Edited by C. R. Mathews

Keywords:
protein evolution;
protein engineering;
structure–function relationship;
loop swapping;
enzyme design

Evolution of proteins involves sequence changes that are frequently localized at loop regions, revealing their important role in natural evolution. However, the development of strategies to understand and imitate such events constitutes a challenge to design novel enzymes in the laboratory. In this study, we show how to adapt loop swapping as semiautonomous units of functional groups in an enzyme with the $(\beta/\alpha)_8$-barrel and how this functional adaptation can be measured *in vivo*. To mimic the natural mechanism providing loop variability in antibodies, we developed an overlap PCR strategy. This includes introduction of sequence diversity at two hinge residues, which connect the new loops with the rest of the protein scaffold, and we demonstrate that this is necessary for a successful exploration of functional sequence space. This design allowed us to explore the sequence requirements to functional adaptation of each loop replacement that may not be sampled otherwise. Libraries generated following this strategy were evaluated in terms of their folding competence and their functional proficiency, an observation that was formalized as a Structure–Function Loop Adaptability value. Molecular details about the function and structure of some variants were obtained by enzyme kinetics and circular dichroism. This strategy yields functional variants that retain the original activity at higher frequencies, suggesting a new strategy for protein engineering that incorporates a more divergent sequence exploration beyond that limited to point mutations. We discuss how this approach may provide insights into the mechanism of enzyme evolution and function.

Introduction

Molecular evolution of proteins involves mutation and recombination events through sequence space to create new functional properties.[1] During the study of these evolutionary processes in the laboratory, point mutations have played an important role, but these are only able to search local regions of sequence space. For

*Corresponding author. E-mail address: soberon@ibt.unam.mx.

Abbreviations used: SFLA, Structure–Function Loop Adaptability; PriA, phosphoribosyl isomerase A; CFU, colony-forming units; PTE, phosphotriesterase; PRA, *N*-(5′-phosphoribosyl)anthranilate.

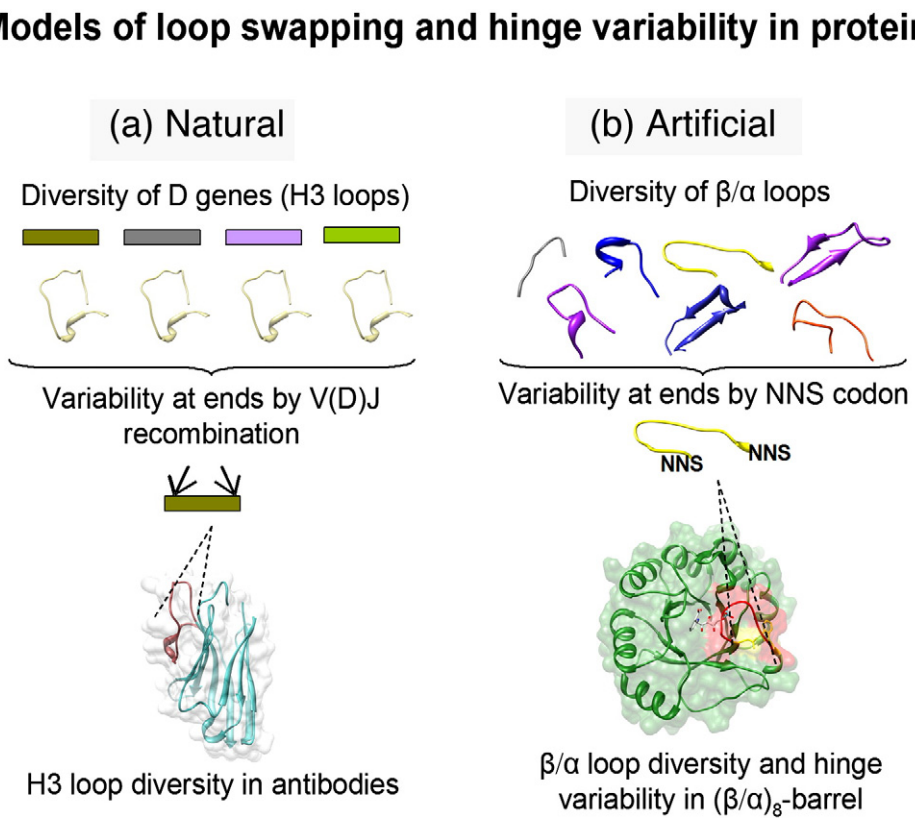

Fig. 1. Models of loop diversity and adaptability in proteins. (a) Natural loop diversity and adaptability in antibodies. The most important loop in antibodies (H3 loop) is encoded by different D genes. Additional variability is introduced by the V(D)J recombination event, generating sequence diversity at both ends of the selected D gene.[5] The diversity of the H3 loops and the introduction of variability in both junctions result in a large number of potential binding sites subsequently selected against the antigens. The three-dimensional (3D) structure of the single domain antibody cAb-Lys2 [Protein Data Bank (PDB) ID: 1rjc][5] is illustrated at the bottom. The H3 loop is shown in brown. (b) Artificial loop diversity and adaptability in $(\beta/\alpha)_8$-barrels proposed in this article. Different β/α loops have embedded functional information representing a diversity source that may represent functional protein modules that can be adapted to a different scaffold through the introduction of variability at both ends. The 3D structure of ecTrpF (PDB ID: 1pii) forming a complex with the product analogue reduced 1′-(2′-carboxyphenylamino)-1′-deoxiribulose 5′-phosphate (rCdRP) is illustrated at the bottom. β/α loop 6 is shown in red. Variability was introduced using the degenerate codon NNS at both ends of the loops (hinges). The 3D structures were rendered using the Chimera package.[8]

exploration of more divergent variability, sequence changes such as insertions, deletions or module exchanges should be taken into account.[2] In natural sequences, these changes are frequently localized at nonstructured specific regions, i.e., loops, revealing their role in protein evolution and in the diversification of numerous enzyme families and superfamilies.[3] Usually, the sequence identity between loops from unrelated enzymes is low or does not exist. Otherwise, there is a high sequence conservation among loops from homologous enzymes, revealing their structure–function relationships.[4] Incorporating such types of sequence changes would be a very valuable asset to the engineering of proteins, but methodologies to successfully mimic such events in the laboratory remain to be developed.

Antibodies are a versatile example of natural loop engineering to design different binding specificities. Recently, it was pointed out that the most important loop (H3 loop) within the antigen binding site is mainly encoded by a particular D gene, inserted between the V and J genes.[5] Picking this gene out of a small pool provides most of the diversity. Additional variability is introduced by the imprecision in V(D)J recombination event, generating sequence diversity at both ends of the selected D gene and, therefore, on the structure of the H3 loop.[5–7] This introduction of variability in both junctions of the H3 loop, in addition with the combination of different D genes, results in a large number of potential binding sites subsequently selected against the antigens. This example provides information about the natural functional adaptation of this loop in antibodies (Fig. 1).

Establishing the structure–function relationship of loops, which in some cases lack electron density in

crystallographic structural models, is a fundamental question in protein science. Loops are known to play critical roles of enzyme function, including catalysis,[9] substrate specificity[10] and protein–protein interactions.[11] Thus, the development of novel methods for engineering of loops would promote the construction of enzymes with novel substrate specificities or functions.[12] In order to face this challenge, our group has recently developed a novel systematic methodology, which was employed to investigate the structural evolvability of a modified $(\beta/\alpha)_8$-barrel enzyme by different loop exchanges.[13] We have originally focused on this protein fold since the $(\beta/\alpha)_8$-barrel is one of the most common folds among protein catalysts and has great functional and structural versatility.[14] Conveniently, the residues of their active sites are located at the C-terminal ends of β-strands as well as in loops that link β-strands with α-helices (i.e., β/α loops), which are considered not to contribute to protein stability.[13,14]

The present work focuses on the effectiveness that loop swapping and hinge variability may have in exploring the functional sequence space of a $(\beta/\alpha)_8$-barrel enzyme, allowing a more divergent sequence exploration,[2] which thus far has largely been limited to point mutations. Additionally, to mimic the natural system of antibodies that produces diversity on junctions of their H3 loop[5] and on the basis of the assumption that junctions might require different conformations to accommodate each loop replacement, we randomized the residues that connect the new loops with the rest of the protein scaffold (Fig. 1). To this end, the β/α loop 6 of monofunctional *N*-(5′-phosphoribosyl)anthranilate (PRA) isomerase from *Escherichia coli* (ecTrpF) was selected.[15] This enzyme catalyzes the Amadori rearrangement of PRA to 1′-(2′-carboxyphenylamino)-1′-deoxiribulose 5′-phosphate (CdRP),[16] which is the third step in the synthesis of tryptophan from chorismic acid. The loop 6 is located just above the catalytic residue Asp126, and it forms a long and flexible lid over the active-site pocket, which is critical for substrate binding and PRA isomerase activity.[17]

Loop replacements were performed with an overlap PCR strategy (Fig. 3a) that includes the introduction of sequence variability at the amino and the carboxy ends, i.e., the hinge residues of each loop replacement, mimicking the natural system of antibodies to generate diversity at both ends of the H3 loop.[5] The final libraries were evaluated in terms of their folding competence and their functional proficiency, an observation that was formalized as a Structure–Function Loop Adaptability (SFLA) value. Details about the function and stability of some variants were obtained by enzyme kinetic studies of PRA isomerase activity, circular dichroism (CD) and size-exclusion chromatography. The observation that varying the hinge residues of loops allows preservation of function on a large number of variants suggests a whole new strategy for protein engineering, incorporating a more divergent sequence exploration, beyond that limited to point mutations.

Results

Loop replacement and hinge variability design

It has been suggested that β/α loop 6, which connects the β-strand 6 to helix 6, fulfills an equivalent functional role, as a lid that closes down upon substrate binding in three different $(\beta/\alpha)_8$-barrel enzymes: TrpF (PRA isomerase, EC 5.3.1.24), TrpC (indole-3-glycerol-phosphate synthase, EC 4.1.1.48) and TrpA (α-subunit of tryptophan synthase, EC 4.2.1.20).[18] These enzymes catalyze consecutive reactions in tryptophan biosynthesis, which share conserved elements of substrate specificity, including the phosphate binding site, supporting the idea that they share a common evolutionary origin.[18] In histidine biosynthesis, the enzyme HisA (ProFAR isomerase, EC 5.3.1.16) uses a reaction mechanism similar to that of TrpF to convert an identical aminoaldose moiety into the corresponding aminoketose.[19] Additionally, PRA isomerase activity can be established in HisA,[20] and after several rounds of mutagenesis, it was observed that conformational changes in its β/α loop 6 are important to increase this activity.[21] Moreover, an enzyme homologue of HisA, dubbed phosphoribosyl isomerase A (PriA), which takes part in both histidine and tryptophan biosynthesis, therefore harboring both ProFAR and PRA isomerase activities, was discovered.[22] PriA is structurally similar to HisA, and conformational changes in β/α loops 5 and 6 have been suggested as important for its PRA isomerase activity.[23,24]

The loops to be used as replacements were selected with the aim of covering a range of functional, structural and evolutionary relationships with loop 6 of ecTrpF (Loop 6 Wt) (Table 1 and Fig. 2). Loop 6 of tryptophan synthase α chain from *E. coli* (Loop 6 ecTrpA), loop 6 of PriA from *Streptomyces coelicolor* (Loop 6 scPriA) and loop 6 of ProFAR isomerase from *Thermotoga maritima* (Loop 6 tmHisA) have all three types of relationships, as was described above. Loop 6 of methyltetrahydrofolate–corrinoid-dependent methyltransferase from *Moorella thermoacetica* (Loop 6 mtMetR) was chosen only for its structural equivalent position. Additionally, loops 1 of scPriA (Loop 1 scPriA) and tmHisA (Loop 1 tmHisA), which do not have functional, structural or evolutionary relationships to Loop 6 Wt, were selected. Moreover, to explore the intrinsic functional role of the original loop 6, we designed a new loop, predicted to be compatible with the same backbone structure but with a different sequence, using the

Table 1. Sequence and structural characteristics of loop replacements

Loop replacement	Sequence[a]	Length	Structural similarity (RMSD)[b]	Sequence identity[c] (%)	Type of relationships with Loop 6 Wt
Loop 6 Wt	**N**GQGGSGQRFD**W**	12	0 Å on 12 residues	100	—
Loop 6 scPriA	**D**IAKDGTLQGP**N**	12	2.7 Å on 8 residues	17	Functional, structural and evolutionary
Loop 6 tmHisA	**E**KDGTLQEHD**F**	11	2.1 Å on 9 residues	27	Functional, structural and evolutionary
Loop 6 ecTrpA	**S**RAGVTGAENRAALP**L**	16	2.3 Å on 12 residues	25	Functional, structural and evolutionary
Loop 6 mtMetR	**P**LILPANVA**Q**	10	2.8 Å on 7 residues	0	Structural equivalent position
Loop 6 Rosetta	**N**GNEGDGVEHD**W**	12	0.2 Å on 12 residues	40	Theoretically calculated with the same backbone structure
Loop 1 scPriA	**D**GQAVRLVHGESGTETSYG**S**	20	—[d]	—[d]	None
Loop 1 tmHisA	**R**GKVARMIKGRKENTIFYEK**D**	21	—[d]	—[d]	None

[a] Sequence of each loop used for the replacement. The hinge residues mutated to saturation are shown in bold and underlined.
[b] Minimal RMSD calculated by a structural superimposition of the α-carbons of loop 6 of ecTrpF with each loop replacement.
[c] Calculated on the basis of the structural superimposition described in footnote b.
[d] Fewer than three residues aligned; loop 6 of ecTrpF does not match with loop replacement.

Rosetta Design web server.[25] The resulting lowest free-energy sequence (Loop 6 Rosetta) differs by 60% from the original loop 6.

The analysis of protein crystal structures has shown that loop movements can take two main forms: (i) hinge motions that are not constrained by tertiary packing interactions, which are characterized by the localization of motion to a few main-chain torsion angle changes, and (ii) shear motions that are manifested by small side-chain torsion angle changes along an entire interface and closely packed segments of polypeptides.[26] Asn127 and Trp138 are the end residues of the β/α loop 6 of ecTrpF (Fig. 2), which, in comparison with the end residues identified after a superimposition of liganded and unliganded TrpF structures from *T. maritima*, comprise two well-defined protein hinges (data not shown). In addition, on the basis of previous studies on loop 6 of triosephosphate isomerase,[27] an experimental model for the $(\beta/\alpha)_8$-barrels that has

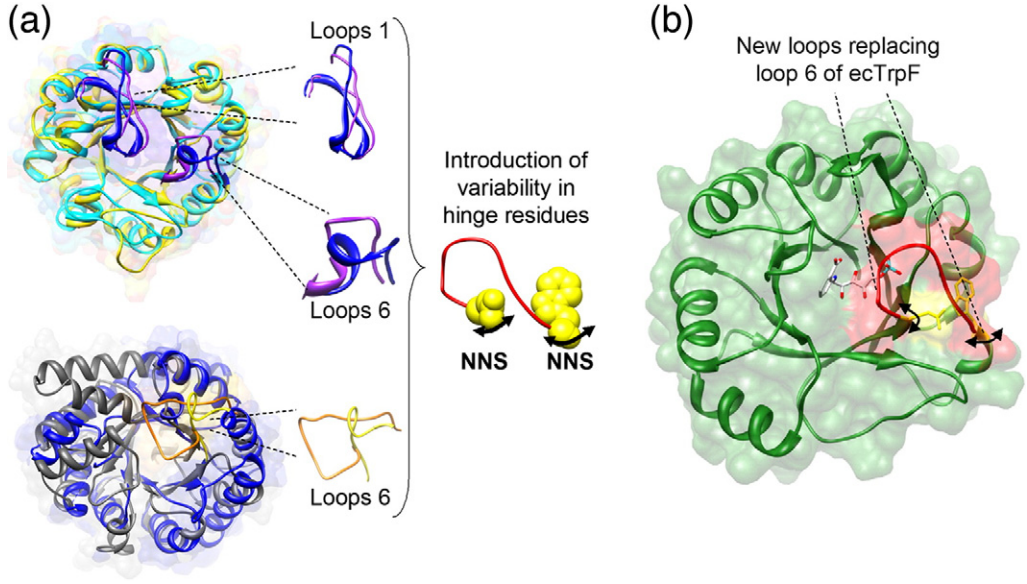

Fig. 2. Loop replacement design. (a) β/α loops from positions 1 and 6 of the scPriA (yellow), tmHisA (cyan), ecTrpA (gray) and mtMetR (blue) enzymes are illustrated. Zoom representation of selected loops is shown; the colors are a follows: blue (tmHisA), purple (scPriA), orange (ecTrpA) and yellow (mtMetR). Sequence variability was introduced using the degenerate codon NNS at both end hinge residues of each loop replacement. (b) The selected loops were used to replace the β/α loop 6 (shown in red) of ecTrpF, which is anchored by hinge residues Asn127 and Trp138. 3D structure of ecTrpF (PDB ID: 1pii) forming a complex with the product analogue rCdRP is illustrated. Docking of rCdRP was inferred from a TrpF structure in complex with rCdRP (PDB ID: 1lbm) and superimposed with the ecTrpF structure. The 3D structures were rendered using the Chimera package from UCSF.[8]

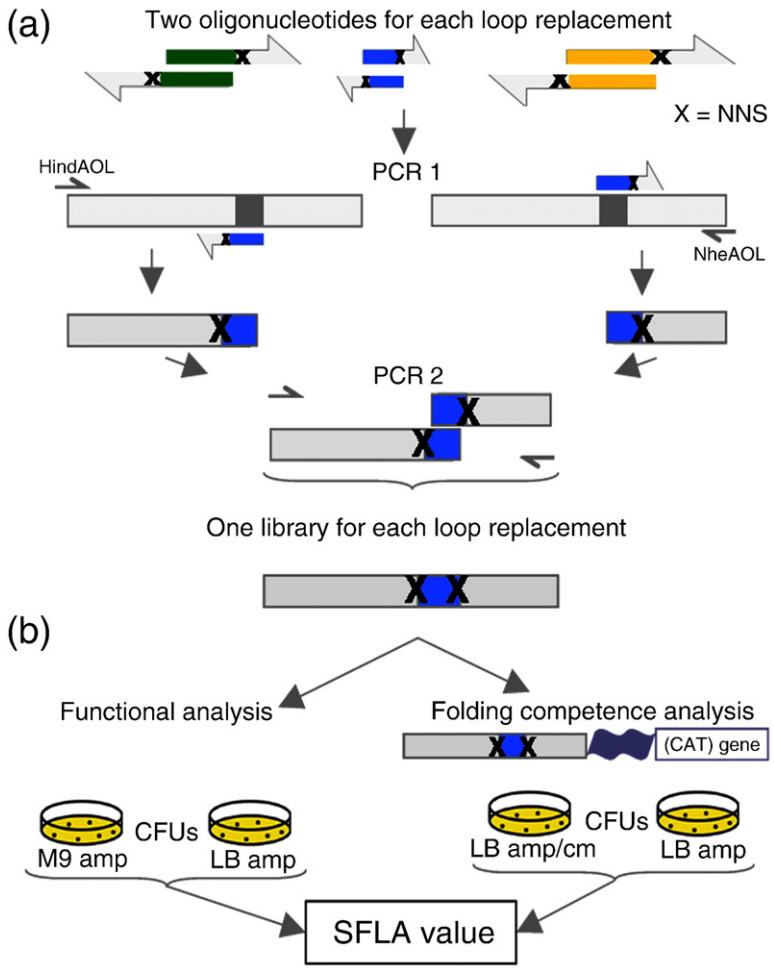

Fig. 3. Library construction and assessment of the SFLA value. (a) One pair of oligonucleotides of different sizes (from 36 bp to 69 bp), which are partially complementary (12 bp), was designed for each loop replacement. The first halves of the *trpF* gene were constructed using the oligonucleotide HindAOL as 5′-primer and the corresponding noncoding loop oligonucleotide as 3′-primer (left, PCR 1). The carboxy halves were constructed using the corresponding coding loop oligonucleotide as 5′-primer and the oligonucleotide NheAOL as 3′-primer (right, PCR 1). Libraries were amplified by overlapping-extension PCR (PCR 2) using the two PCR products from the previous reactions as templates and the oligonucleotides HindAOL and NheAOL. (b) The fraction of functional and folded variants was calculated as described in Materials and Methods. The results of the functional proficiency analysis were the first component of SFLA value, and the results of the folding competence analysis were the second component.

evolutionary relationships with TrpF,[18] these residues would be expected to participate in the movement of the original loop 6. Therefore, each loop replacement could require different hinge residues, such that conformational movements in the new protein scaffold could be coordinated to maintain catalysis. To this end, loop replacements were performed with an overlap PCR strategy that introduces sequence variability at both end residues, allowing the exploration of all 20 amino acids (Fig. 3a).

From the sequence analyses of all libraries, it was concluded that 97% of the generated variants were correctly constructed. The sequence of the hinge residues reveals that there is a true biased representation of some amino acids, which is in agreement with the theoretical distribution expected by an NNS codon (Supplementary Fig. S1).

Scoring the folding competence and functional proficiency

To estimate the folding competence, we fused the libraries to a chloramphenicol acetyl transferase gene using the system that we have previously described as a reliable folding competence reporter.[13] Briefly, if the fusion is stably expressed and soluble in bacteria, they confer resistance to chloramphenicol, thus providing a simple test for folding competence. This folding competence

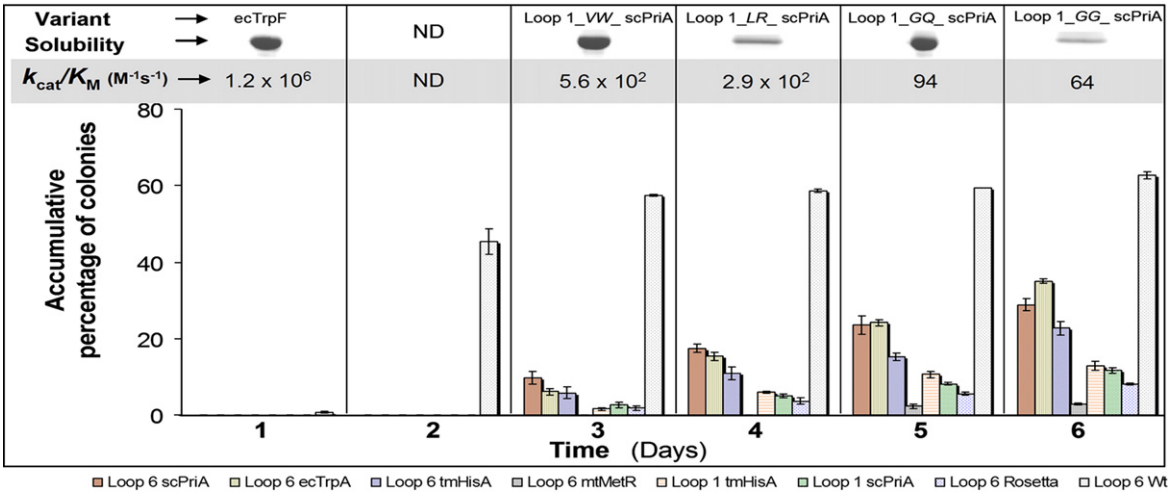

Fig. 4. Kinetics and sensitivity of the *in vivo* functional assay. The percentage values and error bars represent the accumulative average and standard deviation, respectively, of the number of colonies that appeared throughout the functional assay until 6 days. (Top) the catalytic efficiency values (k_{cat}/K_m) of selected variants from the library Loop 1 scPriA, which were isolated at different time points, are shown. The soluble protein level after overexpression and nickel-affinity purification is also shown. ND, values not determined since only library Loop 6 Wt rendered functional variants at this point in time.

selection is independent of the amount of soluble protein.[13] The fraction of folded variants of each library was estimated by comparing the number of colony-forming units (CFU) growing on plates with and without selection pressure for folding competence (Fig. 3b).

The functional proficiency of each library was scored by the ability of their variants to enable growth of the *E. coli* tryptophan auxotroph FBG-Wf (JM101 $trpF^{\Delta}$)[23] on minimal media lacking tryptophan. Hence, the fraction of functional variants was estimated by comparing the number of CFU growing on plates of minimal and rich media (Fig. 3b). The number of CFU that appeared on the selection plates was counted every 24 h for 6 days (Fig. 4). The ecTrpF enzyme complements the *trpF* deficiency overnight. Extraction and sequencing of plasmids that complemented at different time points throughout this period of evaluation, just before false positives started to appear, were used to define the functional range of variants in the library. The function conferred by these constricts was confirmed after transformation *de novo* of fresh *E. coli* FBG-Wf cells and growth on minimal media (data not shown). Interestingly, the number of CFU that appeared after every 24 h varies among the different libraries studied (Fig. 4). The lack of stochastic variation in cell growth suggests that gene amplification and point mutation of chromosomal genes[28,29] were not contributing to the observed phenotypes.

Enzyme kinetic analysis qualitatively confirmed the *in vivo* complementing experiments. Variants selected after 3–4 days of incubation in minimal media have the best k_{cat}/K_m values, while the variants selected after 5–6 days have the poorest k_{cat}/K_m values of all variants analyzed (Fig. 4). Furthermore, for at least one library, a correlation between the observed phenotypes and the decreasing levels of catalytic efficiency (k_{cat}/K_m) was observed (Fig. 4 and Table 2). Long complementing times show that our threshold for selection of true functional variants with PRA isomerase activity is at least higher than $k_{cat}/K_m = 64\ M^{-1}\ s^{-1}$. Furthermore, although the expression or solubility levels may be influencing the functional proficiency reported by our *in vivo* assay, the overexpression analysis of selected variants does not show a correlation between the solubility level and the complementing time (Fig. 4 and Supplementary Fig. 2a). Indeed, the expression of several randomly selected variants scored as not active by our functional proficiency assay but soluble by our folding competence assay suggests that differential solubility levels are not related with the functional proficiency of the variants (Supplementary Fig. 2b). We previously observed in the *E. coli* FBG-Wf strain that loss of PRA isomerase activity of different PriA enzyme variants is unrelated to insolubility problems.[23]

Significance of the SFLA value

On the basis of theoretical and statistical analyses supporting that structural changes in loops are strongly coupled to the evolutionary distance of their sequences and with their functional dependence,[4] we conducted experiments that provide data to estimate an SFLA value (Fig. 3b). This value was scored as the ratio between the fractions of functional variants to the total fractions of folded variants and was used as a metric reporting on the

Table 2. *In vitro* and *in vivo* functional analysis of selected variants

Variant[a]	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	Complementing time (days)[b]
ecTrpF	2028	28.9	1.2×10^6	0.5
Loop 6_*QY*_ecTrpA	0.96	32.4	4.9×10^2	3
Loop 6_*GW*_scPriA	0.78	60.5	2.1×10^2	3
Loop 6_*RL*_tmHisA	1.14	82.4	2.3×10^2	3
Loop 1_*VW*_scPriA	0.84	24.8	5.6×10^2	3
Loop 1_*LR*_scPriA	1.10	63.8	2.9×10^2	4
Loop 1_*GQ*_scPriA	0.96	170.1	94	5
Loop 1_*GG*_scPriA	0.96	250.9	64	6
Loop 6_*EN*_scPriA	ND	ND	ND	~

The overall standard errors of enzyme kinetic parameters are less than 20%.

Plasmidic DNA was extracted from selected variants and used to retransform the *E. coli* strain MC1061*thiE*[4] and plated on LB supplemented with ampicillin and chloramphenicol. Resistance to this latter antibiotic (Cm^R) is shown in the last column.

~, after 1 week of incubation, this variant does not complement.

[a] The N- and C-terminal hinge residues are shown in italic type.

[b] Time elapsed before the appearance of visible colonies (1 mm) on minimal medium plates. Growth was recorded every 24 h.

number of functional enzyme variants of each library. Hence, the final value determines the fraction of random mutations at the hinge residues for each loop replacement that are capable of folding and retaining sufficient PRA isomerase activity to support tryptophan synthesis and, therefore, growth on minimal media. As a reference, we probed the effect of only introducing variability at the hinge residues Asn127 and Trp138 while maintaining the wild-type loop 6 (Loop 6 Wt).

We found that the fraction of folded and functional variants depends on the loop used as replacement. The values obtained from different libraries vary between 49% and 75% for folding competence and from as little as 3% to as much as 60% for functional proficiency (Fig. 5a). The differences between the higher numbers of folded proteins compared to functional proteins imply the non-retention of PRA isomerase activity in several folded variants. The lack of activity was not a consequence of a decrease in expression levels (Supplementary Fig. 2b) and the subsequent enzyme kinetic analysis, for at least one variant showed that it is catalytically dead (Table 2). This suggests that the SFLA value reflects on the ability of loop replacement with hinge variability to retain the PRA isomerase activity on the ecTrpF scaffold.

The values obtained for Loop 6 Wt show that most of the folded variants retain PRA isomerase activity (70% folded *versus* 60% functional), resulting in an SFLA value of 0.84 (Fig. 5b). The library with the second highest value of functional variants and an SFLA value of 0.70 was Loop 6 ecTrpA, which came as a surprise since this enzyme lacks PRA isomerase activity (Fig. 5a and b). As expected, despite the high structural and sequence similarity between Loop 6 scPriA and Loop 6 tmHisA, the former, which has been related to its PRA isomerase activity,[23,24] showed a larger fraction of functional variants on the ecTrpF scaffold (Fig. 5a). Indeed, the SFLA values (0.39 and 0.28, respectively) of their cognate libraries reflected higher structure–function loop adaptability for Loop 6 scPriA than for Loop 6 tmHisA (Fig. 5b).

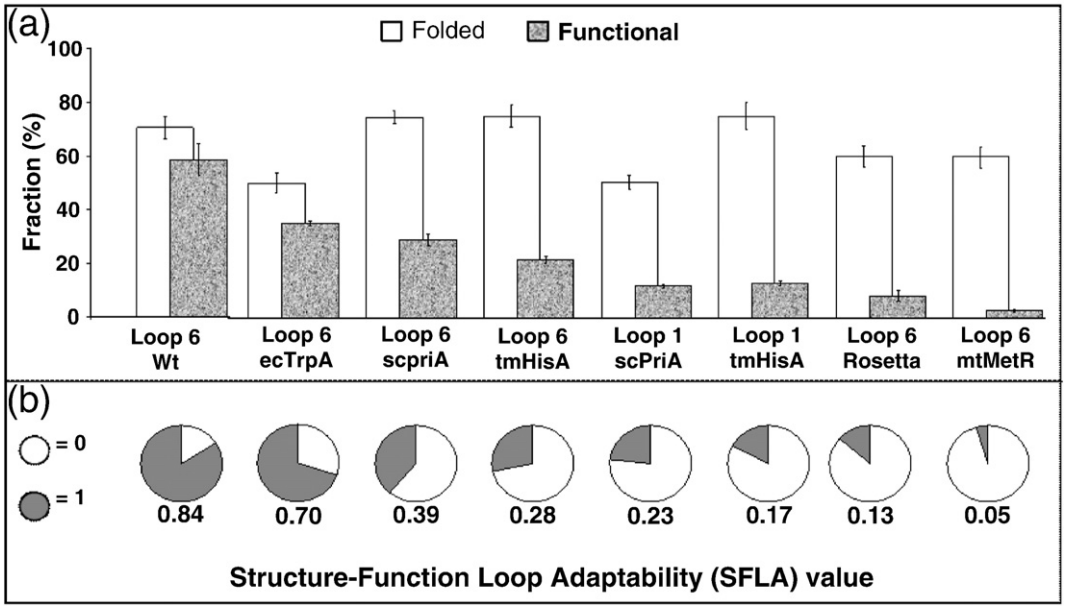

Fig. 5. Analysis of functional proficiency and folding competence and SFLA values of the libraries. (a) Fraction of folded and functional variants, which are the two components of SFLA value. The fraction values and error bars represent the average and standard deviation, respectively. (b) SFLA value (see Materials and Methods).

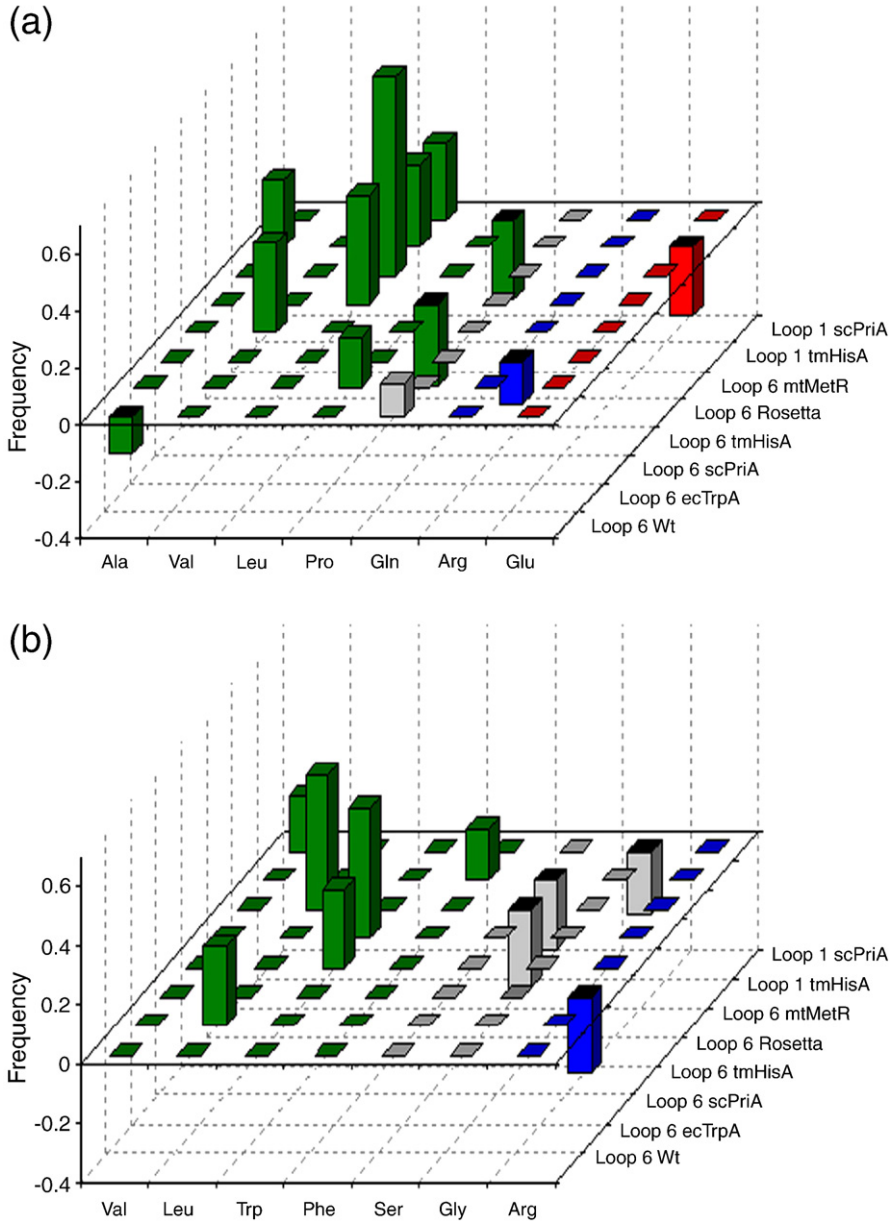

Fig. 6. Sequence analysis of the selected residues in the two hinges.. Sequence analysis of the selected residues in the two hinges. The positive or negative overrepresentation of the normalized frequency for amino acids found on the selected *versus* the nonselected conditions is shown in the amino (a) and carboxy (b) hinges. Amino acids are colored according to their physicochemical properties as follows: green, nonpolar and hydrophobic; gray, polar and uncharged; blue, polar basic; and red, polar acidic. The figure shows only the amino acids with a frequency either higher or lower (2 SD) than expected by the NNS design.

The functional relevance of the results described above is emphasized when it is noted that, despite having a high fraction of folded variants, libraries Loop 6 mtMetR, Loop 6 Rosetta, Loop 1 scPriA and Loop 1 tmHisA have a lower fraction of functional variants, with consequent lower SFLA values of 0.05, 0.13, 0.23 and 0.17, respectively (Fig. 5b). Loop 6 Rosetta, which shares 40% sequence identity with Loop 6 Wt, was compatible with the target protein backbone in folding competence terms, as shown by 60% folded variants; however, only 8% of the variants were functional, yielding a poor SFLA value of 0.13 (Fig. 5a and b).

Hinge variability for each loop replacement

The different SFLA values imply that certain hinge residues that maintain PRA isomerase activity may be

selected. To investigate if there is a pattern of functionally compatible residues in our assay, we performed a sequence analysis for each library complementing the *trpF* deficiency in minimal media. There is a true biased representation of some amino acids without selective pressure as a result of the variability introduced by the NNS codon (Supplementary Fig. S1). Therefore, the frequencies observed under selective pressure *versus* nonselective pressure were normalized per amino acid and for each library. Despite the small number of the samples sequenced (about 20 per library and for each condition), we found a statistically significant overrepresentation, indicating that some hinge residues are more compatible than others with the new loop and/or vice versa (Fig. 6). In the ecTrpF structure, the original hinge residues Asn127 and Trp138 anchor β/α loop 6, and the observed recurrence of residue selection at these positions suggests that the reorganization of each loop replacement with hinge residues is necessary. This analysis also showed a general preference for hydrophobic residues at both hinges, independently of the loop replacement, suggesting the importance of maintaining hydrophobic packing in these regions. Additionally, as suggested by the number of residues that was positively or negatively selected under functional constraints, we found that there is a more restricted acceptance of variability at the amino hinge with respect to the carboxy hinge (Fig. 6).

Enzyme kinetic and structural analyses of a set of variants

To obtain a broader understanding of the functional and structural effects on the ecTrpF scaffold caused by (i) different loop replacements and (ii) different residues at hinge positions maintaining the same loop replacement, we performed steady-state enzyme kinetics of PRA activity (Table 2) and structural studies by CD in a set of variants (Fig. 7). One of the fastest-growing variants complementing the tryptophan auxotrophy, which formed colonies after 3 days on minimal media, was selected for different libraries (Table 2). Also, four variants from library Loop 1 scPriA, which covered the window of our functional assay (3–6 days, Fig. 4), were selected. This selection was complemented by a variant from Loop 6 scPriA (Table 2) that was not able to complement the tryptophan auxotrophy but was shown to be soluble (Supplementary Fig. 2b).

The enzyme kinetic analysis revealed a minimum decrease of 4 orders of magnitude in k_{cat}/K_m with respect to the wild-type ecTrpF upon loop replacement, largely due to a reduction on the k_{cat} (Table 2). The effect of different hinge residues maintaining loop 1 from scPriA was also explored, and the enzyme kinetic analysis of the variants with glycine at both hinges (loop1_GG_scPriA) and with only one glycine (loop1_GQ_scPriA) suggests that binding to the transition state was mainly affected (Table 2). These results also show that there is a relationship between the N-terminal and C-terminal hinges, illustrating their important role in the loop structure–function adaptability.

Since the catalytic activity of variant Loop 6_*EN*_scPriA could not be detected *in vitro*, which is consistent with the lack of functional activity *in vivo* (Table 2), it is safe to state that our functional proficiency assay is able to select for soluble and nonfunctional variants independently of their

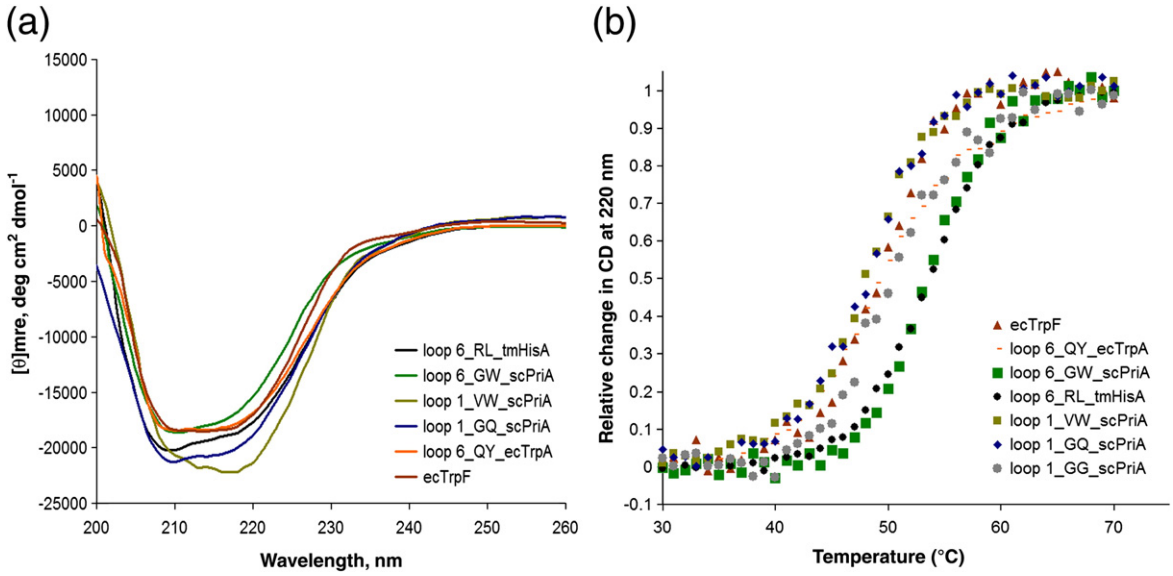

Fig. 7. Structural analysis of selected variants. (a) Far-UV CD spectra of ecTrpF and variants with different loop replacements and with different residues at their hinge positions and (b) thermal unfolding curves.

solubility levels (Supplementary Fig. 2b). These results suggest that variants with a negative functional phenotype could be explained by adoption of incorrect conformations of the catalytic active site. Indeed, K_m values of variants complementing the tryptophan auxotrophy after 5 or 6 days tended to increase, which contrasted with the observation that k_{cat} parameters throughout all the variants analyzed were similar (Table 2).

The far-UV CD spectra analyses suggest only slight structural changes at the secondary structure level as a consequence of loop replacements or modification of hinge residues maintaining the same loop (Fig. 7a and Supplementary Table S1). These slight differences in the CD spectra and in the estimated secondary structure content among the variants analyzed could reflect changes between interactions of secondary structure elements surrounding the original β/α loop 6.[13] The thermal denaturation process (Fig. 7b) showed that all variants form stable structures, with apparent thermal melting temperatures ($T_{m,app}$) that are very similar to those of ecTrpF, around 49 °C, except for the variants with loop 6 from scPriA (Loop 6_GW_scPriA) and loop 6 from tmHisA (Loop 6_RL_tmHisA) that were more thermostable by 4 °C (Supplementary Table S2). This may be related to the fact that both variants exist predominantly as dimers (Supplementary Fig. 3 and Supplementary Table 2), which is in agreement with the previous observations in TrpF from *T. maritima*, where the dimerization state increases the intrinsic thermal stability of the protein.[30]

Discussion

In a manner reminiscent to that of the H3 loop of the antibodies, the β/α loops of $(\beta/\alpha)_8$-barrels vary immensely in length, sequence and conformation (Fig. 1), and it has been suggested that their modularity is a convenient device in the evolution of $(\beta/\alpha)_8$-barrels, as semiautonomous units of critical functional groups.[31,32] Although we are not aware of systematic evidence for a similar mechanism operating in the natural history of $(\beta/\alpha)_8$-barrels, it is tempting to speculate that events resulting in loop swapping have been captured by selective pressure, as previously suggested after phylogenetic analysis.[4] In the case of ecTrpF, it is apparent that β/α loop 6 satisfies the requirements for its functional evolvability, suggesting that, in $(\beta/\alpha)_8$-barrels, the separation of their tightly packed scaffolds (stability face) from their floppy active sites (catalytic face)[33] could be simultaneously promoting structural plasticity[13] as well as functional versatility via active-site loop swapping. Indeed, loops may include structural and dynamic information that is not readily interpreted from their sequences, and these could be translated among different scaffolds. This is implied by the high SFLA values of Loop 6 ecTrpA, which has been postulated to fulfill a function similar to that of Loop 6 ecTrpF,[18] and Loop 6 scPriA, which is an enzyme with PRA isomerase activity.[23] In support of this hypothesis, it was recently observed that different loop sequences may reflect the natural selection of not only chemical properties but also dynamic modes that augment substrate specificity.[34] Our results show that the sequence of any given loop replacement is more important in obtaining functional variants and with less influence in obtaining folded variants (Fig. 5a). Thus, a relationship between loop sequences and function was more prominently found than with proper folding (Fig. 5a).

It is clear that the introduction of variability at hinge residues connecting the loops with the ecTrpF scaffold has a noticeable effect for obtaining variants that retain the PRA isomerase activity. Importantly, an arbitrary pair of connecting residues would have failed to produce a functional protein 40–97% of the time, depending on the particular loop used as replacement. Perhaps not surprisingly, in some cases, for example, the library Loop 6 Wt, the obtained SFLA value is indicative of their high structure–function adaptation (Fig. 5b). Furthermore, the higher number of fastest-growing colonies observed for this library, which only includes variability at hinge residues, supports this notion (Fig. 4). This is in accordance with the fact that loop 6 was highly replaceable by other loops without affecting the folding competence. In contrast, the concomitant loss of functional proficiency of ecTrpF suggests that the high sequence conservation of loop 6 TrpF homologues was evolved mainly under functional constraints. These constraints were probably also related in maintaining an active conformational dynamics of the active site, as was recently observed in another fold.[34]

The SFLA value obtained for Loop 6 Rosetta (0.13), in which the functional constraints were not considered, confirms the lack of functional information embedded on this loop in spite of its 40% sequence identity with Loop 6 Wt (Fig. 5b and Table 1). A comparison of this value with that obtained for Loop 6 Wt (0.84) also highlights the importance of the functional information embedded in the original sequence to retain the PRA isomerase activity of ecTrpF. Low SFLA values were also observed for Loop 6 mtMetR, which only maintains a structurally equivalent position, and Loop 1 scPriA and Loop 1 tmHisA, which lack functional and structural relationships with Loop 6 Wt (Fig. 5b). These results emphasize a recognized challenge in computational enzyme design, recently identified as the main reason for failure to conceive, a priori, an active-site architecture.[35] Within this context, our loop replacement strategy could offer a complementary approach for exploring active-site architectures,

which could then feed back on the enzyme design approaches.

Among heterologous loops, the SFLA value of Loop 6 ecTrpA indicates that it is the most adapted, structurally and functionally, to retain the PRA isomerase activity (Fig. 5b). This result is in agreement with the early suggestion that loop 6 is fulfilling an equivalent function during the catalysis in both TrpF and TrpA, closing down the binding site.[18] Despite the high structural and sequence similarity between Loop 6 scPriA and Loop 6 tmHisA (Fig. 2 and Table 1), we found that Loop 6 scPriA shows a better structure–function adaptability than Loop 6 tmHisA, which could be rationalized assuming that this loop participates in the substrate specificity of the PRA isomerase activity of scPriA,[24] while this activity is absent in tmHisA.[20] Interestingly, in a recent study, wild-type levels of PRA isomerase activity were established in tmHisA after mutations that induce a conformational change in its β/α loop 6.[21]

The selective pressure of our *in vivo* assays (Fig. 3b) reveals a functional and structural association between the hinge residues and each loop replacement (Fig. 6). Consistent with this, the motions of loop 6 in triosephosphate isomerase are controlled by their N- and C-terminal residues, and their mutation increases the appearance of nonfunctional conformations of the original loop, which may not be competent for ligand binding or catalysis[36] or induce nonproductive loop–loop interactions, drastically affecting the kinetic parameters.[37] Therefore, our results highlight the importance of variability at hinge residues to functional adaptation of loops and their probable role on the evolution of the structure–function relationship in $(\beta/\alpha)_8$-barrel enzymes. Interestingly, in another fold, the variability at the end residues of secondary structure elements, including the loops, correlates with their movement across the catalytic cycle and is being subjected to natural selective pressure.[38]

It has previously been shown that only modest levels of PRA isomerase enzymatic activity are sufficient to suppress tryptophan auxotrophy.[20,39] The catalytic efficiency (k_{cat}/K_m) of the wild-type ecTrpF is around 1.2×10^6 M^{-1} s^{-1}, and the catalytic efficiency of our most deficient variant, isolated after 6 days of growth, is in the range of 64 M^{-1} s^{-1}, that is, 5 orders of magnitude less. Assuming that complementing time *in vivo* could be related to the activity *in vitro* (Fig. 4), we decided to establish 6 days as the threshold of a true functional proficiency analysis. This window of observation avoids false positives, and it implies that the selected variants with PRA isomerase activity have a catalytic efficiency better than 64 M^{-1} s^{-1}. In addition, the sensitive assay for PRA isomerase activity is limited to substrate concentrations ≤ 0.8 mM[40] because of the extremely high fluorescence quantum yields of anthranilate and PRA. Recently, it has been estimated that the minimum PRA isomerase catalytic efficiency that can reliably be identified *in vivo* and confirmed *in vitro* is around 0.3 M^{-1} s^{-1}, suggesting that the lower limit for biologically relevant PRA isomerase activity in *E. coli* could be in the range of 25- to 30-fold less.[39] This figure is 7 orders of magnitude below that found for ecTrpF and 2 orders of magnitude below our most deficient variant, demonstrating that the results reported herein are biologically meaningful.

The *in vitro* analysis of a set of variants contributes to a general understanding of the molecular effects of loop swapping and hinge variability in the enzyme used as scaffold. Despite structural and sequence differences between loop replacements and hinge residues, the decrease in the k_{cat} for all analyzed variants shows a similar trend (Table 2), most likely due to disruption of the array of catalytic residues and/or the phosphate binding sites that are located nearby (Fig. 2). Additionally, we found that the range of K_m values among the analyzed variants depends on the restrictions imposed by different loop replacements and also by different hinge residues. Given the structural position and the high importance of the original loop 6 in the substrate binding and catalysis of ecTrpF, loop replacements presented in this work modified drastically the framework of the active site of ecTrpF, which, consequently, affected their kinetic parameters (Table 2). All these new variants could be selected for altered specificities without necessarily abolishing the original activity of the protein, as it was suggested early for insertions in the β/α loops of TrpF.[41]

Recently, two attempts of loop swapping in the $(\beta/\alpha)_8$-barrel have been reported. The first attempt involved the transplantation of the eight β/α loops from a phosphotriesterase (PTE) by the eight corresponding loops of the Dr0390 protein, a member of the amidohydrolase superfamily. However, the chimeric protein was largely insoluble, and the PTE activity to paraoxon could not be detected on the soluble chimeric protein only obtained after refolding.[42] In the second attempt, two β/α loops were grafted from lactonase from *Rhodococcus erythropolis* and lactonase from *Mycobacterium tuberculosis* to lactonase from *Mycobacterium avium subsp. paratuberculosis* K-10 enzyme, which are PTE-like lactonases. However, loop grafting from AhlA to MCP did not result in a change in the substrate preference, and loop grafting from PPH to MCP had no detectable activity toward the substrate of interest.[43] The authors concluded that the effect of loop grafting is still both unpredictable and uncertain and that there are some rather subtle issues that govern the folding of the chimeric protein that are not well understood.[42,43] It could be that hinge variability to adapt the functional information embedded within the β/α loops of these $(\beta/\alpha)_8$-barrels is necessary, as shown by our results.

The observation that varying the hinges of loops allows preservation of function on a large number of variants suggests a whole new strategy for protein engineering, incorporating a more divergent sequence exploration, beyond that limited to point mutations. The SFLA concept can also be used to study other types of structure–function relationships, including the replacement of other secondary structure elements, such as β-sheet or α-helix, where the introduction of sequence variability at their hinge residues may also prove to be important for obtaining folded and functional proteins.

Materials and Methods

Construction of libraries

All oligonucleotides used in this study are described in Supplementary Table S3. Two oligonucleotides, which are partially complementary (12 bp), were designed for each loop replacement: one that corresponds to the noncoding DNA strand and the other, to the coding DNA strand. In both of them, an NNS codon that replaced the hinge residues was introduced. Libraries were independently constructed by PCR using the corresponding pair of oligonucleotides for each loop replacement, as depicted in Fig. 3a. For each library, the first half of the *trpF* gene was constructed using a *trpF*-containing pDAN5 plasmid as template and the oligonucleotide Hind3AOL as 5′-primer with a HindIII restriction site and the noncoding loop oligonucleotide as 3′-primer (Fig. 3a, PCR 1). The second half was constructed using the same plasmid as template and their corresponding coding loop oligonucleotide as 5′-primer and the oligonucleotide Nhe1AOL with an NheI restriction site as 3′-primer (Fig. 3a, PCR 1). The amplification products were purified from a 1% agarose gel. Libraries were finally constructed by overlapping-extension PCR using the corresponding two PCR products from the previous reactions as templates and the oligonucleotides Hind3AOL as 5′-primer and NheAOL as 3′-primer (Fig. 3a, PCR 2). Amplified products were purified from a 1% agarose gel and then digested with HindIII and NheI, and the purified reaction products were ligated into the pDAN5 plasmid. The resulting libraries contained 10^5 to 10^6 different variants. Approximately 20 plasmids, for each library, were sequenced to confirm loop replacement and to analyze the statistical distribution of the variability introduced at both hinges.

Analysis of functional proficiency

The fraction of variants that retain PRA isomerase activity was selected by complementing the tryptophan auxotroph *E. coli* JM101*ΔtrpF*[23] in M9 minimal media. To this end, each library was used to independently transform electrocompetent cells in triplicate. Dilutions of approximately 1000–1500 viable transformed cells were spread on two different media with ampicillin (200 μg/ml): LB media agar and M9 minimal media agar. Plates were incubated at 30 °C for 6 days, and the CFU were counted every 24 h. The fraction of functional variants was calculated as the ratio of the CFU grown under selective pressure for PRA isomerase activity (M9 media) to the CFU grown without this selective pressure (LB media), and this value was the first component of the SFLA value (Fig. 3b). For each library, approximately 20 colonies were selected from the M9 plates, and their plasmids were sequenced.

Analysis of folding competence

The fraction of variants with folding competence was estimated by fusing all libraries to the chloramphenicol acetyl transferase gene as an *in vivo* folding reporter, using the method previously described.[13] These libraries, as well as the positive and negative controls for selection of folded proteins, were independently used to transform *E. coli* MC1061*ΔthiE* electrocompetent cells in triplicate. Dilutions of approximately 1000–1500 viable transformed cells were spread on two different conditions with ampicillin (200 μg/ml): LB media agar with and without 20 μg/ml of chloramphenicol. Plates were incubated at 30 °C for 18 h. The fraction of variants with folding competence was calculated as the ratio of the number of CFU grown under selective pressure for folding (LB amp plus chloramphenicol) to the number of CFU grown without this selective pressure (LB amp), and this figure was the second component of the SFLA value (Fig. 3b).

Statistical analysis of sequences

The amino acid frequencies observed at mutagenized hinge positions were scored. The plasmid sequences from the colonies grown without selective pressure for PRA isomerase activity were compared with the sequences found under this pressure for each library. The discrepancy in the abundance of each residue between these two conditions was calculated. The average and the standard deviation of the frequencies were used to determine, with 95% confidence, the negative or positive selection for specific amino acids.

Protein production

Genes encoding the selected variants were subcloned from pDAN5 into pET28b(+) plasmid and sequenced. To this end, genes were amplified by PCR using oligonucleotides 5′-GCCATACCATGGGGGAGAA-TAAGGTATGTGGC-3′ with an NcoI site (underlined) as 5′-primer and 5′-GTCCGAAAGCTTTCATTA-GTGGTGGTGGTGGTGGTGGGATCCATATG-CGCGCA-3′ with a HindIII site (underlined) as 3′-primer. The amplified product was ligated into pET28b(+). All constructs were sequenced entirely to exclude inadvertent mutations. Overexpression was performed in *E. coli* Rosetta2 cells (Novagen) after transforming the different pET28b(+) plasmids housing the encoding sequences. For each variant, 1 l of LB media supplemented with kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml) was inoculated with a 5-ml preculture and incubated at 37 °C. After an OD_{600} of 0.7 was reached, expression was induced by adding 0.5 mM IPTG, and growth continued for another 14 h at 20 °C. The cells were harvested by centrifugation, for 5 min, at 4000 rpm and 4 °C.

Protein purification and steady-state enzyme kinetics

Cell pellets were resuspended in 25 ml of 10 mM potassium phosphate buffer at pH 7.6, 50 mM NaCl, 5% glycerol, 0.1 mM DTT, 0.1 mM PMSF and 2.5 mg of lysozyme, lysed by sonication (Branson Sonifier 450; 20 s, six times in 30-s intervals, 50% pulse, 4 °C) and centrifuged again (20 min, 11,000 rpm at 4 °C) to separate the soluble from the insoluble fraction of the cell extract. Variants were purified from the soluble cell fraction by loading the cell extract into a nickel Sepharose column (HisTrap FF crude, 5 ml; GE Healthcare) previously equilibrated with 50 mM potassium phosphate buffer, pH 7.6, and 300 mM NaCl. The bound His6-tagged protein was eluted by applying a linear gradient from 1 mM to 300 mM imidazole. Fractions with pure protein were pooled, and these were concentrated in the Amicon Ultra-15 system (MILLIPORE) to a final volume of 3 ml. Next, the samples were loaded on a Superdex 200 column (GE Healthcare) that was previously equilibrated with 50 mM Hepes buffer (pH 7.6) and 100 mM NaCl. Fractions with pure protein were pooled and concentrated in the Amicon Ultra-15 system to a final volume of 3 ml and dialyzed against 3× 1 l degassed 10 mM sodium phosphate buffer (pH 7.6), 1 mM ethylenediaminetetraacetic acid and 1 mM 2-mercaptoethenol at 4 °C. Protein concentrations were quantified using the Bradford method using the BIO-RAD protein assay. Michaelis–Menten enzyme kinetics of PRA isomerase activity were determined using reported protocols[16] with minor modifications.[24] Each data point represents the average of at least three independent experiments using freshly purified enzyme.

Structural studies by CD

Measurements were carried out using a J-715 spectropolarimeter (JASCO) equipped with a Peltier temperature control supplied by JASCO. Eight replicate spectra were collected on each sample to improve the signal-to-noise ratio. The far-UV CD spectra were collected from 190 to 260 nm at 25 °C. Proteins were measured at 0.3 mg/ml and dissolved in 10 mM potassium phosphate buffer at pH 7.6, 1 mM ethylenediaminetetraacetic acid and 1 mM 2-mercaptoethenol. Eight replicate spectra were collected on each sample to improve the signal-to-noise ratio. The spectra were collected in a 0. 01-cm-path-length cell. The thermal denaturation process was analyzed by measuring the change in ellipticity at 220 nm as a function of temperature. The temperature was increased at a rate of 0.3 °C min^{-1}. Thermal denaturation curves were normalized assuming a linear temperature dependence of the baselines for native and denaturated states. The apparent thermal melting temperature ($T_{m,app}$) was determined by finding a midpoint temperature between the native form (linear interpolation of the native region) and the denatured form (either the lowest point or linear interpolation of the unfolded region) on thermal unfolding curves.

Analytical gel filtration

The intermolecular associations were analyzed by size-exclusion chromatography in an Akta FPLC and a Superdex 200 column (GE Healthcare). Purified protein in an initial volume of 0.15 ml was eluted at a flow rate of 0.4 ml min^{-1} over a Superdex 200 column that was equilibrated with 50 mM Hepes buffer (pH 7.6) and 100 mM NaCl. The apparent molecular masses were determined by comparing the protein elution volumes to a calibration curve, which was obtained using proteins from Gel Filtration Standard of BIO-RAD (151-1901).

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2011.05.027

Acknowledgements

We are indebted to Birte Hocker and Frances Arnold for useful discussions and critical reading of the manuscript and to the anonymous reviewers for their constructive comments and very critical suggestions. The authors thank Jorge Yañez, Eugenio López, Santiago Becerra and Paul Gaytán from Unidad de Síntesis y Secuenciación (Instituto de Biotecnología, Universidad Nacional Autónoma de México) for technical assistance on the synthesis of oligonucleotides and sequencing. This work was supported by Mexican Science and Technology Research Council grants 50952-Q and 83039. A.O.-L. thanks doctoral scholarship 201145 from Mexican Science and Technology Research Council.

References

1. Smith, J. M. (1970). Natural selection and the concept of a protein space. *Nature*, **225**, 563–564.
2. Bogarad, L. D. & Deem, M. W. (1999). A hierarchical approach to protein molecular evolution. *Proc. Natl Acad. Sci. USA*, **96**, 2591–2595.
3. Seibert, C. M. & Raushel, F. M. (2005). Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry*, **44**, 6383–6391.
4. Panchenko, A. R. & Madej, T. (2005). Structural similarity of loops in protein families: toward the understanding of protein evolution. *BMC Evol. Biol.* **5**, 10.
5. De Genst, E., Silence, K., Ghahroudi, M. A., Decanniere, K., Loris, R., Kinne, J. *et al.* (2005). Strong *in vivo* maturation compensates for structurally restricted H3 loops in antibody repertoires. *J. Biol. Chem.* **280**, 14114–14121.
6. Weill, J. C. & Reynaud, C. A. (1996). Rearrangement/hypermutation/gene conversion: when, where and why? *Immunol. Today*, **17**, 92–97.
7. Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, **302**, 575–581.
8. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
9. Venkitakrishnan, R. P., Zaborowski, E., McElheny, D., Benkovic, S. J., Dyson, H. J. & Wright, P. E. (2004).

Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry*, **43**, 16046–16055.
10. Heinis, C., Schmitt, S., Kindermann, M., Godin, G. & Johnsson, K. (2006). Evolving the substrate specificity of O^6-alkylguanine-DNA alkyltransferase through loop insertion for applications in molecular imaging. *ACS Chem. Biol.* **1**, 575–584.
11. Akiva, E., Itzhaki, Z. & Margalit, H. (2008). Built-in loops allow versatility in domain–domain interactions: lessons from self-interacting domains. *Proc. Natl Acad. Sci. USA*, **105**, 13292–13297.
12. Bornscheuer, U. (2009). Meeting report: protein design and evolution for biocatalysis. *Biotechnol. J.* **4**, 443–445.
13. Ochoa-Leyva, A., Soberon, X., Sanchez, F., Arguello, M., Montero-Moran, G. & Saab-Rincon, G. (2009). Protein design through systematic catalytic loop exchange in the (beta/alpha)8 fold. *J. Mol. Biol.* **387**, 949–964.
14. Sterner, R. & Hocker, B. (2005). Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem. Rev.* **105**, 4038–4055.
15. Eberhard, M., Tsai-Pflugfelder, M., Bolewska, K., Hommel, U. & Kirschner, K. (1995). Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. *Biochemistry*, **34**, 5419–5428.
16. Hommel, U., Eberhard, M. & Kirschner, K. (1995). Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry*, **34**, 5429–5439.
17. Patrick, W. M. & Blackburn, J. M. (2005). *In vitro* selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. *FEBS J.* **272**, 3684–3697.
18. Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel beta/alpha-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry*, **30**, 9161–9169.
19. Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K. & Sterner, R. (2002). Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry*, **41**, 12032–12042.
20. Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. & Sterner, R. (2000). Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl Acad. Sci. USA*, **97**, 9925–9930.
21. Claren, J., Malisi, C., Hocker, B. & Sterner, R. (2009). Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds. *Proc. Natl Acad. Sci. USA*, **106**, 3704–3709.
22. Barona-Gomez, F. & Hodgson, D. A. (2003). Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* **4**, 296–300.
23. Wright, H., Noda-Garcia, L., Ochoa-Leyva, A., Hodgson, D. A., Fulop, V. & Barona-Gomez, F. (2008). The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. *Biochem. Biophys. Res. Commun.* **365**, 16–21.
24. Noda-Garcia, L., Camacho-Zarco, A. R., Verdel-Aranda, K., Wright, H., Soberon, X., Fulop, V. & Barona-Gomez, F. (2010). Identification and analysis of residues contained on beta→alpha loops of the dual-substrate (beta alpha)8 phosphoribosyl isomerase A specific for its phosphoribosyl anthranilate isomerase activity. *Protein Sci.* **19**, 535–543.
25. Liu, Y. & Kuhlman, B. (2006). RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, W235–W238.
26. Gerstein, M., Lesk, A. M. & Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry*, **33**, 6739–6749.
27. Sun, J. & Sampson, N. S. (1998). Determination of the amino acid requirements for a protein hinge in triosephosphate isomerase. *Protein Sci.* **7**, 1495–1505.
28. Hastings, P. J., Slack, A., Petrosino, J. F. & Rosenberg, S. M. (2004). Adaptive amplification and point mutation are independent mechanisms: evidence for various stress-inducible mutation mechanisms. *PLoS Biol.* **2**, e399.
29. Hendrickson, H., Slechta, E. S., Bergthorsson, U., Andersson, D. I. & Roth, J. R. (2002). Amplification–mutagenesis: evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proc. Natl Acad. Sci. USA*, **99**, 2164–2169.
30. Thoma, R., Hennig, M., Sterner, R. & Kirschner, K. (2000). Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Structure*, **8**, 265–276.
31. Babbitt, P. C. & Gerlt, J. A. (1997). Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591–30594.
32. Gerlt, J. A. & Babbitt, P. C. (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* **70**, 209–246.
33. Hocker, B., Jurgens, C., Wilmanns, M. & Sterner, R. (2001). Stability, catalytic versatility and evolution of the (beta alpha)(8)-barrel fold. *Curr. Opin. Biotechnol.* **12**, 376–381.
34. Peng, T., Zintsmaster, J. S., Namanja, A. T. & Peng, J. W. (2007). Sequence-specific dynamics modulate recognition specificity in WW domains. *Nat. Struct. Mol. Biol.* **14**, 325–331.
35. Baker, D. (2010). An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817–1819.
36. Kempf, J. G., Jung, J. Y., Ragain, C., Sampson, N. S. & Loria, J. P. (2007). Dynamic requirements for a functional protein hinge. *J. Mol. Biol.* **368**, 131–149.
37. Wang, Y., Berlow, R. B. & Loria, J. P. (2009). Role of loop–loop interactions in coordinating motions and enzymatic function in triosephosphate isomerase. *Biochemistry*, **48**, 4548–4556.
38. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, **450**, 913–916.
39. Patrick, W. M. & Matsumura, I. (2008). A study in molecular contingency: glutamine phosphoribosyl-

pyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* **377**, 323–336.
40. Leopoldseder, S., Claren, J., Jurgens, C. & Sterner, R. (2004). Interconverting the catalytic activities of (betaalpha)(8)-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis. *J. Mol. Biol.* **337**, 871–879.
41. Urfer, R. & Kirschner, K. (1992). The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. *Protein Sci.* **1**, 31–45.
42. Xiang, D. F., Kolb, P., Fedorov, A. A., Meier, M. M., Fedorov, L. V., Nguyen, T. T. *et al.* (2009). Functional annotation and three-dimensional structure of Dr0930 from *Deinococcus radiodurans*, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry*, **48**, 2237–2247.
43. Chow, J. Y., Wu, L. & Yew, W. S. (2009). Directed evolution of a quorum-quenching lactonase from *Mycobacterium avium* subsp. *paratuberculosis* K-10 in the amidohydrolase superfamily. *Biochemistry*, **48**, 4344–4353.