Protein Science (1998), 7:1757–1767. Cambridge University Press. Printed in the USA.

Probing enzyme quaternary structure by combinatorial mutagenesis and selection

GAVIN MACBEATH, PETER KAST, AND DONALD HILVERT

Departments of Chemistry and Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

(RECEIVED February 23, 1998; ACCEPTED May 18, 1998)

Abstract

Genetic selection provides an effective way to obtain active catalysts from a diverse population of protein variants. We have used this tool to investigate the role of loop sequences in determining the quaternary structure of a domain-swapped enzyme. By inserting random loops of four to seven residues into a dimeric chorismate mutase and selecting for functional variants by genetic complementation, we have obtained and characterized both monomeric and hexameric enzymes that retain considerable catalytic activity. The low percentage of active proteins recovered from these selection experiments indicates that relatively few loop sequences permit a change in quaternary structure without affecting active site structure. The results of our experiments suggest further that protein stability can be an important driving force in the evolution of oligomeric proteins.

Keywords: chorismate mutase; directed evolution; domain swapping; *Escherichia coli*; four-helix bundle; genetic selection; hinge loop; quaternary structure

Many natural proteins are composed of more than one copy of the same polypeptide. For some of these proteins, quaternary structure appears to have arisen through the process of "domain swapping" (Bennett et al., 1994, 1995). This involves the replacement of a unit of structure, or "domain," in one polypeptide with the same domain from an identical polypeptide. The result is an intertwined dimer or higher order oligomer (Fig. 1). Several proteins have now been found that can exist in both monomeric and domain-swapped oligomeric forms (Piccoli et al., 1992; Parge et al., 1993; Bennett et al., 1994), lending support to the proposal that domain swapping may have played a role in the evolution of some multimeric proteins (Bennett et al., 1994, 1995).

In a domain-swapped protein, the sequence of amino acids that connects the swapped domain with the rest of the polypeptide is referred to as the "hinge loop" (Bennett et al., 1995) (Fig. 1). Several studies have highlighted the role of the hinge loop in controlling the oligomerization state of the protein. For example, changes in the length or sequence of hinge loops have been described that convert monomeric proteins into stable dimers (Di Donato et al., 1994; Green et al., 1995), or dimeric proteins into stable monomers (Mossing & Sauer, 1990; Trinkl et al., 1994; Dickason & Huston, 1996; MacBeath et al., 1998b). Importantly, a high-resolution crystal structure of one such variant—a monomeric version of λ cro (Mossing & Sauer, 1990)—revealed significant conformational changes relative to the wild-type dimer (Albright et al., 1996). This result suggests that in the case of enzymes, where structural changes are often poorly tolerated, mutations that alter quaternary structure without causing a significant loss in catalytic activity may be relatively rare.

In nature, infrequent but beneficial mutations are identified and propagated by genetic selection. Here, we exploit this process to investigate the role of the hinge loop in controlling quaternary structure. As a model protein, we chose the chorismate mutase domain of the bifunctional chorismate mutase-prephenate dehydratase enzyme from *Escherichia coli* (*pheA* gene product). The N-terminal portion of this enzyme folds independently from the C-terminal moiety (Stewart et al., 1990) to form a homodimeric with an unusual six-helix-bundle topology (Lee et al., 1995) (Fig. 2). We have previously shown that residues 1-96 of the *pheA* gene product are sufficient for full chorismate mutase (CM) activity (MacBeath et al., 1998c) and refer to this domain as EcCM (for *E. coli* chorismate mutase).

EcCM catalyzes the conversion of chorismate to prephenate with a rate acceleration of $>10^6$ over the uncatalyzed reaction (Zhang et al., 1996). The rearrangement of chorismate to prephenate represents the first committed step in the biosynthesis of L-tyrosine (Tyr) and L-phenylalanine (Phe) in bacteria, fungi, and higher plants (Haslam, 1993). We have previously prepared a recombinant strain of *E. coli* (KA12/pKIMP-UAUC) that lacks CM activity but has all the other enzymic functions necessary for the biosynthesis of Tyr and Phe (Kast et al., 1996b). As a result, the strain is unable to grow on media lacking these amino acids but regains prototrophy when transformed with an expression plasmid (pKECMB-W) that harbors the gene for EcCM (*pheA'*) (Mac-

Reprint requests to: Donald Hilvert, Laboratorium für Organische Chemie, ETH Zentrum, Universitätstrasse 16, CH-8092 Zürich/Switzerland; e-mail: hilvert@org.chem.ethz.ch.

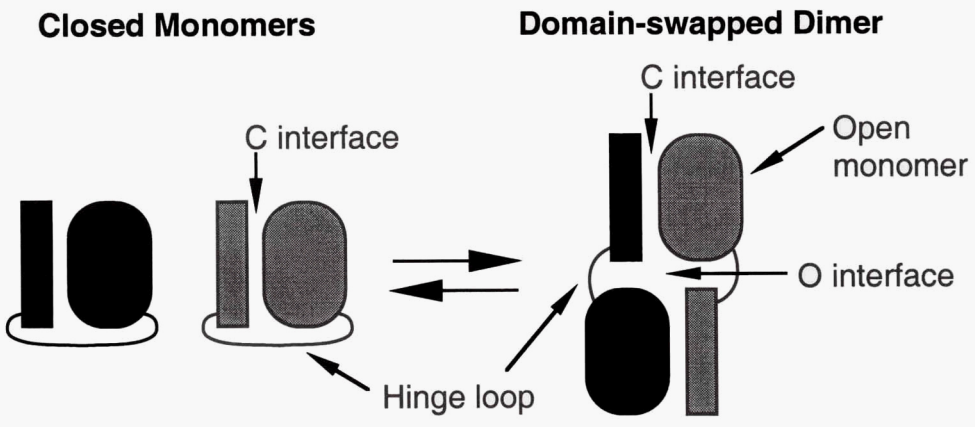

Fig. 1. Domain swapping. A domain-swapped dimer forms when one domain of a monomeric protein is replaced by the same domain from an identical polypeptide. "Domains" can be as large as an entire globular protein domain, as in the diphtheria toxin metastable dimer (Bennett et al., 1994), or as small as a β-strand, as in the CksHs2 dimer (Parge et al., 1993). According to the nomenclature of Bennett et al. (1995), the C interface refers to the contact surface between domains in the closed monomer (preserved in the dimer), the O interface refers to the contact surface between open monomers in the domain-swapped dimer (not present in the closed monomer), and the hinge loop refers to a segment of polypeptide that links the swapped domain to the rest of its subunit.

Beath et al., 1998a). This complementation system thus provides a sensitive test for the conformation of protein variants. If a mutation in EcCM leads to significant structural perturbations in the active site of the enzyme, cells producing this variant will be unable to grow on media lacking Phe and Tyr. If, on the other hand, the mutation does not affect the catalytic function of the enzyme or interfere with protein folding and stability, the cells will grow unhindered. This permits us to direct the evolution of functional variants of EcCM with altered quaternary structure.

Crystallographic studies on EcCM (Lee et al., 1995) show a domain-swapped, dimeric topology (Fig. 2) in which the long, N-terminal helix (H1) spans the two catalytic domains and contributes residues to both active sites. Unlike other domain-swapped dimers, however, EcCM lacks a flexible hinge loop sequence; H1 is a long, unbroken α-helix. We reasoned that if the appropriate loop were inserted into the middle of this helix, the normal dimeric structure could be disrupted and other oligomeric states populated (Figs. 1, 2). Consistent with this notion, we have recently reported the insertion of an 8 amino acid loop into helix H1 of MjCM, a thermostable homolog of EcCM from *Methanococcus jannaschii*, to yield a highly active and monomeric variant (MacBeath et al., 1998b).

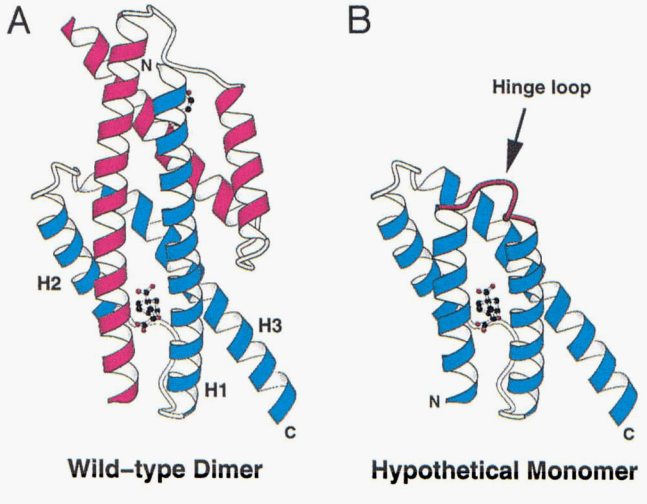

Fig. 2. Changing the quaternary structure of EcCM. **A:** A MOLSCRIPT representation (Kraulis, 1991) of the wild-type, dimeric enzyme is shown on the left, with transition state analog inhibitor bound in the active sites. Alternative oligomerization states may be induced by inserting randomized hinge loop sequences between residues Ala_{23} and Leu_{24} in the middle of helix H1. **B:** A hypothetical structure of a monomeric variant is shown on the right, with the inserted hinge loop colored red.

How important is the specific composition of the inserted hinge loop? Studies on interhelical turns in cytochrome *b*-562 (Brunet et al., 1993; Ku & Schultz, 1995) and ROP (Castagnoli et al., 1994; Vlassi et al., 1994; Predki & Regan, 1995; Nagi & Regan, 1997) have shown that there are relatively few constraints on the sequence or length of interhelical turns. This suggests that a high percentage of loop sequences could be tolerated in EcCM. However, we have shown that, if residues in an interhelical turn are involved in long-range tertiary interactions, the fraction of acceptable sequences may be dramatically reduced (MacBeath et al., 1998a). Also, in the aforementioned experiments with MjCM (the thermostable homolog of EcCM), less than 0.05% of turn sequences gave monomeric proteins with high catalytic activity (MacBeath et al., 1998b). Given the exacting conformational control required for efficient catalysis, hinge loops that change the quaternary structure of EcCM without causing a significant decrease in its catalytic activity are, likewise, expected to be rare. In the study with MjCM, we employed genetic complementation to obtain functional variants; here, we use these same methods to determine how the length of the inserted loop affects quaternary structure in this less stable *E. coli* enzyme.

Results

Design and construction of the hinge loop libraries

Computer modeling suggested inserting a hinge loop between residues Ala_{23} and Leu_{24} of EcCM. This site is in the middle of the H1

helix, dividing the molecule into two equal catalytic domains along the H3-H3′ interface (Fig. 2). Because these residues are relatively exposed to solvent, steric clashes between the insert and the underlying H3 helix are also minimized. To explore the effect of loop length on the quaternary structure of this enzyme, we prepared four different libraries (L4, L5, L6, and L7), with randomized loops of four, five, six or seven residues, respectively.

A unique cloning strategy, designed to avoid several problems associated with PCR-based methods, was employed to construct the libraries. When PCR is used to incorporate oligonucleotides with randomized codons into the gene of interest, a strong bias toward the wild-type sequence may be introduced in the primer annealing step. Heteroduplex DNA may also be formed at this stage, resulting in the production of mixed clones upon transformation. Introduction of unwanted point mutations elsewhere in the gene, by error-prone polymerases employed for PCR, is another common problem (polymerases with 3′ to 5′ exonuclease activity, such as *pfu* polymerase, cannot be used since they may degrade part of the library primer and replace the degraded portion with wild-type sequence). Furthermore, trimming of the ends of PCR products with restriction enzymes needed for cloning often reduces the diversity of the random libraries.

The alternative strategy is illustrated in Figure 3. First, we prepared an inactive version of plasmid pKECMB-W by inserting a piece of "stuffer" DNA into *pheA′*. The resulting plasmid, pKECMB-S2, did not complement the CM-deficiency of KA12/pKIMP-UAUC cells and was used to reduce the occurrence of "false positives" in the libraries. Two adapter primers, ECADB-S and ECADN-S, as well as the library primer, ECL*n*-N (n = 4, 5, 6, or 7), were phosphorylated, mixed in equimolar amounts, and annealed. The annealed primers were then ligated with the appropriate *Bss*HII/*Nhe*I fragment of pKECMB-S2, simultaneously restoring the gene and introducing the randomized hinge loops. The gap encompassing the randomized sequence (Fig. 3) was filled in and the DNA introduced directly into KA12/pKIMP-UAUC cells for genetic selection. About 10^7 transformants were obtained for each library, enabling the assessment of a large number of hinge loops of varying lengths.

Evaluation by genetic selection

Following amplification of the libraries under nonselective conditions, the cells were washed thoroughly, diluted, and plated in duplicate on either LB agar plates (nonselective conditions) or M9c agar plates (selective conditions). The M9c plates were incubated at 30 °C and colonies counted after 6 days and then again after 12 days (Fig. 4; Table 1). It was immediately apparent that only a small percentage of clones in any of the libraries were able to grow on the selective plates, and that the size of colonies varied dramatically under these conditions. The low percentage of complementing clones supports the notion that relatively few hinge loops permit a change in quaternary structure without some perturbation of the protein's tertiary structure. This result also illustrates the value of genetic selection for obtaining rare, functional variants.

It was also clear that the four libraries gave dramatically different results from each other. In libraries L5 and L6, less than 2×10^{-3}% of the clones were able to complement the CM-deficiency, even after 12 days at 30 °C. Library L4 initially appeared similar to L5 and L6. However, after prolonged incubation, 30 to 50 times as many clones were able to grow in library L4 than in either L5 or L6 (Fig. 4; Table 1). By far the highest percentage of complementing clones was observed in the L7 library, where 0.47% of the transformants formed visible colonies on selective plates after 12 days at 30 °C.

A simple explanation of this result is that neither monomeric nor multimeric variants of EcCM are easily formed. Instead, the dimeric structure is maintained and the long, N-terminal α-helix is simply lengthened by the number of inserted residues. This process would move the two helices, H3 and H3′, apart and result in an elongated dimer with unaltered active sites. Since seven residues corresponds exactly to two turns of an α-helix (with a left-handed superhelical twist), the L7 library yields the highest number of functional variants. The insertion of four residues is less well tolerated, since it is slightly out of register with the ideal 3.6 residues per turn, while the insertion of either five or six residues is very poorly tolerated since this would throw the helix severely out of register.

Sequences of selected hinge loops

For sequence analysis, selected clones were picked from each of the four libraries and plasmid DNA recovered. To our surprise, restriction analysis revealed that the gene sequences for most of the complementing clones from libraries L5 and L6 did not contain the expected number of inserted codons. Instead, they matched the length of the wild-type *pheA′*. DNA sequences were obtained for 14 such clones (dubbed "escape mutants") from library L5 (data not shown). All 14 mutants had no inserted bases at the expected site but contained various point mutations in the vicinity. In most cases, the corresponding protein sequences differed from that of wild-type EcCM only in the identity of residue 24. Restriction

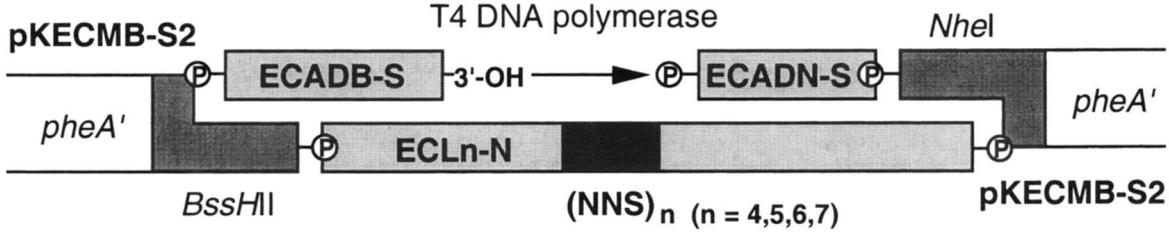

Fig. 3. Construction of the hinge loop libraries. Oligonucleotides ECADB-S, ECADN-S, and ECL*n*-N (n = 4, 5, 6, or 7) were phosphorylated, annealed and ligated into the appropriate *Bss*HII/*Nhe*I fragment of pKECMB-S2. The resulting gap was filled in using T4 DNA polymerase and the DNA introduced into KA12/pKIMP-UAUC cells for genetic selection.

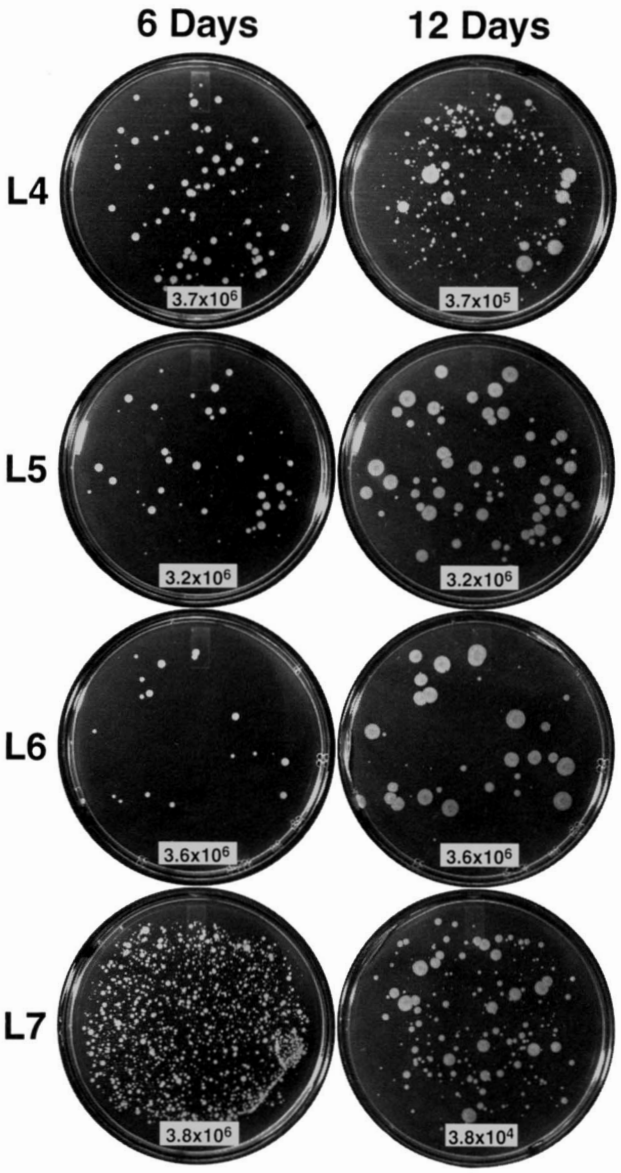

Fig. 4. Evaluation of the hinge loop libraries by genetic selection. Cells were washed, plated on minimal media (M9c), and incubated at 30 °C for the indicated length of time. The number of transformed cells that were spread on each plate (indicated in the white box) was determined by spreading an appropriate dilution of cell suspension on rich media (LB) and counting the colonies after incubation for 3 days at 30 °C (not shown).

analysis of 30 selected clones from library L5 (both large and small colonies) revealed only two poorly complementing variants with an appropriate insert. No plasmids of the correct size were identified among 30 selected L6 clones. This means that the true percentages of five- and six-residue hinge loops that permit in vivo activity are at least 10-fold lower than the estimates reported in Table 1, further emphasizing the deleterious effects of these insertions. Restriction analysis of selected clones from library L4 showed that the fastest growing clones were also predominantly escape mutants. However, many more clones appear after 12 days of incubation and the majority of these variants contain the expected 4-codon insert. No escape mutants were observed among the selected clones from library L7. Based on these observations, we

Table 1. *Evaluation of the hinge loop libraries by genetic complementation*

Library	% Complementing clones visible after 6 days	% Complementing clones visible after 12 days
L4	0.0014	0.056
L5	0.0016	0.0018
L6	0.00060	0.0011
L7	0.036	0.47

estimate that about 1 in 10^5 clones are escape mutants in each library. This low level of background only becomes limiting for libraries with an extremely small number of active clones. The mechanism by which these mutants are formed is unclear, but likely involves error-prone repair of gapped DNA within the cell. Additional experiments will be necessary to clarify the extent to which the new library construction strategy contributes to these background levels.

Sequences were obtained for the two marginally active L5 variants and for 12 active full-length variants from each of the L4 and L7 libraries (Table 2). A fairly strong consensus was observed among the L4 sequences: [small Ile (or similar) Val (or similar) hydrophilic], with a few exceptions. The strong consensus is not surprising given the low percentage of complementing clones (Table 1). While only two L5 clones were obtained, their sequences are also fairly similar, particularly in the central three residues of the insert (Table 2). Less similarity was apparent among the L7 variants, consistent with the higher percentage of complementing clones in this library (Table 1). Nevertheless, a weak consensus is observed: [*X*-hydrophobic-small-hydrophilic-hydrophobic-small-hydrophilic] (where *X* is any amino acid). Interestingly, the consensus does not clearly match the pattern expected for a simple extension of helix H1 by two helical turns. In the wild-type dimer, H1 forms a coiled-coil with H1′, and the first and fifth residues of the inserted loop lie at the "d" and "a" positions of this motif respectively (Creighton, 1993). In natural coiled-coils, the "d" position is predominantly occupied by Leu or Ala, the "a" position by Leu, Ile, Ala, or Val, and the other positions by more hydrophilic residues (Cohen & Parry, 1990). However, most of the clones in Table 2 do not conform to this expectation, suggesting that the simple explanation for the statistical trends of Table 1 may be incorrect. We therefore elected to study one randomly chosen variant from each of the L4, L5, and L7 libraries in greater detail.

Protein production, purification, and refolding

Efficient production of wild-type EcCM and the three hinge loop variants L4-1, L5-1, and L7-1 (Table 2) was achieved using the T7 expression system (Studier et al., 1990). The genes for EcCM and the three variants were each subcloned into pET-22b-pATCH, a derivative of pET-22b (Novagen, Madison, Wisconsin) that has been modified to prevent unwanted translational read-through at the TGA stop codon (MacBeath & Kast, 1998). The resulting plasmids (pET-EcCM-pATCH, pET-L4-1-pATCH, pET-L5-1-pATCH, and pET-L7-1-pATCH, respectively) were introduced into host strain KA13,

Table 2. *Hinge loop sequences of selected clones from libraries L4, L5, and L7*

Clone	DNA sequence	Protein sequence[a]
L4-1	TCG ATC GTC CAG	Ser Ile Val Gln
L4-2	GCG ATC GTG CAG	Ala Ile Val Gln
L4-3	TGC TTG TGC TGC	Cys Leu Cys Cys (+Leu_{24}Phe)
L4-4	ACG ATC GTC CGG	Thr Ile Val Arg
L4-5	ACC ATC GTC CAG	Thr Ile Val Gln
L4-6	GCC ATA GTG AGG	Ala Ile Val Arg (+Ala_{23}Val)
L4-7	ACC ATC ATC GAA	Thr Ile Ile Glu
L4-8	AAC ATC GCC GCG	Asn Ile Ala Ala
L4-9	TGT CTT GTC CTG	Cys Leu Val Leu (+Leu_{21}Ile)
L4-10	TGC ATC TGC CGC	Cys Ile Cys Arg
L4-11	TCC GTC GTG CTG	Ser Val Val Leu
L4-12	TTG ATC GTC CGG	Leu Ile Val Arg
L5-1	TGC TTC CCG TGG GAC	Cys Phe Pro Trp Asp (+Ala_{16}Glu)
L5-2	ATC TAC CCC TTC GCC	Ile Tyr Pro Phe Ala
L7-1	ACC GCC GCG AGC CTC ACC AGC	Thr Ala Ala Ser Leu Thr Ser
L7-2	GCG TGC GCC ACC AGC CTC AGG	Ala Cys Ala Thr Ser Leu Arg
L7-3	CAC TTC TCG CAC ATG TCG AAG	His Phe Ser His Met Ser Lys
L7-4	TTG TTC TCC AGC ATG GCG CAG	Leu Phe Ser Ser Met Ala Gln
L7-5	AGC TTC GCG ACC TGC ACC AAG	Ser Phe Ala Thr Cys Thr Lys
L7-6	ACC GCG CAC AAC CTC GCC CTG	Thr Ala His Asn Leu Ala Lys
L7-7	TGC AGC GCG ACC ATC CTG ACC	Cys Ser Ala Thr Ile Leu Thr
L7-8	CTC CTC GGC CAG GTC TCC TCG	Leu Leu Gly Gln Val Ser Ser
L7-9	ATC CTC GCG GAG CTC AGC CAG	Ile Leu Ala Glu Leu Ser Gln
L7-10	CTC GCG GCG AAC ATG GCC CAG	Leu Ala Ala Asn Met Ala Gln
L7-11	GAG TTC TCG CTC ATC TGC TCG	Glu Phe Ser Leu Ile Cys Ser
L7-12	GTG CTG GTG ACG CTC ACG GAG	Val Leu Val Thr Leu Thr Glu

[a]Clones L4-3, L4-6, L4-9, and L5-1 also contained single point mutations producing the indicated amino acid changes.

a *recA*-deleted K-12 strain of *E. coli*, which is devoid of endogenous *E. coli* CM activity. Upon induction of the chromosomally integrated T7 RNA polymerase gene with IPTG, high yields of soluble protein were obtained for EcCM and L5-1. In contrast, both L4-1 and L7-1 were recovered as insoluble inclusion bodies.

Proteins produced using pET-22b-pATCH derivatives contain a C-terminal $(His)_6$-tag that enables efficient purification by affinity chromatography on a matrix containing chelated Ni^{2+} ions (Van Dyke et al., 1992). All four proteins were purified in this way (EcCM and L5-1 under native conditions and L4-1 and L7-1 under denaturing conditions), yielding samples of >98% purity as judged by SDS-polyacrylamide gel electrophoresis and Coomassie blue staining. Analysis by electrospray ionization mass spectrometry (ESI-MS) confirmed the identity of each protein and indicated that the N-terminal methionine residue (Met_1) had been removed in the EcCM and L5-1 samples but was intact in the L4-1 and L7-1 samples. All experimentally determined masses correlated well with the calculated masses (EcCM [-Met_1]: calcd 11,886.7, obsd 11,885; L4-1: calcd 12,445.4, obsd 12,444; L5-1 [-Met_1]: calcd 12,593.5, obsd 12,592; L7-1: calcd 12,649.6, obsd 12,651).

Analytical size-exclusion column chromatography showed that wild-type EcCM preparations were composed of a mixture of correctly folded, dimeric protein (~50%) and various misfolded, higher-order aggregates (remaining 50%). The higher-order aggregates were converted into fully active, dimeric protein by denaturation with guanidinium chloride and subsequent renaturation upon dilution into native buffer. The efficiency of refolding was >95% as judged by analytical gel filtration and native polyacrylamide gel electrophoresis. Kinetic analysis of the refolded EcCM at 20 °C indicated a k_{cat} of 16.3 s^{-1} and a K_m of 290 μM, which is in general agreement with the reported values of 9.0 s^{-1} and 300 μM, respectively (Galopin et al., 1996). (NB: The value for k_{cat} of 9.0 s^{-1} was calculated from the reported values of $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ with T = 293.15 K.)

Based on the successful refolding of EcCM, the three hinge loop variants were treated in the same way. A significant amount of precipitated protein (~25–40%) accumulated during the refolding of L4-1 and L7-1, but L5-1 remained entirely in solution. After removing the precipitated material by centrifugation, the variants were examined by gel filtration. In all three cases, a predominant species was observed (>80% of the sample), which was then isolated for further analysis. (NB: Samples of L5-1 purified under either native or denaturing conditions behave similarly when subjected to nondenaturing polyacrylamide gel electrophoresis and size exclusion chromatography, suggesting that the refolding procedure does not alter the predominant oligomerization state.)

Structural and kinetic characterization

The oligomerization states of EcCM and the three variants were determined by analytical ultracentrifugation. Figure 5 shows sedimentation equilibrium data for all four proteins, loaded at an initial concentration of 10 μM. As expected, EcCM behaves as a dimer. The data fit best to a single ideal species model (Equation 1)

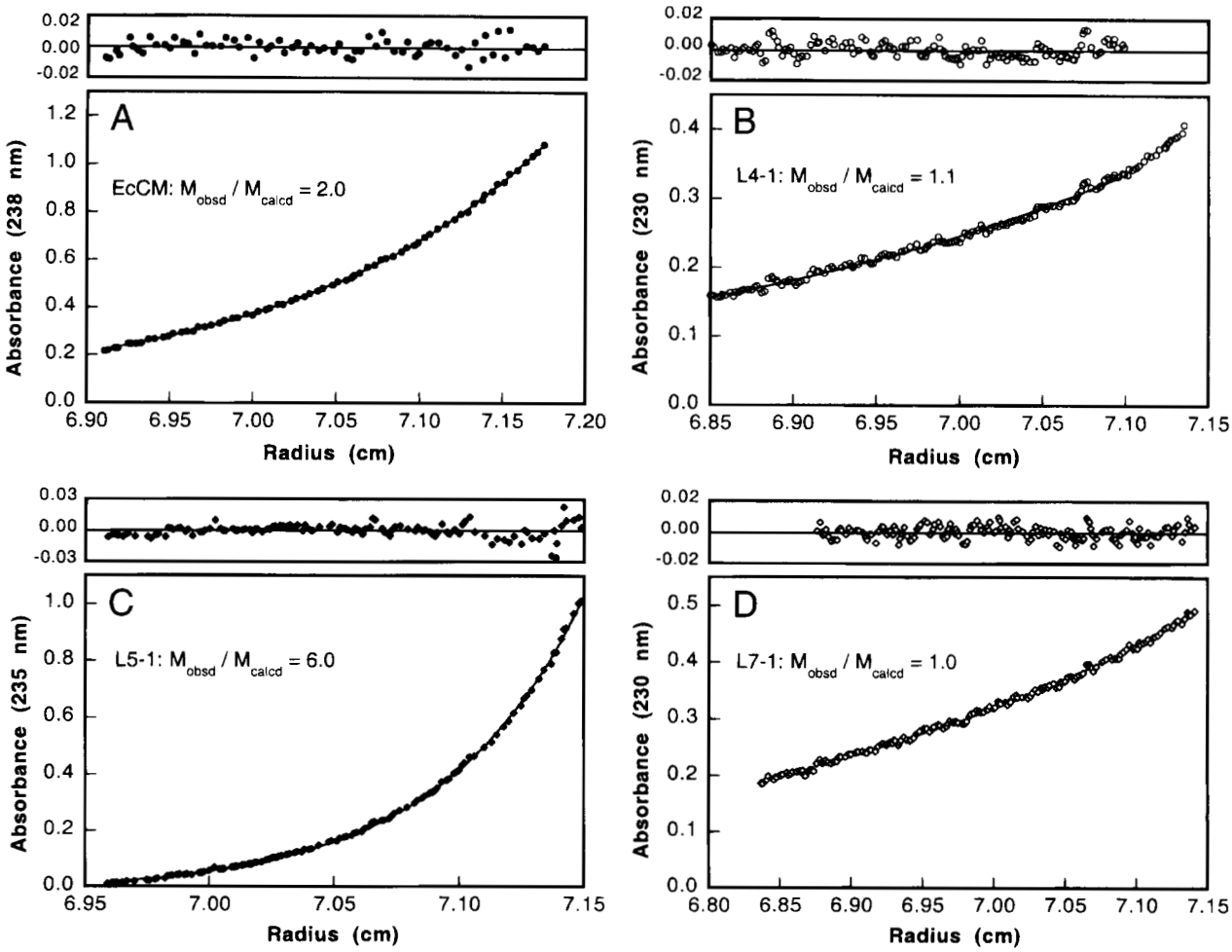

Fig. 5. Sedimentation equilibrium analysis of EcCM and variants. Data were obtained at 17,000 rpm, 20 °C, with 10 μM protein in PBS. Lines indicate least-squares fitting to a single ideal species model (see Materials and methods). The residuals from these fits are shown in the windows above the plots. (**A**) EcCM, (**B**) L4-1, (**C**) L5-1, (**D**) L7-1.

with an average molecular mass of 23,800 g/mol (Fig. 5A). This compares very well with the expected value of 23,774 g/mol, calculated for the EcCM dimer based on amino acid composition.

Sedimentation of L4-1 and L7-1, on the other hand, gave significantly different results. In both cases, protein material was lost during the course of the experiment (about 50% loss), presumably because of aggregation and subsequent removal by centrifugation. This is consistent with the observation that both L4-1 and L7-1 precipitate slowly upon prolonged storage, particularly at high concentration. Nevertheless, sufficient material remained in solution to permit analysis of the soluble protein. Contrary to our expectation that the L4 and L7 variants are dimeric (see above), both proteins were clearly monomeric at the concentrations employed for the sedimentation analysis (initial concentration: ~10 μM). In each case, the data fit best to a single ideal species model, yielding average molecular masses of 13,700 g/mol for L4-1 (calculated: 12,445 g/mol) and 12,900 g/mol for L7-1 (calculated: 12,650 g/mol). Thus, it appears that monomeric versions of EcCM have been obtained. However, these monomers are relatively unstable and aggregate readily.

In contrast to L4-1 and L7-1, L5-1 did not precipitate upon prolonged storage. Likewise, the sedimentation data did not show the same loss of material during the course of the experiment. Unlike L4-1 and L7-1, however, L5-1 is not monomeric. In fact, the data fit best to a single ideal species model with an average molecular mass of 75,800 g/mol (Fig. 5C). This corresponds well to the molecular mass expected for a hexameric species (75,564 g/mol, calculated from amino acid composition). Apparently, insertion of this five-residue hinge loop favors the formation of a well-defined, higher-order oligomer rather than a marginally stable monomer (as with L4-1 and L7-1).

To investigate the secondary structure of the hinge loop variants, far UV circular dichroism (CD) spectra were obtained for each mutant as well as for the wild-type enzyme (Fig. 6). The spectra of L4-1, L5-1, and L7-1 are roughly similar in shape to that of EcCM and exhibit features indicative of high α-helical content (Greenfield & Fasman, 1969; Chen et al., 1974). Nevertheless, the intensities of the spectra are substantially diminished in all three variants relative to EcCM. This likely arises from a combination of effects including the insertion of non-helical residues, perturbations in the secondary structure around the inserted hinge loops, and decreased protein stability (L4-1 and L7-1). Potential inaccuracies in the determination of protein concentration may also contribute to intensity differences. Since EcCM has no Trp residues and only one

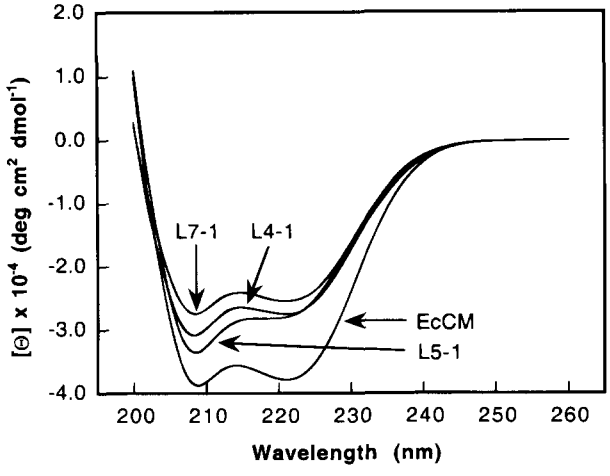

Fig. 6. Circular dichroism spectra of EcCM and variants. Spectra were obtained at 20 °C with 5 μM protein in PBS.

Tyr, the concentration could not be determined accurately by UV absorption due to the potential contribution from minor, strongly absorbing protein contaminants. All concentrations were consequently determined using the Micro BCA Protein Assay (Pierce, Rockford, Illinois) with bovine serum albumin (BSA) as the calibration standard. In terms of spectral shape, L5-1 appears the most dissimilar to EcCM, suggesting larger structural perturbations in this variant. This is consistent with its dramatically altered oligomeric state.

When assayed for catalytic activity, all three topological variants demonstrated saturation kinetics. Steady-state parameters, k_{cat} and K_m, were determined for each mutant at 20 °C and are given in Table 3. For all measurements, 0.1 mg/mL of BSA was included in the reaction buffer to stabilize the proteins at low concentrations. Interestingly, the monomeric variant L7-1 is almost as active as EcCM, indicating very little structural perturbation in the active site of the enzyme despite its topological transformation and structural instability. A threefold decrease in k_{cat} was observed for variant L4-1 relative to EcCM, while L5-1 suffered a drop of 100-fold in activity. For each variant, K_m did not deviate substantially from that observed for the wild-type enzyme (no more than 2.4-fold). The reduced activity of L5-1 is not surprising, given its dramatically altered quaternary structure. It is also consistent with the less effective genetic complementation observed for this mutant (data not shown).

Since relatively low enzyme concentrations (50–400 nM) were employed to determine the kinetic parameters of the hinge loop variants (see Materials and methods), it is important to establish the relevance of these data to the oligomerization states determined at higher concentrations. This was done by measuring the activity of each variant (under k_{cat}/K_m conditions) over a wide range of protein concentrations (Fig. 7). For EcCM, L5-1, and L7-1, the observed activity was essentially independent of protein concentration, suggesting that their oligomerization states remain unchanged in this regime, or that activity is independent of the state of oligomerization. L4-1 exhibited a slight decrease in activity with decreasing concentration. However, the effect is small (less than twofold change between 400 nM and 25 nM) and probably reflects the instability of the protein at low concentrations.

Table 3. *Kinetic parameters for EcCM and the three hinge loop variants*[a]

Protein	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
EcCM	16	290	56,000
L4-1	5.7	700	8,100
L5-1	0.16	600	270
L7-1	9.8	230	42,000

[a]Kinetic parameters were determined at 20 °C in PBS.

Discussion

Based on the observed data, a model for the oligomeric forms of the EcCM variants is presented in Figure 8. In the wild-type protein (Fig. 8A), the absence of a hinge loop leads to the formation of a stable dimer. Hydrophobic interactions at the subunit interface, particularly between residues on helices H3 and H3′ (Fig. 2), serve to stabilize the structure. In fact, most of the hydrophobic interactions in the wild-type dimer occur at this interface since the highly polar active sites of the enzyme are located in the core of each four-helix-bundle motif (Fig. 2).

In variants L4-1 and L7-1 (Fig. 8B,D), the insertion of a hinge loop leads to the formation of a marginally stable monomer. Since the extensive interactions at the subunit interface of the wild-type enzyme are lost in the transformation, the equilibrium between native and denatured protein is shifted significantly toward the unfolded state. The unfolded polypeptide then aggregates, either nonspecifically or by a mechanism involving various domain-swapped intermediates. Aggregation is essentially irreversible, pulling the equilibrium toward the unfolded protein. This explains the progressive precipitation of both L4-1 and L7-1.

In the case of L5-1, insertion of a five-residue hinge loop may lead to the formation of a stable domain-swapped hexamer (Fig. 8C). Topological constraints require the hinge loops to be positioned in the interior of the complex, where they could interact to form a stable core. Consistent with this notion, the selected loop of L5-1 (Cys-Phe-Pro-Trp-Asp) is strongly hydrophobic (the Asp

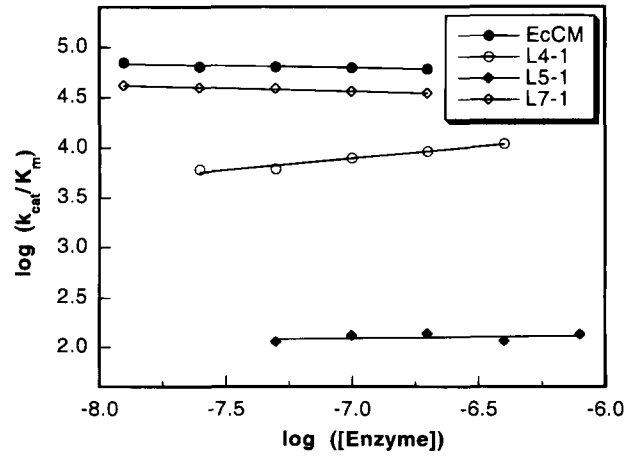

Fig. 7. Dependence of catalytic activity on protein concentration for EcCM and the three hinge loop variants.

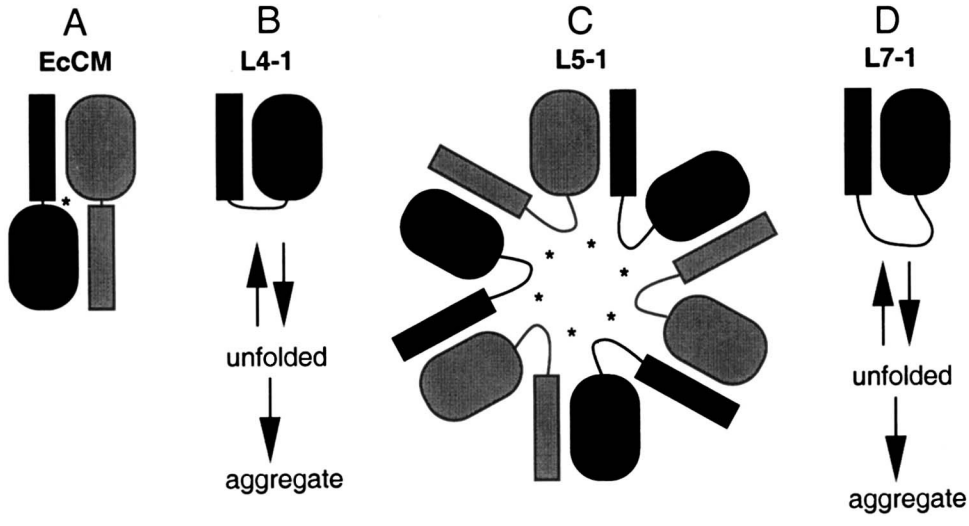

Fig. 8. Proposed model for the structural organization of the hinge loop variants. * designates stabilizing, hydrophobic interactions.

occurs at the end of the loop and hence may not be involved in this interaction). It is interesting to note that the selected loop of L5-2 (Ile-Tyr-Pro-Phe-Ala; Table 2) is also very hydrophobic and purified samples of this protein co-elute with L5-1 on an analytical size exclusion column (data not shown).

The hexameric structure of L5-1 is particularly interesting for two reasons. First, the protein exists predominantly in one oligomerization state. This means that the free energy of the hexamer is significantly lower than that of other oligomeric forms (assuming thermodynamic control). Evidently, specific stabilizing interactions are formed in the hexamer that are absent in alternative structures. Secondly, it is clear that the hexamer is more stable to aggregation and precipitation than the L4-1 or L7-1 monomers. This has intriguing implications concerning the evolution of oligomeric proteins (Janin et al., 1988; Goodsell & Olson, 1993; Bennett et al., 1995). It appears that, in this case, oligomerization offers an easy way to increase the stability of the enzyme. Since the highly polar active site is located in the core of the four-helix-bundle domain, the isolated monomer is relatively unstable. However, when several monomeric units are brought together, a new hydrophobic core can form that serves to stabilize the overall structure. The hexamer, whose efficiency is significantly decreased relative to the wild-type dimer, may also serve as an excellent starting point for the evolution of more active catalysts using our genetic complementation system.

What, then, is the origin of the different percentages of complementing clones in the four libraries (Table 1)? Based on the model presented in Figure 8, we can offer the following (speculative) explanation. Insertion of a four-residue hinge loop in EcCM is insufficient to form a higher-order oligomeric structure. Either the loop is too short to span the distance between domains, or it is insufficient to form a stable core in the center of the complex. As a result, active variants are forced to adopt the unstable monomeric topology. The short loop is not ideal for this purpose, yielding a relatively low percentage of active clones. In contrast, five- (and perhaps six-) residue loops are able to form stable oligomeric structures (Fig. 8C). However, complexation results in structural perturbations that render the protein less active. Extremely few variants have sequences appropriate for the formation of a catalytically competent complex. In addition, assembly of this complex may be inefficient in the cell, resulting in a very low percentage of complementing clones. When the loop is extended to seven residues, it becomes too long to fold into a compact core in the center of the complex. As a result, the protein is forced to adopt the less stable, but more active, monomeric topology. Since it is not constrained by short loop length, a high percentage of clones are able to complement the CM deficiency.

The marginal stability of L4-1 and L7-1 is particularly interesting in the context of a closely related experiment with MjCM (a thermostable homolog of EcCM). We have recently reported the design of a monomeric variant of MjCM, which contains an eight-residue hinge loop in the middle of helix H1 (MacBeath et al., 1998b). This monomer is much better behaved than both L4-1 and L7-1. Denaturation studies have shown, however, that it is only stabilized by 2.7 kcal/mol (MacBeath et al., 1998b). Since EcCM is about 5 kcal/mol less stable than MjCM (MacBeath et al., 1998c), it is perhaps not surprising that monomeric variants of EcCM are so unstable. Nevertheless, we note that the wild-type dimeric structure was not adopted by either of these clones. Like many analogous proteins (Neet & Timm, 1994), EcCM derives most of its stability from quaternary, rather than tertiary, structure (MacBeath et al., 1998c). Since many of the stabilizing interactions formed between its subunits are disrupted by the extension of helix H1, the dimeric topology is no longer favored in the hinge loop variants.

In summary, we have exploited genetic selection to explore the role of the hinge loop in controlling the process of domain swapping. By combining structure-based analysis with in vivo selection, we have altered the quaternary structure of a dimeric enzyme while retaining considerable catalytic efficiency. The low percentage of complementing clones in each library indicates that relatively few sequences permit a change in oligomerization state without compromising the activity of the resulting enzyme. Moreover, the different oligomerization states obtained in this study highlight the importance of the hinge loop length in determining quaternary structure. Most significantly, it appears that the major reason for the emergence of oligomers, at least in this case, is the stability of the resulting protein. It is interesting to note that all

other natural CMs studied to date are multimeric and our results suggest that increasing protein stability by formation of a larger hydrophobic core absent in the monomer may have played an important role in the evolution of these, and perhaps other, oligomeric proteins.

Materials and methods

E. coli strains

General cloning was carried out in *E. coli* strain XL1-Blue (Stratagene, La Jolla, California). Genetic selection experiments were performed using the CM-deficient *E. coli* strain KA12/pKIMP-UAUC (Kast et al., 1996a, 1996b). The genotype of KA12 is *Δ(srlR-recA)306::Tn10*, *Δ(pheA- tyrA-aroF)*, *thi-1*, *endA-1*, *hsdR17*, *Δ(argF lac)U169*, *supE44*. The plasmid pKIMP-UAUC confers chloramphenicol resistance and provides expression of genes encoding monofunctional forms of prephenate dehydrogenase and prephenate dehydratase from *Erwinia herbicola* and *Pseudomonas aeruginosa*, respectively. For biophysical and kinetic analysis, proteins were produced in *E. coli* strain KA13 (MacBeath & Kast, 1998). This strain was constructed by site-specific integration of λDE3 prophage into the chromosome of strain KA12 and so carries an IPTG-inducible gene for T7 RNA polymerase (Studier & Moffatt, 1986).

DNA manipulations

All nucleic acid manipulations were according to standard procedures (Sambrook et al., 1989). Restriction endonucleases, T4 DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs (Beverly, Massachusetts). Oligonucleotides were obtained by custom synthesis from Life Technologies (Gaithersburg, Maryland) and purified by polyacrylamide gel electrophoresis. DNA sequencing was performed on an Applied Biosystems 373 Automated DNA Sequencer using dye terminator nucleotides and chain termination chemistry (Sanger et al., 1977). DNA was prepared for sequencing using a QIAGEN Miniprep kit (Hilden, Germany).

Construction of plasmids

The construction of plasmid pKECMB-W has been described previously (MacBeath et al., 1998a). It confers ampicillin resistance and harbors the 3′-truncated *pheA* gene (*pheA′*) of *E. coli* (Stewart et al., 1990) under control of the *bla* promoter. The *pheA′* gene encodes the first 96 residues (CM domain) of the bifunctional chorismate mutase-prephenate dehydratase enzyme of *E. coli*.

To eliminate any wild-type background from re-ligated vector during construction of the hinge loop libraries, a plasmid was constructed in which the 32 bp *Nhe*I-*Hin*dIII fragment of pKECMB-W (MacBeath et al., 1998a) was replaced by a 466 bp *Nhe*I-*Hin*dIII fragment of unrelated DNA. The 466 bp fragment serves as a piece of "stuffer DNA," intended to destroy the function of the EcCM gene. The resulting plasmid pKECMB-S2 (3,346 bp) showed no detectable complementation of the CM auxotrophy in KA12/pKIMP-UAUC cells, as expected.

For production of proteins on a large scale, the relevant genes were expressed from the T7 promoter. pET-EcCM (5,652 bp) was constructed by ligating the 288 bp *Nde*I-*Xho*I fragment from pKECMT-W (MacBeath et al., 1998a) with the 5,364 bp *Nde*I-*Xho*I fragment from the T7-promoter vector pET-22b(+) (Novagen, Madison, Wisconsin). We have subsequently found that many pET vectors cause unwanted translational read-through at the TGA stop codon (MacBeath & Kast, 1998). To circumvent the problem of heterogeneous protein production, we have constructed pET-22b-pATCH, a derivative of pET-22b(+) by following the protocol outlined elsewhere (MacBeath & Kast, 1998). pET-EcCM-pATCH (5,668 bp) was obtained in an analogous way from pET-EcCM. pET-L4-1-pATCH (5,680 bp), pET-L5-1-pATCH (5,683 bp), and pET-L7-1-pATCH (5,689 bp) were constructed by ligating the respective 300 bp, 303 bp, and 309 bp *Nde*I/*Xho*I fragments from pKECMB-L4-1, pKECMB-L5-1, and pKECMB-L7-1 (obtained from libraries L4, L5, and L7) with the 5,380 bp *Nde*I/*Xho*I fragment from pET-22b-pATCH.

Construction of the hinge loop libraries

The strategy for library construction is shown in Figure 3. Codons were randomized using the scheme NNS (N = A/C/G/T; S = G/C). Primers ECADB-S [CGCGCTGGATGAAAAATTATTAG], ECADN-S [TGGCAGAACGGCGCGAG], and ECL*n*-N [CTAG CTCGCGCCGTTCTGCCAGTAA(**SNN**)$_n$CGCTAATAATTTTC ATCCAG; n = 4, 5, 6, or 7] were phosphorylated with T4 polynucleotide kinase for 2 h at 37 °C (300 pmol primer, 1 mM ATP in a 50 μL reaction). Phosphorylated primers were mixed at a final concentration of 2 μM each and annealed by boiling for 3 min, followed by slow cooling to <25 °C over 1 h. The annealed primers were mixed with the 2,866 bp *Bss*HII/*Nhe*I fragment of pKECMB-S2 at a molar ratio of 10:1 and ligated with T4 DNA ligase at 16 °C overnight. Following heat inactivation of the ligase at 65 °C for 5 min, the gapped DNA was filled in with T4 DNA polymerase and 2 mM of each dNTP (4 °C, 5 min; 25 °C, 5 min; 37 °C, 2 h; 75 °C, 10 min). The nicked DNA was ligated with T4 DNA ligase and prepared for electroporation by desalting with a QIAquick PCR purification column (Hilden, Germany). Transformation of electrocompetent KA12/pKIMP-UAUC cells gave ~10^7 transformants for each library (L4: 9.3×10^6, L5: 8.8×10^6, L6: 1.4×10^7, L7: 1.3×10^7). The transformed cells were amplified for 16 h at 30 °C in 100 mL LB broth (Miller, 1972) supplemented with 150 μg/mL sodium ampicillin and 35 μg/mL chloramphenicol and then aliquotted and frozen in 20% glycerol at −80 °C.

Selection experiments

Selection experiments were conducted using M9c medium (Kast et al., 1996b), which is based on M9 minimal medium (Miller, 1972). It consisted of Na_2HPO_4 (60 mg/mL)/KH_2PO_4 (30 mg/mL)/NH_4Cl (10 mg/mL)/NaCl (5 mg/mL)/0.2% (wt/vol) D-(+)-Glc/1 mM $MgSO_4$/0.1 mM $CaCl_2$/thiamin-HCl (5 μg/mL)/4-hydroxybenzoic acid (5 μg/mL)/4-aminobenzoic acid (5 μg/mL)/2,3-dihydroxybenzoic acid (1.6 μg/mL)/L-Trp (20 μg/mL)/sodium ampicillin (150 μg/mL)/chloramphenicol (35 μg/mL), pH 7.0. For Petri dishes, 15 g agar was added per liter. Where required, M9c was supplemented with 20 μg of L-Phe and L-Tyr per mL. For complementation tests on individual clones, well-isolated colonies were picked from either M9c plates or LB plates (supplemented with 150 μg/mL sodium ampicillin and 35 μg/mL chloramphenicol) and streaked on M9c plates to yield individual colonies. For selection experiments with the libraries, 1 mL aliquots of cells were thawed, washed four times with M9c, diluted with M9c and plated in duplicate both on M9c plates and on LB

plates supplemented with 150 μg/mL sodium ampicillin and 35 μg/mL chloramphenicol. In all cases, plates were incubated at 30 °C. Colonies on M9c plates were counted and evaluated after 6 days and then again after 12 days. Colonies on LB plates were counted after 3 days.

Protein production, purification, and refolding

EcCM, L4-1, L5-1, and L7-1 were produced with plasmids pET-EcCM-pATCH, pET-L4-1-pATCH, pET-L5-1-pATCH, and pET-L7-1-pATCH, respectively. The host strain for protein production was KA13. Cells from a single colony were grown in 500 mL LB medium (supplemented with 150 μg/mL sodium ampicillin) at 37 °C up to an OD_{600} of 0.8. The cultures were cooled to room temperature and isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 0.4 mM. After 16 h induction at room temperature, the cells were harvested and resuspended in 20 mL PBS (10 mM phosphate, 160 mM NaCl, pH 7.5) supplemented with 100 μM phenylmethylsulfonyl fluoride (PMSF) and 2 μg/mL pepstatin A and aprotinin. Following cell lysis by passage through a French press, insoluble material was separated from soluble material by centrifugation (28,000 × g, 20 min, 4 °C).

Both EcCM and L5-1 were found entirely in the soluble fraction (supernatant) and were purified as follows. 200 μL of PBS containing 500 mM imidazole was added to the supernatant (final concentration of 5 mM imidazole) and the solution was loaded onto a column packed with 5 mL of Ni-NTA agarose (QIAGEN, Hilden, Germany) that had been pre-equilibrated with PBS containing 5 mM imidazole. The column was thoroughly washed with PBS containing 30 mM imidazole and the bound protein was subsequently eluted with PBS containing 250 mM imidazole.

Proteins L4-1 and L7-1 were found primarily in the insoluble fraction following cell lysis. The proteins were dissolved in 20 mL PBS containing 6 M urea, 5 mM imidazole, and 10 mM 2-mercaptoethanol and any remaining insoluble material was removed by centrifugation, as before. The supernatant containing the solubilized protein was then loaded onto a column packed with 5 mL of Ni-NTA agarose that had been pre-equilibrated with PBS containing 6 M urea and 5 mM imidazole. The column was washed thoroughly with PBS containing 6 M urea and 30 mM imidazole and the bound protein was subsequently eluted with PBS containing 6 M urea and 250 mM imidazole.

Prior to refolding, EcCM and L5-1 were denatured by the addition of 4 volumes of PBS containing 8 M guanidinium chloride (GdmCl). Since L4-1 and L7-1 were purified under denaturing conditions, no further treatment was necessary. All four proteins were refolded by 20-fold dilution into PBS (final concentration < 6 μM protein). The proteins were concentrated by ultrafiltration (Amicon stirred-cell concentrator, Millipore, Bedford, Massachusetts) and dialyzed extensively against PBS. Purified, refolded proteins were stored in PBS supplemented with 100 μM PMSF, 2 μg/mL pepstatin A, 2 μg/mL aprotinin, and 0.02% sodium azide. Prior to biophysical or kinetic analysis, the protease inhibitors, preservatives, and higher-order aggregates of the proteins were removed by size exclusion column chromatography using a Superose 12 (10/30) FPLC column (Amersham Pharmacia Biotech, Uppsala, Sweden) with PBS as the running buffer. Protein concentration was determined using the Pierce Micro BCA Protein Assay Reagent (Pierce, Rockford, Illinois) with bovine serum albumin as the calibration standard.

Mass spectrometry

Proteins were prepared for mass spectrometry by desalting on a NAP-5 column (Amersham Pharmacia Biotech, Uppsala, Sweden) that had been pre-equilibrated in 1% aqueous acetic acid. Electrospray ionization mass spectrometry (ESI-MS) was performed on an API III Perkin Elmer SCIEX triple quadrupole mass spectrometer.

Circular dichroism spectroscopy

All circular dichroism (CD) experiments were performed on an Aviv Circular Dichroism Spectropolarimeter, Model 61DS, equipped with a single position thermoelectric cuvette holder. CD spectra were recorded at 20 °C in PBS, with either 5 μM protein (d = 0.2 cm) or 500 nM protein (d = 1.0 cm). Spectra were obtained by averaging three wavelength scans taken in 0.5 nm steps, with a signal averaging time of 2 s and a bandwidth of 1.5 nm.

Analytical ultracentrifugation

Sedimentation equilibrium was performed on a temperature-controlled Beckman XL-A analytical ultracentrifuge equipped with an An60Ti rotor and photoelectric scanner. Data were collected on 10 μM protein samples in PBS from 3,000–17,000 rpm using double-sector cells with charcoal-filled Epon centerpieces and sapphire windows. Scans were performed at the wavelengths indicated in Figure 5 with a step size of 0.001 cm and 30–50 averages. Samples were allowed to equilibrate over 24–36 h and duplicate scans 3 h apart were overlaid to determine if equilibrium had been reached. The partial specific volume of each protein was calculated based on its amino acid compositions using the program Sednterp (Laue et al., 1992). The data were analyzed by a nonlinear least-squares method using the ORIGIN software provided by Beckman. The data were then fitted to two classes of models. First, the data were fitted to a single ideal species model using Equation 1,

$$A_r = \exp[\ln(A_0) + (M(1 - \bar{\nu}\rho)\omega^2/2RT)\cdot(x^2 - x_0^2)] + E \quad (1)$$

where A_r is the absorbance at radius x, A_0 is the absorbance at a reference radius x_0 (the meniscus), M is the molecular weight of the single species, $\bar{\nu}$ is the partial specific volume of the protein, ρ is the density of the solvent, ω is the angular velocity of the rotor, R is the gas constant, T is the absolute temperature, and E is a baseline error correction factor. The data were then fitted to a self-associating system of monomer to n-mer equilibria using Equation 2,

$$A_r = \exp[\ln(A_0) + (M(1 - \bar{\nu}\rho)\omega^2/2RT)\cdot(x^2 - x_0^2)] + \sum(\exp[n\ln(A_0) + \ln(K_{a,n}) + n(M(1 - \bar{\nu}\rho)\omega^2/2RT)\cdot(x^2 - x_0^2)]) + E \quad (2)$$

where $K_{a,n}$ is the association constant for the formation of n-mer from monomers, and M is the monomeric molecular weight. A series of models was tested to examine various possibilities and the best fit was determined based on the randomness and the magnitude of the residuals.

Kinetics

All kinetic measurements were performed at 20 °C in PBS supplemented with 0.1 mg/mL of bovine serum albumin (to stabilize the enzymes at low concentrations). Initial rates were determined by monitoring the disappearance of chorismate spectrophotometrically at 274 nm ($\epsilon_{274\,nm}$ = 2,630 M^{-1} cm^{-1}) using a Cary 3 Bio UV-Visible Spectrophotometer equipped with a thermoelectric cuvette holder. All initial rates were corrected for the corresponding temperature-specific background reaction. Kinetic parameters k_{cat} and K_m were calculated from the initial rates as described (Görisch, 1978), using 20 nM EcCM, 100 nM L4-1, 400 nM L5-1, or 50 nM L7-1. To study the dependence of activity on protein concentration, the variants were assayed with ~60 μM chorismate and k_{cat}/K_m calculated directly from the initial rates ([chorismate] $\ll K_m$).

Acknowledgments

We thank Professor Bruce Ganem for providing the gene for *E. coli* chorismate mutase. We also thank Hilal Lashuel and Professor Jeffery Kelly for analytical ultracentrifugation studies. G.M. is the recipient of a Natural Sciences and Engineering Research Council of Canada 1967 Centennial Postgraduate Scholarship and an Eli Lilly Graduate Student Fellowship. This work was supported by the Skaggs Institute for Chemical Biology at The Scripps Research Institute.

References

Albright RA, Mossing MC, Matthews BW. 1996. High-resolution structure of an engineered cro monomer shows changes in conformation relative to the native dimer. *Biochemistry 35*:735–742.

Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci USA 91*:3127–3131.

Bennett MJ, Schlunegger MP, Eisenberg D. 1995. 3D Domain swapping: A mechanism for oligomer assembly. *Protein Sci 4*:2455–2468.

Brunet AP, Huang ES, Huffine ME, Loeb JE, Weltman RJ, Hecht MH. 1993. The role of turns in the structure of an α-helical protein. *Nature* 364:355–358.

Castagnoli L, Vetriani C, Cesareni G. 1994. Linking an easily detectable phenotype to the folding of a common structural motif. *J Mol Biol 237*:378–387.

Chen Y-H, Yang JT, Chau KH. 1974. Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*:3350–3359.

Cohen C, Parry DAD. 1990. α-Helical coiled coil and bundles: how to design an α-helical protein. *Proteins Struct Funct Genet 7*:1–15.

Creighton TE. 1993. *Proteins: Structures and molecular properties*. New York: W. H. Freeman and Company.

Dickason RR, Huston DP. 1996. Creation of a biologically active interleukin-5 monomer. *Nature 379*:652–655.

Di Donato A, Cafaro V, D'Alessio G. 1994. Ribonuclease A can be transformed into a dimeric ribonuclease with antitumor activity. *J Biol Chem 269*:17394–17396.

Galopin CC, Zhang S, Wilson DB, Ganem B. 1996. On the mechanism of chorismate mutases: Clues from wild-type *E. coli* enzyme and a site-directed mutant related to yeast chorismate mutase. *Tetrahedron Lett 37*:8675–8678.

Goodsell DS, Olson AJ. 1993. Soluble proteins: Size, shape and function. *Trends Biochem Sci 18*:65–68.

Görisch H. 1978. On the mechanism of the chorismate mutase reaction. *Biochemistry 17*:3700–3705.

Green SM, Gittis AG, Meeker AK, Lattman EE. 1995. One-step evolution of a dimer from a monomeric protein. *Nat Struct Biol 2*:746–751.

Greenfield N, Fasman GD. 1969. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry 8*:4108–4116.

Haslam E. 1993. *Shikimic acid metabolism and metabolites*. New York: John Wiley.

Janin J, Miller S, Chothia C. 1988. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol 204*:155–164.

Kast P, Asif-Ullah M, Hilvert D. 1996a. Is chorismate mutase a prototypic entropy trap?—Activation parameters for the *Bacillus subtilis* enzyme. *Tetrahedron Lett 37*:2691–2694.

Kast P, Asif-Ullah M, Jiang N, Hilvert D. 1996b. Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis. *Proc Natl Acad Sci USA 93*:5043–5048.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Ku J, Schultz PG. 1995. Alternate protein frameworks for molecular recognition. *Proc Natl Acad Sci USA 92*:6552–6556.

Laue TM, Shah BD, Ridgeway TM, Pelletier SL. 1992. Computer-aided interpretation of analytical sedimentation data for protein. In: Harding SE, Rowe AJ, Horton JC, eds. *Analytical ultracentrifugation in biochemistry and polymer science*. Cambridge, UK: Royal Society of Chemistry. pp 90–125.

Lee AY, Karplus PA, Ganem B, Clardy J. 1995. Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J Am Chem Soc 117*:3627–3628.

MacBeath G, Kast P. 1998. UGA read-through artifacts: When popular gene expression systems need a pATCH. *BioTechniques 24*:789–794.

MacBeath G, Kast P, Hilvert D. 1998a. Exploring sequence constraints on an interhelical turn using *in vivo* selection for catalytic activity. *Protein Sci 7*:325–335.

MacBeath G, Kast P, Hilvert D. 1998b. Redesigning enzyme topology by directed evolution. *Science 279*:1958–1961.

MacBeath G, Kast P, Hilvert D. 1998c. A small, thermostable and monofunctional chorismate mutase from the archeon *Methanococcus jannaschii*. *Biochemistry 37*:10062–10073.

Miller JH. 1972. *Experiments in molecular genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Mossing MC, Sauer RT. 1990. Stable, monomeric variants of λ cro obtained by insertion of a designed β-hairpin sequence. *Science 250*:1712–1715.

Nagi AD, Regan L. 1997. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des 2*:67–75.

Neet KE, Timm DE. 1994. Conformational stability of dimeric proteins: Quantitative studies by equilibrium denaturation. *Protein Sci 3*:2167–2174.

Parge HE, Arvai AS, Murtari DJ, Reed SI, Tainer JA. 1993. Human CksHs2 atomic structure: A role for its hexameric assembly in cell cycle control. *Science 262*:387–395.

Piccoli R, Tamburrini M, Piccialli G, Di Donato A, Parente A, D'Alessio G. 1992. The dual-mode quaternary structure of seminal RNase. *Proc Natl Acad Sci USA 89*:1870–1874.

Predki PF, Regan L. 1995. Redesigning the topology of a four-helix-bundle protein: Monomeric Rop. *Biochemistry 34*:9834–9839.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Plainview, NY: Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74*:5463–5467.

Stewart J, Wilson DB, Ganem B. 1990. A genetically engineered monofunctional chorismate mutase. *J Am Chem Soc 112*:4582–4584.

Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol 189*:113–130.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60–89.

Trinkl S, Glockshuber R, Jaenicke R. 1994. Dimerization of βB2-crystallin: The role of the linker peptide and the N- and C-terminal extensions. *Protein Sci 3*:1392–1400.

Van Dyke MW, Sirito M, Sawadogo M. 1992. Single-step purification of bacterially expressed polypeptides containing an oligo-histidine domain. *Gene 111*:99–104.

Vlassi M, Steif C, Weber P, Tsernoglou D, Wilson KS, Hinz H-J, Kokkinidis M. 1994. Restored heptad pattern continuity does not alter the folding of a four-α-helix bundle. *Nat Struct Biol 1*:706–716.

Zhang S, Kongsaeree P, Clardy J, Wilson DB, Ganem B. 1996. Site-directed mutagenesis of monofunctional chorismate mutase engineered from the *E. coli* P-protein. *Bioorg Med Chem 4*:1015–1020.