Exploring sequence constraints on an interhelical turn using in vivo selection for catalytic activity*

GAVIN MACBEATH, PETER KAST, AND DONALD HILVERT

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

(RECEIVED July 16, 1997; ACCEPTED August 28, 1997)

Abstract

The role of interhelical turns in determining protein structure has been investigated previously in relatively simple four-helix-bundle proteins using combinatorial mutagenesis coupled with screening for functional variants. To assess the tolerance to sequence substitution of a short, interhelical turn in a larger, more complicated protein, we have exploited a more sensitive in vivo selection for catalytic activity. Randomization of three solvent-exposed turn residues in *Escherichia coli* chorismate mutase (Ala₆₅, His₆₆, and His₆₇), followed by selection, indicated that >63% of tripeptides, including some with significantly altered backbone conformations, can functionally replace the native sequence. The increased sensitivity of the catalytic assay allowed optimal sequences to be distinguished from less appropriate ones, revealing a statistically significant preference for hydrophilic residues in solvent-exposed positions. It also enabled investigation of the extent to which either secondary structure or tertiary interactions influence substitution patterns. Randomization of an α -helical residue (Lys₆₄), together with the adjacent solvent-exposed tripeptide, Ala₆₅-His₆₆-His₆₇, showed that the secondary structure at position 64 does not limit the range of side chains allowed at this site. In contrast, randomization of a buried turn residue (Leu₆₈), together with the same tripeptide, revealed an extremely strict requirement for hydrophobic aliphatic amino acids at this position. The strong constraint imposed by the tertiary interaction, in contrast to the weak influence of secondary structure, has important implications for protein design.

Keywords: chorismate mutase; combinatorial mutagenesis; *Escherichia coli*; four-helix-bundle; genetic selection; interhelical turn; protein engineering; random libraries

In the past few years there has been much debate over the role of turns in determining protein structure. Are specific turn sequences required to direct the folding process, or do turns play a more passive role, simply connecting, by default, contiguous segments of secondary structure? If the former is true, then turn sequences should be relatively intolerant to amino acid substitutions. Conversely, if turns play a passive role, they should be correspondingly more tolerant to mutation.

The problem can be addressed experimentally by replacing the turn in a given protein with random sequences and then evaluating which new turns still enable the protein to fold correctly (Reidhaar-Olson & Sauer, 1988). This approach was employed by Hecht and coworkers to study a three-residue, interhelical turn in the four-helix-bundle protein cytochrome *b*-562 (Brunet et al., 1993). Using a simple color assay to identify correctly folded proteins (cell extracts of cytochrome *b*-562 are colored red), they found that all of the 31 variants they studied passed their folding test. In a similar

experiment, Cesareni and coworkers demonstrated a high tolerance to amino acid substitution in the interhelical turn of ROP (another four-helix-bundle protein), using the ability of the protein to mediate dimerization of the N-terminal domain of the λ repressor as their criterion for determining correctly folded structure (Castagnoli et al., 1994). They concluded that the turn, like the three-residue turn of cytochrome *b*-562, is of "low informational importance." These two studies, along with other mutagenesis experiments (Vlassi et al., 1994; Predki & Regan, 1995; Nagi & Regan, 1997), support the notion that four-helix-bundle domains are relatively tolerant to sequence substitutions in their interhelical connections.

Recently, a more rigorous study by Regan and coworkers showed that, although a solvent-exposed residue in the turn of ROP can be functionally replaced by any of the other 19 amino acids, each substitution stabilizes or destabilizes the protein to a different extent (Predki et al., 1996). Evidently, not all interhelical turns are equivalent. Preference for one turn sequence over another may be accentuated by destabilizing the host protein so that it is less able to compensate for energetically unfavorable sequences. This approach was employed successfully by DeGrado and coworkers to study a β -turn in the B1 domain of protein G (Zhou et al., 1996). As an alternative to this approach, we propose that a more sensitive assay for structural integrity may reveal subtle sequence prefer-

*This paper is dedicated to Professor Dieter Seebach on the occasion of his 60th birthday.

Reprint requests to Donald Hilvert at his present address: Laboratorium für Org. Chemie, ETH Zürich, Universitätsstr. 16, CH-8092 Zürich, Switzerland; e-mail: hilvert@org.chem.ethz.ch.

ences without the need to employ destabilizing mutations. Like the approach of DeGrado and coworkers, this would enable a more accurate evaluation of alternative sequences and facilitate the investigation of how additional structural features affect the tolerance of interhelical turns to mutation.

With this premise, we turned to one of the most exacting probes of structural integrity: the ability to catalyze a chemical transformation. As a model protein for these studies, we chose the chorismate mutase domain of the bifunctional chorismate mutase–prephenate dehydratase enzyme from *Escherichia coli* (*pheA* gene product). The N-terminal domain folds independently from the C-terminal domain (Stewart et al., 1990) to form a homodimer, each subunit of which adopts a four-helix-bundle structure (Lee et al., 1995) (Fig. 1). We have shown that residues 1–96 of the *pheA* gene product are sufficient for full chorismate mutase (CM) activity (G. MacBeath, unpubl. results) and refer to this domain as EcCM (for *E. coli* chorismate mutase).

Unlike many four-helix-bundle structures, EcCM is an enzyme; it catalyzes the conversion of chorismate to prephenate with a rate acceleration of $>10^6$ -fold 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 (L-Tyr) and L-phenylalanine (L-Phe) in bacteria, fungi, and higher plants (Haslam, 1993). We have previously prepared a strain of *E. coli* that lacks CM activity, but has all the other enzymic func-

tions necessary for the biosynthesis of these two amino acids (Kast et al., 1996b). As a result, the strain is unable to grow on media lacking L-Tyr and L-Phe, but regains prototrophy when transformed with a plasmid expressing a functional CM enzyme (Kast et al., 1996b). This complementation system provides a very sensitive and accurate test for the structural integrity of protein variants. If a mutation in EcCM renders the protein unable to fold correctly, cells expressing this variant will not grow on minimal media without L-Tyr and L-Phe. However, if a mutation still permits correct folding, cells expressing this variant should grow unhindered. Because enzyme function, unlike ligand binding (cytochrome *b*-562) or dimerization (ROP), may be extremely sensitive to structural perturbations, we are able to detect relatively small changes that might go unnoticed when less exacting assays are employed. Moreover, because the cells require functional protein in order to grow, we can screen large numbers of variants very rapidly by genetic selection, simply plating the cells in duplicate on minimal media, either supplemented with or lacking L-Tyr and L-Phe.

Using this system, we can extend the studies on interhelical turns by investigating sequence preferences in a number of contexts. Previous studies have focused on turns comprised of either solvent-exposed residues or residues not involved in extensive interactions in the protein. In this way, the investigators sought to assess the intrinsic propensities of the amino acids to form the

Fig. 1. Ribbon structure of EcCM, the chorismate mutase domain of *Escherichia coli* chorismate mutase–prephenate dehydratase (Protein Data Bank entry 1ecm). **A:** View of the homodimeric enzyme with transition-state analogue inhibitor bound in the active sites. **B:** Detailed view of turn L2, showing the five residues under investigation. The figure was generated with the program MOLSCRIPT (Kraulis, 1991).

corresponding turns, independent of the surrounding protein. In fact, in the study by Regan and coworkers, the two residues flanking the position of interest were changed to alanine in order to simplify further the local environment (Predki et al., 1996). In this study, we not only build on this work using our more sensitive assay, but also investigate what role, if any, the protein context plays in restricting the sequence of a turn. Specifically, we determine to what extent either secondary structural requirements or long-range tertiary interactions render interhelical turns less tolerant to sequence substitution.

Results and discussion

Engineering of the chorismate mutase complementation system

The *E. coli*-based complementation system for CM activity has been described previously (Kast et al., 1996b). The system is comprised of three elements: (1) the RecA[−] *E. coli* host strain KA12, which lacks both endogenous CM genes (*tyrA* and *pheA*); (2) the auxiliary plasmid pKIMP-UAUC, which restores the deleted prephenate dehydrogenase and prephenate dehydratase activities also encoded by the bifunctional *tyrA* and *pheA* genes, respectively; and (3) a second, compatible plasmid for expression of the protein of interest.

In order to study variants of EcCM, we constructed plasmid pKECMT-W (Fig. 2A), in which *pheA'*, a 3'-truncated version of *pheA* encoding the monofunctional CM domain, was constitutively expressed by the relatively strong *trc* promoter (see the Materials and methods). However, when KA12/pKIMP-UAUC cells were transformed with this plasmid, cell growth was severely impaired on minimal media, even when supplemented with L-Phe and L-Tyr. Evidently, high levels of EcCM are toxic to the cells under these conditions. To reduce the level of protein production, the *trc* promoter was replaced by the less efficient *bla* promoter (Deuschle

et al., 1986). In addition, an *Mlu* I site was introduced into *pheA'* to facilitate DNA manipulation, yielding pKECMB-Wm. Upon transformation of KA12/pKIMP-UAUC cells with this plasmid, no toxic effects were noted. Moreover, pKECMB-Wm provided efficient complementation of the L-Tyr/L-Phe auxotrophy. All further experiments were conducted using this construct.

Design of the random turn libraries

EcCM is a homodimer composed of two identical four-helix-bundle domains (Lee et al., 1995) (Fig. 1). Each polypeptide folds into three helices, with the N-terminal helix (H1) spanning both domains. H1 is connected to H2 by a long, irregular loop (L1) that comprises part of the active site of the enzyme. H2, on the other hand, is connected to H3 by a relatively short stretch of amino acids (L2) that is distant from the active site. Five residues connect H2 to H3: Ala₆₅ assumes a 3_{10} -helical conformation at the C-terminus of H2 (ϕ/ψ angle of $-63.8^\circ/-32.2^\circ$ with a 3_{10} -hydrogen bond from Ala₆₅ NH to Leu₆₂ O), whereas residues His₆₆, His₆₇, Leu₆₈, and Asp₆₉ adopt irregular conformations (ϕ/ψ angles of $-89.0^\circ/-5.1^\circ$, $71.1^\circ/31.5^\circ$, $-102.9^\circ/128.4^\circ$, and $-80.4^\circ/154.3^\circ$, respectively).

Of the five residues in L2, all but Leu₆₈ are solvent exposed. Conveniently, the first three residues (Ala₆₅, His₆₆, and His₆₇) are not involved in any tertiary interactions with the rest of the protein. This enabled us to evaluate the generality of Hecht's conclusions (Brunet et al., 1993) using a different four-helix-bundle protein (albeit one more complex than cytochrome *b*-562). Assuming their result is general (i.e., >91% of all possible tripeptides are tolerated in interhelical turns of four-helix-bundle proteins), we should observe a similar result (i.e., >91% of EcCM variants should complement the L-Tyr/L-Phe auxotrophy). However, if their assay for correctly folded protein (colored cell extract) was too insensitive to detect subtle structural perturbations in cytochrome *b*-562, our more stringent requirement of catalytic activity may reveal a lower percentage of tolerated sequences in EcCM and/or a subset of sequences that are only partially tolerated (i.e., result in reduced levels of complementation). Hence, to establish a benchmark for our subsequent experiments on the effects of protein context, we prepared a library in which only the tripeptide Ala₆₅-His₆₆-His₆₇ was completely randomized. This was designated library [AHH].

Immediately preceding this tripeptide sequence in EcCM is Lys₆₄. Like the following three residues, Lys₆₄ is completely solvent exposed. However, it is not part of turn L2, but rather adopts an α -helical conformation at the C-terminal end of H2. To investigate the extent to which the identity of a residue involved in secondary structure affects tolerance to sequence substitution in the adjacent turn, we prepared library [KAHH] in which Lys₆₄ was randomized, along with Ala₆₅, His₆₆, and His₆₇.

Finally, to explore the effects of tertiary interactions on the sequence dependence of interhelical turns, we turned our attention to Leu₆₈. In contrast to the residues discussed above, the side chain of Leu₆₈ is sequestered from solvent, buried in the interface between L2 and H1' (helix H1 of the other subunit). It is not tightly packed in the interface (no side-chain atom of Leu₆₈ is within 3.7 Å of any other side-chain atom in the protein), but is nevertheless enclosed in a hydrophobic pocket, formed by atoms from Leu_{22'}, Leu_{25'}, Ala_{26'}, Arg_{29'}, His₆₆ (C β atom), Tyr₇₂, and Ile₇₃. This hydrophobic cluster appears to mediate the association of L2 with H1'. Is this association essential to the structural integrity of L2 and to the folding and catalytic proficiency of the enzyme,

Fig. 2. Construction of the random turn libraries. **A:** A section of pKECMT-W, including part of the *pheA'* gene, was amplified by PCR. In each case, the 5' primer (5' P) was PTRC-S, whereas the 3' primer (3' P) was either P_{AHH}, P_{KAHH}, or P_{AHHL} (for libraries [AHH], [KAHH], or [AHHL], respectively). The 507-bp PCR product was digested subsequently with *Nhe* I and *Mlu* I to yield a 130-bp fragment. **B:** pKECMB-S, the recipient plasmid for the libraries, contained an 832-bp piece of stuffer DNA, which was removed by digestion with *Nhe* I and *Mlu* I (NB: drawing not to scale). **C:** Ligation of the 130-bp *Nhe* I-*Mlu* I fragment of the PCR product with the 2,782-bp *Nhe* I-*Mlu* I fragment of pKECMB-S yielded each library in pKECMB-Wm.

despite its distance (~ 18 Å) from the active site? If Leu₆₈ plays an important role in the interhelical turn, it should be relatively intolerant to substitution by dissimilar residues. Alternatively, if the association is not critical, then other turn sequences may be found that do not retain this tertiary interaction. To investigate the effect of this interaction on sequence restriction in the loop, we prepared library [AHHL], in which the tetrapeptide Ala₆₅-His₆₆-His₆₇-Leu₆₈ was completely randomized.

Construction of the libraries

Before constructing the three different libraries, we prepared a functionally inactive derivative of pKECMB-Wm by replacing the 130-bp *Nhe* I-*Mlu* I fragment of pKECMB-Wm (which encompasses L2) with an 832-bp piece of stuffer DNA, yielding pKECMB-S (Fig. 2B). As expected, this plasmid offers no detectable complementation of the CM deficiency. The strategy for preparing the libraries then involved amplifying a DNA region of pKECMT-W by PCR, using mutagenic primers to randomize the codons for the appropriate residues in L2 (Fig. 2A). Codons were mutated using the scheme NNS, where N is an equimolar mixture of A, C, G, and T, and S is an equimolar mixture of G and C. This yields 32 possible codons, including codons for all 20 amino acids, and only a single stop codon (TAG). The primers also included a mismatched base (silent mutation) following the mutagenic region to reduce any bias toward wild-type sequence that may occur in the primer-template hybridization step of the PCR.

Following amplification and restriction enzyme digestion, the 130-bp *Nhe* I-*Mlu* I fragment was ligated with the appropriate fragment of pKECMB-S (Fig. 2B), simultaneously restoring the *pheA'* gene and introducing the mutagenized turn sequence (Fig. 2C). Because the host vector (pKECMB-S) is functionally inactive, any background interference from this vector will not yield "false positives." The ligated DNA was introduced first into XL1-Blue cells (Stratagene), yielding $>10^8$ transformants for each library. Following amplification of the libraries, plasmid DNA was recovered and subsequently introduced into the auxotrophic strain KA12/pKIMP-UAUC, again yielding $>10^8$ transformants for each library. The two-step process eliminates the possibility of transforming the auxotrophic cells with heteroduplex DNA, which may be present following the PCR amplification step.

Because library [AHH] has a potential diversity of $32^3 = 3.28 \times 10^4$ and libraries [KAHH] and [AHHL] each have a potential diversity of $32^4 = 1.05 \times 10^6$, the >100 -fold excess of transformants in each library ensures that most of the possible sequence variants are represented. To verify the strategy and evaluate the quality of the primers, 24 clones were randomly chosen from library [AHHL] and sequenced. Of the 192 bases that were derived from equimolar mixtures of A, C, G, and T, 48 were A, 81 were C, 34 were G, and 29 were T (48 expected for each base). Of the 96 bases that were derived from equimolar mixtures of C and G, 65 were C and 31 were G (48 expected for each base). Although there appears to be an overabundance of C's, it is clear that all four bases are present in approximately the intended proportions.

Evaluation by genetic selection

Following amplification of the libraries under nonselective conditions, the cells were washed thoroughly, diluted, and plated in triplicate on M9c agar plates, either supplemented with or lacking L-Tyr and L-Phe (see the Materials and methods). The plates were

incubated at 30 °C and colonies were counted after 3 days and then again after 7 days (Table 1). It was immediately apparent that colony size varied dramatically on the selective plates, ranging from clones with wild-type-like growth rates to clones that produced only pin-prick-sized colonies after 7 days (Fig. 3). To distinguish the different levels of complementation, clones yielding large colonies (ranging from wild-type size to half that diameter) were designated "fully active" whereas those forming smaller colonies were designated "partially active." The percent of viable clones in each library that were fully active is also given in Table 1.

Based on this simple but stringent test for functional protein, we find that 58% of library [AHH] is composed of viable clones. Because the TAG stop codon occurs at a frequency of 1 in 32, we would, at most, expect only $(31/32)^3 = 91\%$ of the library to be composed of clones encoding full-length EcCM sequences. In practice, this percentage may be slightly smaller due to the presence of cloning artifacts in the libraries. Although no alternative ligation products were encountered in the screening and sequencing of clones (see below), some mutants bearing single-base deletions (leading to frame-shift mutations) were observed. Regardless of this adjustment, it is clear that the majority ($>63\%$) of tripeptide sequences can functionally substitute for Ala₆₅-His₆₆-His₆₇ in EcCM, broadly consistent with the results of Hecht and coworkers using cytochrome *b*-562 (Brunet et al., 1993). However, it is also clear that not all sequences are equivalent. The range in colony size among viable clones (Fig. 3) strongly suggests that some tripeptides are preferable to others. Although different sequences may simply result in different levels of protein production, it is more likely that structural changes induced by the alternative turns affect protein stability, susceptibility to proteases, or catalytic efficiency.

When the experiment is extended to include the α -helical residue Lys₆₄ (library [KAHH]), the fraction of viable clones drops only slightly, to 50%. When adjusted for the presence of stop codons, both library [AHH] and library [KAHH] exhibit a tolerance to sequence substitution of about 86% per residue. Evidently, the secondary structural involvement of position 64 does not impose a significant constraint on sequence requirements. Lys₆₄, like the three solvent-exposed turn residues that follow, is of relatively low informational importance. In marked contrast, when Leu₆₈ is included in the randomization (library [AHHL]), the percentage of viable clones drops dramatically, to 6.3% (Table 1; Fig. 3). This is almost a 10-fold drop from library [AHH], attributable solely to the inclusion of Leu₆₈. Correcting for the presence of stop codons,

Table 1. Evaluation of random turn libraries by genetic complementation

Library	Supplemented ^a	Selective ^b	% Viable ^c	% Fully active ^d
[AHH]	2,376	1,388	58	55
[KAHH]	2,645	1,325	50	50
[AHHL]	2,225	1,403 ^e	6.3	37

^aNumber of colonies counted on M9c media supplemented with L-Tyr and L-Phe.

^bNumber of colonies counted on M9c media lacking L-Tyr and L-Phe.

^cPercentage of clones able to grow on selective media.

^dPercentage of viable clones with colonies over half the diameter of wild-type colonies grown on selective media.

^eThese clones were plated at a 10-fold higher concentration than those on the corresponding supplemented plates.

Fig. 3. Screening for functional variants of EcCM by genetic selection. Top row of plates shows colonies from library [AHH]; middle row, from library [KAHH]; and bottom row, from library [AHLH]. Plates in the left column are supplemented with L-Tyr and L-Phe and were photographed after 3 days at 30 °C. Plates in the middle column lack L-Tyr and L-Phe and were also photographed after 3 days at 30 °C. The right column shows the same plates as the middle column, photographed after 7 days at 30 °C to illustrate the range in colony size. For each library, the same volume of cell suspension was spread on selective and nonselective plates. NB: In the right column, colonies on the [AHH] and [KAHH] plates are smaller than on the [AHLH] plate due to inhibition of growth resulting from overcrowding.

position 68 has a tolerance to substitution of only 11%. Because 3 of the 32 possible codons specify Leu, it is clear that there are very few alternatives for this position in the turn.

Permissive and nonpermissive turn sequences

For sequence analysis, clones were divided into three categories based on their ability to complement the L-Phe/L-Tyr auxotrophy: fully active, partially active, and inactive (unable to grow on selective media). Sequences were obtained for each class of clones as follows. For the fully active and partially active clones, well-isolated colonies were picked from the initial plating experiment and streaked on fresh selective plates to verify their phenotype. They were then recovered from the fresh plates and their relevant DNA sequences were determined. For the inactive clones, random colonies from supplemented plates were streaked in parallel on both supplemented and selective plates. Those clones that showed no growth on the selective plates after 7 days were recovered from the supplemented plates and subjected to sequence analysis.

Table 2 shows the sequences of clones that were deemed fully active based on their fast growth rates on selective media. Al-

though many different amino acid residues are found in the solvent-exposed positions (64–67) in all three libraries, it is clear that the range of residues permitted at position 68 is extremely limited. Among the 10 fully active clones from the [AHLH] library, 6 have retained the Leu at position 68. Interestingly, all three available codons for Leu are observed among the clones, indicating a strong selection for phenotype, but lack of bias in the genotype. The only other residues found at position 68 are Val, Ile, and Met, all similar in size and hydrophobicity to Leu. The low tolerance of Leu₆₈ to substitution highlights its importance in the interhelical turn. Apparently, the docking of L2 onto H1', mediated by Leu₆₈, is an essential structural feature of the folded, active enzyme.

The low informational importance of Lys₆₄ (indicated by the complementation data above) is further illustrated by the sequencing results. Although only six fully active clones were sequenced from library [KAHH], we found (besides Glu, Arg, Gln, and Lys) both Gly and Thr at position 64. With the exception of Pro, Gly is generally considered to have the lowest α -helical propensity of the commonly occurring amino acids, with Thr following closely behind (O'Neil & DeGrado, 1990; Horovitz et al., 1992; Blaber et al., 1993; Chakrabarty et al., 1994; Muñoz & Serrano, 1994). Evi-

Table 2. Sequences of fully active clones from each of the libraries

Wild type	DNA sequences					Protein sequences				
	AAA	GCG	CAC	CAT	CTG	Lys ₆₄	Ala ₆₅	His ₆₆	His ₆₇	Leu ₆₈
[AHH]										
AHH-fa-1		GGG	CAG	CAC			Gly	Gln	His	
AHH-fa-2		AAG	CGG	AAC			Lys	Arg	Asn	
AHH-fa-3		CCC	TAC	GAG			Pro	Tyr	Glu	
AHH-fa-4		ACC	AGG	CAG			Thr	Arg	Asn	
AHH-fa-5		CAC	CGG	AGC			His	Arg	Ser	
AHH-fa-6		AGG	GCG	CAG			Arg	Ala	Gln	
[KAHH]										
KAHH-fa-1	GAG	GAG	AAG	CAC		Glu	Glu	Lys	His	
KAHH-fa-2	AGG	CAG	CGG	AAG		Arg	Gln	Arg	Lys	
KAHH-fa-3	GGG	CCC	AAG	AAG		Gly	Pro	Lys	Lys	
KAHH-fa-4	CAG	GAC	GCG	CAC		Gln	Asp	Ala	His	
KAHH-fa-5	ACC	AGG	CTC	GGG		Thr	Arg	Leu	Gly	
KAHH-fa-6	AAG	TGC	TTG	CAG		Lys	Cys	Leu	Gln	
[AHHL]										
AHHL-fa-1		GCC	AAG	AAC	GTG		Ala	Lys	Asn	Val
AHHL-fa-2		CAG	CAC	AAG	ATG		Gln	His	Lys	Met
AHHL-fa-3		CCC	CGC	GGG	ATG		Pro	Arg	Gly	Met
AHHL-fa-4		GCC	TGG	CAC	CTG		Ala	Trp	His	Leu
AHHL-fa-5		CAC	GGC	GGC	CTC		His	Gly	Gly	Leu
AHHL-fa-6		ACC	CAG	GAC	CTC		Thr	Gln	Asp	Leu
AHHL-fa-7		CCC	CAG	AAG	CTG		Pro	Gln	Lys	Leu
AHHL-fa-8		ACC	CAG	CAC	TTG		Thr	Gln	His	Leu
AHHL-fa-9		GCC	TAC	AAC	ATC		Ala	Tyr	Asn	Ile
AHHL-fa-10		GCC	AGG	CAG	TTG		Ala	Arg	Gln	Leu

dently, the protein is able to compensate for these disfavored substitutions without a noticeable loss in activity. This result is consistent with the observation that helix propensities may be overwhelmed by more important factors, such as the appropriate patterning of polar and nonpolar amino acids (Xiong et al., 1995).

A closer analysis of the residues found in the solvent-exposed positions reveals a more subtle bias: there is a statistically relevant overrepresentation of hydrophilic residues in positions 64–67. We can divide the 20 amino acids into two equal groups: 10 hydrophobic residues (Trp, Phe, Ile, Leu, Cys, Met, Val, Tyr, Pro, and Ala) and 10 hydrophilic (Thr, His, Gly, Ser, Gln, Asn, Glu, Asp, Lys, and Arg). With this classification, the number of codons of the form NNS is almost evenly split as well: 15 encode hydrophobic amino acids, 16 encode hydrophilic, and 1 specifies chain termination. In addition, C appears the same number of times (16) among the codons in each group. Thus, the previously noted overrepresentation of C's in the library oligonucleotides does not, in itself, result in a bias toward hydrophobic or hydrophilic residues. However, of the 72 codons in Table 2 that code for residues in solvent-exposed positions, 56 encode hydrophilic residues (78%). This is 4.7 standard deviations higher than the expected value of 36, indicating a very strong statistical bias. We can show that this preference arises from the genetic selection of fully active clones by comparing the results to those obtained from 24 randomly chosen, *unselected* clones (see above). In that case, of the 96 randomized codons, 51 encode hydrophobic residues, 42 encode hydrophilic residues, and 3 specify stops (expected values are 45, 48, and 3, respectively). These data are within 1.2 standard deviations of their expected values, indicating no significant bias.

This striking preference for hydrophilic residues was not observed previously by Hecht and coworkers using cytochrome *b*-562 (Brunet et al., 1993). (They observed 45 hydrophilic residues and 51 hydrophobic residues using the scheme NNN, where 32 codons specify hydrophilic residues and 29 specify hydrophobic ones.) One explanation for the discrepancy is that their assay for correctly folded protein was too insensitive to detect subtle structural perturbations resulting from the placement of hydrophobic residues in solvent-exposed positions. By employing a more sensitive assay, we were able to uncover more subtle sequence preferences. Alternatively, the inclusion of too many hydrophobic residues in a turn may favor intermolecular aggregation. Small protein aggregates are probably inconsequential in an assay based on color formation, but may prevent catalytic function or may be poorly tolerated in the cytoplasm of a cell.

The preference that we observe is intriguing in light of a wealth of evidence that surface-exposed residues are generally highly tolerant to mutation, including mutation to hydrophobic amino acids (Perutz et al., 1965; Miller et al., 1979; Lesk & Chothia, 1980; Reidhaar-Olson & Sauer, 1988; Shortle, 1989; Bowie et al., 1990). From a theoretical standpoint, surface-exposed hydrophobic residues should not be particularly destabilizing per se, because their side chains are exposed to water in both the folded and unfolded states. However, surface-exposed hydrophobic residues become deleterious when they promote alternative, misfolded conformations of the protein or lead to aggregation (Bowie et al., 1990). In support of this notion, an analysis of 226 globin sequences by Bashford et al. (1987) revealed a statistical bias against the occurrence of too many large, hydrophobic surface residues. It is clear

from the present study that EcCM similarly disfavors a preponderance of hydrophobic residues in the solvent-exposed positions of L2; every fully active clone that was sequenced contained at least one hydrophilic residue. Presumably, when too many hydrophobic residues are introduced, other conformations and/or aggregation states predominate.

The bias toward hydrophilic amino acids in positions 64–67 diminishes when we examine the partially active clones (Table 3). Excluding residues from AHH-pa-1 and AHH-pa-3 (which contain suppressible stop codons; see below), 61% of solvent-exposed residues are hydrophilic. In addition, clone AHH-pa-2 has a tripeptide sequence composed entirely of hydrophobic residues (Met₆₅-Tyr₆₆-Met₆₇). We are reluctant to overinterpret this result, however, because KAHH-pa-3 exhibits the same partially active phenotype and yet contains four hydrophilic residues (Thr₆₄-Asn₆₅-Thr₆₆-Asp₆₇). Other factors must account for the reduced activity of clones such as this. Interestingly, a nonstructural mechanism accounts for the reduced activity of two clones in the [AHH] library. Both AHH-pa-1 and AHH-pa-3 contain a TAG stop codon. This codon is suppressed by Gln in KA12/pKIMP-UAUC cells (*supE44*) (Bachmann, 1972; Eggertsson & Söll, 1988). However, the efficiency of suppression is low (~1–10%) when the codon is followed by a pyrimidine (Miller & Albertini, 1983). Thus, although the tripeptide His₆₅-Arg₆₆-Gln₆₇ (clone AHH-pa-1) may be tolerated well by the protein, the decreased concentration of full-length enzyme in the cell can lead to a reduced growth rate under selective conditions.

The sequence of partially active clone AHHL-pa-3 (Table 3) provides further insight into the tolerance of this interhelical turn to substitution. Because a Leu is found at position 68, this clone differs from the wild-type protein by the replacement of Ala₆₅-His₆₆-His₆₇ with Thr₆₅-Thr₆₆-Pro₆₇. In the wild-type protein, His₆₇ adopts a ϕ angle of +71.1°. Because the ϕ angles of Pro residues are strongly constrained to -60° (Creighton, 1993), this tripeptide must adopt a dramatically different backbone conformation to accommodate the Pro. Among the fully active clones of Table 2, Pro

was only found in position 65 (Ala₆₅ adopts a ϕ angle of -63.8° in the wild-type protein). The observation of Pro at position 67 in a partially active clone indicates that not only are alternative side chains allowed in the turn, but even alternative backbone conformations are tolerated (although perhaps at greater cost).

Examining the partially active clones from library [AHHL], we also see a strong selection for Leu at position 68 (four of the six clones). Although Ile in position 68 (AHHL-pa-6) was observed previously among the fully active clones, the occurrence of Phe in AHHL-pa-1 is novel. We can attribute the reduced activity to the presence of this Phe: although the clone also has a Pro at position 65 and a Lys at position 67, we know that these two substitutions are well-tolerated given the full activity of AHHL-fa-7 (Table 2). Although hydrophobic, Phe is considerably larger than Leu and so should not be tolerated as well as more isosteric residues, such as Ile, Val, and Met. Nevertheless, it is surprising that even the fairly conservative replacement of Leu for Phe has a significant effect on the enzyme's ability to provide efficient complementation. Clearly, the participation of Leu₆₈ in various tertiary interactions greatly increases its informational content (Reidhaar-Olson & Sauer, 1988).

This point is further emphasized when we consider the sequences of the inactive clones (Table 4). Although frame shifts and stops account for most of the inactive clones in libraries [AHH] and [KAHH], 9 of the 10 clones in library [AHHL] contain legitimate tetrapeptide sequences. Unlike the fully or partially active clones from this library, however, none of these have a large, hydrophobic residue at position 68. Interestingly, even Ala and Cys are unacceptable. Their side chains are apparently too small to mediate the interaction between L2 and H1' that appears to be so critical for the structural integrity of EcCM.

As mentioned above, most of the inactive clones in libraries [AHH] and [KAHH] contain either stop codons or single-base deletions (frame-shift mutations). Nevertheless, both libraries also contain clones with legitimate sequences that fail to produce functional protein. For example, AHH-ia-4 has the sequence Trp₆₅-

Table 3. Sequences of partially active clones from each of the libraries^a

Wild type	DNA sequences					Protein sequences				
	AAA	GCG	CAC	CAT	CTG	Lys ₆₄	Ala ₆₅	His ₆₆	His ₆₇	Leu ₆₈
[AHH]										
AHH-pa-1		CAC	AGG	TAG			His	Arg	*	
AHH-pa-2		ATG	TAC	ATG			Met	Tyr	Met	
AHH-pa-3		TAG	TTC	AAC			*	Phe	Asn	
[KAHH]										
KAHH-pa-1	TTG	ACG	GAC	AAC		Leu	Thr	Asp	Asn	
KAHH-pa-2	CCG	TCC	CTC	GAG		Pro	Ser	Leu	Glu	
KAHH-pa-3	ACG	AAC	ACG	GAC		Thr	Asn	Thr	Asp	
[AHHL]										
AHHL-pa-1		CCC	CAC	AAG	TTC		Pro	His	Lys	Phe
AHHL-pa-2		CGC	GCG	ACG	TTG		Arg	Ala	Thr	Leu
AHHL-pa-3		ACC	ACC	CCC	CTC		Thr	Thr	Pro	Leu
AHHL-pa-4		GTC	ACC	ACC	CTG		Val	Thr	Thr	Leu
AHHL-pa-5		ATC	GAG	GAG	TTG		Ile	Glu	Glu	Leu
AHHL-pa-6		CCC	GCC	CGG	ATC		Pro	Ala	Arg	Ile

^aAsterisk specifies a stop codon (suppressible in KA12/pKIMP-UAUC by Gln).

Table 4. Sequences of inactive clones from each of the libraries^a

Wild type	DNA sequences					Protein sequences				
	AAA	GCG	CAC	CAT	CTG	Lys ₆₄	Ala ₆₅	His ₆₆	His ₆₇	Leu ₆₈
[AHH]										
AHH-ia-1		CGC	CCG	AAC			Frame shift ^b			
AHH-ia-2		TAG	GCC	TCG			Frame shift ^b			
AHH-ia-3		ACG	TGC	AG_			Frame shift			
AHH-ia-4		TGG	CAG	CCC			Trp	Gln	Pro	
[KAHH]										
KAHH-ia-1	CCC	ACG	CCC	AAG		Pro	Thr	Pro	Lys	
KAHH-ia-2	AAC	TAG	TAG	AAG		Asn	*	*	Lys	
KAHH-ia-3	ACG	CGC	TTC	CG_		Frame shift				
KAHH-ia-4 ^c	ATG	AGC	ACG	TTG		Met	Ser	Thr	Leu	
[AHLH]										
AHLH-ia-1	ACC	CCC	AAG	GCG			Thr	Pro	Lys	Ala
AHLH-ia-2	GAC	GAC	ACC	TGC			Asp	Asp	Thr	Cys
AHLH-ia-3	CAC	AAG	CCG	GGG			His	Lys	Pro	Gly
AHLH-ia-4	GAC	GTC	ACC	AAG			Asp	Val	Thr	Lys
AHLH-ia-5	CCC	GAG	CAC	GCC			Pro	Glu	His	Ala
AHLH-ia-6	CTC	GAC	CCC	CCG			Leu	Asp	Pro	Pro
AHLH-ia-7	GCC	AAC	GCG	GCG			Ala	Asn	Ala	Ala
AHLH-ia-8	ACC	CGC	CCG	CAC			Thr	Arg	Pro	His
AHLH-ia-9	CCC	GCC	CGG	TAG			Pro	Ala	Arg	*
AHLH-ia-10	CAC	GCC	GCC	AAC			His	Ala	Ala	Asn

^aAsterisk specifies a stop codon (suppressible in KA12/pKIMP-UAUC by Gln); _ specifies a single-base deletion (results in a frame-shift mutation).

^bFrame shifts in these clones result from single-base deletions in the library oligonucleotides outside the region of the randomized codons (sequences not shown).

^cThis clone also had a point mutation (resulting in a Gly₆₃-to-Glu exchange).

Gln₆₆-Pro₆₇ and yet shows absolutely no growth on selective media. Although Hecht and coworkers found no such sequences among 45 variants of cytochrome *b*-562 (Brunet et al., 1993), an extensive search by Cesareni and coworkers identified three tripeptide sequences that were unable to substitute for the interhelical turn in ROP (Castagnoli et al., 1994). Our results confirm that, although solvent-exposed positions in interhelical turns are highly tolerant to sequence substitution, not all turns are acceptable.

Implications for design

Two important conclusions can be drawn from this study. First, although some hydrophobic residues are tolerated in solvent-exposed positions of the turn, there is an overall bias toward hydrophilic residues at these locations. Presumably, a preponderance of exposed, hydrophobic side chains can promote either intermolecular aggregation or alternative conformations, both of which prevent optimal activity of the enzyme. The high sensitivity of our assay enabled us to observe this preference, which would have been missed had we used a less stringent criterion to evaluate structural integrity. The preference for hydrophilic residues in the exposed positions, together with an absolute requirement for hydrophobic residues at position 68, emphasizes the importance of preserving the binary patterning of polar and nonpolar residues along the turn. The principle of binary patterning, although clearly an oversimplification, has been applied successfully to the design

of novel proteins over the past several years (Kamtekar et al., 1993; Xiong et al., 1995; West & Hecht, 1995; Beasley & Hecht, 1997).

The second important conclusion concerns the role of the protein context in determining which turn sequences are preferred. Hecht and coworkers have already shown that sequence substitution is much less tolerated in turns connecting two β -strands than in turns connecting α -helices; only 6% of tetrapeptide sequences can functionally substitute for the central β -turn of plastocyanin (Ybe & Hecht, 1996). This indicates that different secondary structural contexts can place different requirements on the intervening turn. Our results now show that long-range tertiary interactions between residues in an interhelical turn and residues elsewhere in the protein can have an even greater constraining effect. Leu₆₈, which serves to dock L2 onto an adjacent helix, is almost completely intolerant to substitution, even when alternative turn sequences are explored. In terms of protein design, this means that serious attention must be paid to the choice of turn sequence when interaction with other elements in the protein is possible. In EcCM, the interaction arises from the quaternary organization of the protein. Thus, as more complex structures are explored, long-range tertiary interactions mediated by turn residues must be considered carefully. In this study, we used the power of genetic selection to evaluate turn sequences in a sensitive and efficient manner. We are currently exploring how to apply these same methods to the process of design.

Materials and methods

Strains and plasmids

The construction of *E. coli* strain KA12/pKIMP-UAUC has been described previously (Kast et al., 1996a, 1996b). The genotype of KA12 is $\Delta(\text{srlR-recA})306::\text{Tn10}$, $\Delta(\text{pheA-tyrA-aroF})$, thi-1 , endA-1 , hsdR17 , $\Delta(\text{argF-lac})\text{U169}$, supE44 . The plasmid pKIMP-UAUC confers chloramphenicol resistance and provides constitutive expression of genes encoding monofunctional forms of prephenate dehydrogenase and prephenate dehydratase from *Erwinia herbicola* and *Pseudomonas aeruginosa*, respectively. General cloning and library construction was performed in *E. coli* strain XL1-Blue (Stratagene). Plasmid pJS47 (Zhang et al., 1996), which harbors the chorismate mutase domain encoded by the 3'-truncated *pheA* gene (*pheA'*) of *E. coli* (Stewart et al., 1990), was obtained from Professor B. Ganem (Cornell University, Ithaca, New York).

DNA manipulations

All nucleic acid manipulations were done according to standard procedures (Sambrook et al., 1989). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs or Boehringer Mannheim. Oligonucleotides were obtained by custom synthesis from Life Technologies (GIBCO BRL) or Cruachem (FisherBiotech). Oligonucleotides employed in library construction were purified by PAGE. PCRs were performed using the Stoffel fragment of *AmpliTaq* DNA polymerase (Perkin-Elmer). All relevant portions of the constructed plasmids were confirmed by DNA sequencing 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.

Construction of plasmids

Plasmid pKECMT-W was constructed as follows. First, three overlapping PCR products were generated separately, incorporating various restriction endonuclease sites in the process (underlined in the primer sequences). PCR1T (287 bp) was obtained with primers PTRC-S (CGTTCTACATGTTTGACAGCTTATCATCGACT) and PTRC-N (TCCGATGTATATGCTGTTTCTCTGTGTGAAATTGTTAT) using pK1F7T-0 as the template DNA. Plasmid pK1F7T-0 (P. Kast, unpubl.) is identical in all relevant portions to the previously reported plasmid pKCMT-W (Kast et al., 1996b). PCR2 (80 bp) was obtained with primers ECM1-S (GGAAACAGCATATGACATCGGAAAACCCGTTA) and ECM1-N (CGCTAAAAGCTTTTCATCCAGCGCGCTGATT) using pJS47 as the template. PCR3 (269 bp) was obtained with primers ECM2-S (GATGAAAAGCTTTTAGCGTTACTGGCAGAACGGCGGAGCTAGCCGTCGAGGTG) and ECM2-N (GGACTAGTTCAGTGTTGGTGGTGGTGGTGTGCTCGAGATGTTGTTGGAGCAAA) using pJS47 as the template. Following gel purification, the three overlapping PCR products were combined in equimolar ratios (220 fmol each) and assembled into one 596-bp product (PCR4) in a single 50- μL PCR. The conditions were as follows: hot start (94 °C, 4 min) followed by 25 cycles of denaturation (94 °C, 1 min), annealing (42 °C, 2 min), and extension (72 °C, 30 s with a 7-s increase per cycle), followed by a final extension period (72 °C, 4 min). The outside primers PTRC-S and ECM2-N (25 pmol each) were added after the first 5 cycles. The 582-bp *Nsp* I-*Spe* I fragment of PCR4

was ligated into the 2,507-bp *Nsp* I-*Spe* I fragment of pK1F7T-0 (identical to that of pKCMT-W) to yield pKECMT-W (3,089 bp) (Fig. 2A).

Following transformation of pKECMT-W into KA12/pKIMP-UAUC cells, it was found that cell growth was severely impaired on minimal media (M9c, see below), even when supplemented with L-Phe and L-Tyr. Presumably, the relatively high expression of *pheA'* under control of the *trc* promoter is toxic to the cells under these conditions. Consequently, the *trc* promoter in pKECMT-W was replaced by the much less efficient *bla* promoter, which usually drives expression of the β -lactamase gene in plasmids like pK1F7T-0. PCR1B (118 bp) was obtained using primers PBLA-S (ACCCCTACATGTTTATTTTCTAAATACATTCA) and PBLA-N (GGAAATGTGGAATACTCATATGCTTCCTTTTTC AA) on template pK1F7T-0. The 90-bp *Nsp* I-*Nde* I fragment of PCR1B was then ligated into the 2,822-bp *Nsp* I-*Nde* I fragment of pKECMT-W to yield pKECMB-W (2,912 bp). KA12/pKIMP-UAUC cells transformed with this plasmid showed unimpaired growth on M9c media, with or without L-Phe and L-Tyr.

In order to randomize the codons for EcCM residues 64–68, a downstream *Mlu* I site was incorporated into the gene (at residues 74 and 75). The *Mlu* I site does not alter the protein sequence. Two overlapping PCR products were generated on template pKECMB-W: PCR5 (247 bp) was obtained with primers ECM1-S and ECMN-N (AGCTGGAACAGACGCGTAATGTAATGG), whereas PCR6 (111 bp) was obtained with primers ECMN-S (CCATTACATTACGCGTCTGTTCCAGCT) and ECM2-N (*Mlu* I site underlined in primers ECMN-N and ECMN-S). The two PCR products 5 and 6 were combined in approximately equimolar ratios and outside primers ECM1-S and ECM2-N were employed to yield PCR7 (331 bp). The 196-bp *Nhe* I-*Xho* I fragment of this product was subsequently ligated into the 2,716-bp *Nhe* I-*Xho* I fragment of pKECMB-W to yield pKECMB-Wm (2,912 bp). In complementation tests, pKECMB-Wm was indistinguishable from pKECMB-W.

To eliminate any wild-type background from re-ligated vector during library construction, a plasmid was constructed in which the 130-bp *Nhe* I-*Mlu* I fragment of pKECMB-Wm was replaced by the 832-bp *Nhe* I-*Mlu* I fragment of pET-24d (Novagen). This 832-bp fragment serves as a piece of "stuffer DNA," intended to destroy the function of the EcCM gene. The resulting plasmid pKECMB-S (3,614 bp) (Fig. 2B) showed no detectable complementation of KA12/pKIMP-UAUC cells, as expected.

Construction of libraries

The three libraries were prepared by PCR, incorporating oligonucleotides with randomized codons in the appropriate positions. PCR8 (for library [AHH]), PCR9 (for library [KAHH]), and PCR10 (for library [AHHL]) were each performed with template pKECMT-W using the Stoffel fragment of *AmpliTaq* DNA polymerase (Perkin-Elmer). Following a hot start (94 °C, 4 min), 25 cycles of denaturation (94 °C, 1 min), primer annealing (50 °C, 1 min), and extension (72 °C, 2 min) were performed, ending with an additional period of extension (72 °C, 10 min). The 507-bp PCR products 8, 9, and 10 were obtained using primer PTRC-S as the 5' primer and primers P_{AHH} (GAACAGACGCGTAATGTAATGGGCGTCCAGSNNNSNNNTTTTCCGAGCGTAATTAATCTTTC), P_{KAHH} (GAACAGACGCGTAATGTAATGGGCGTCCAGSNNNSNNNSNNNTCCGAGCGTAATTAATCTTTC), and P_{AHHL} (GAACAGACGCGTAATGTAATGGGCGTCCAGSNNNSNNNSNNNTTTTCCGAGCGTAATTAATCTTTC) as the 3' primers,

respectively. The randomized codons are underlined (N is an equimolar mixture of A, C, G, and T; S is an equimolar mixture of G and C). In bold is shown a mismatched T that results in a silent mutation in the gene; it is introduced to minimize bias toward the wild-type sequence in the annealing step of the PCR. In addition, the concentration of the primers containing the randomized codons was limiting in the PCR and the reaction was driven to completion to incorporate all the available 3' primer. The PCR products were purified (QIAquick PCR purification; QIAGEN), digested with *Nhe* I and *Mlu* I, and purified by agarose gel electrophoresis (recovery using QIAquick gel extraction; QIAGEN). The resulting 130-bp library fragments were then ligated into the 2,782-bp *Nhe* I–*Mlu* I fragment of pKECMB-S, which was gel-purified as above. Following ligation, the DNA was concentrated and desalted using a QIAquick PCR purification column (QIAGEN) and introduced, by electroporation, into competent XL1-Blue cells. For all three libraries, $>10^8$ transformants were obtained. The libraries were amplified in 50 mL LB broth (Miller, 1972), supplemented with 150 μ g/mL ampicillin, for 16 h at 30 °C. Plasmid DNA was then recovered (QIAGEN Midi prep) and introduced into KA12/pKIMP-UAUC cells by electroporation. Again, $>10^8$ transformants were obtained for all three libraries. The transformed cells were amplified in 100 mL LB broth, supplemented with 150 μ g/mL ampicillin and 35 μ g/mL chloramphenicol, for 16 h at 30 °C and then aliquotted and frozen in 20% glycerol at –80 °C.

Selection experiments

Selection experiments were conducted using M9c media (Kast et al., 1996b), which is based on M9 minimal medium (Miller, 1972). It consisted of Na₂HPO₄ (60 mg/mL), KH₂PO₄ (30 mg/mL), NH₄Cl (10 mg/mL), NaCl (5 mg/mL), 0.2% (wt/vol) D-(+)-Glc, 1 mM MgSO₄, 0.1 mM CaCl₂, thiamin-HCl (5 μ g/mL), 4-hydroxybenzoic acid (5 μ g/mL), 4-aminobenzoic acid (5 μ g/mL), 2,3-dihydroxybenzoic acid (1.6 μ g/mL), L-Trp (20 μ g/mL), sodium ampicillin (150 μ g/mL), and chloramphenicol (20 μ g/mL), pH 7.0. For Petri dishes, 15 g agar was added per liter. Where required, M9c was supplemented with 20 μ g of L-Phe and L-Tyr per mL. Because the gene for the *lac* repressor is deleted in KA12 cells, IPTG was not needed to induce expression of *pheA'* from the *trc* promoter. For complementation tests on individual clones, well-isolated colonies were picked from LB plates (supplemented with 150 μ g/mL sodium ampicillin and 35 μ g/mL chloramphenicol) and streaked to yield individual colonies both on M9c plates and on M9c plates supplemented with L-Phe and L-Tyr. For complementation tests on the libraries, 1-mL aliquots of cells were thawed, washed four times with M9c, diluted with M9c, and plated in triplicate both on M9c plates and on M9c plates supplemented with L-Phe and L-Tyr. In all cases, plates were incubated at 30 °C. Colonies were counted and evaluated after 3 days and then again after 7 days.

Acknowledgments

We thank Robin J. Rosenfeld for initial experiments. We also thank Professor Bruce Ganem for providing the gene for *E. coli* chorismate mutase (plasmid pJS47) and Professor Michael Hecht for helpful comments on the manuscript. Support of this work by the Skaggs Institute for Chemical Biology is gratefully acknowledged. 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.

References

- Bachmann BJ. 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev* 47:180–230.
- Bashford D, Chothia C, Lesk AM. 1987. Determinants of a protein fold: Unique features of the globin amino acid sequences. *J Mol Biol* 196:199–216.
- Beasley JR, Hecht MH. 1997. Protein design: The choice of de novo sequences. *J Biol Chem* 272:2031–2034.
- Blaber M, Zhang Xj, Matthews BW. 1993. Structural basis of amino acid α -helix propensity. *Science* 260:1637–1640.
- Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT. 1990. Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* 247:1306–1310.
- 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.
- Chakrabarty A, Kortemme T, Baldwin RL. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci* 3:843–852.
- Creighton TE. 1993. *Proteins: Structures and molecular properties*. New York: W.H. Freeman and Company.
- Deuschle U, Kammerer W, Gentz R, Bujard H. 1986. Promoters of *Escherichia coli*: Hierarchy of in vivo strength indicates alternate structures. *EMBO J* 5:2987–2994.
- Eggertsson G, Söll D. 1988. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol Rev* 52:354–374.
- Haslam E. 1993. *Shikimic acid metabolism and metabolites*. New York: John Wiley.
- Horovitz A, Matthews JM, Fersht AR. 1992. α -Helix stability in proteins. II. Factors that influence stability at an internal position. *J Mol Biol* 227:560–568.
- Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262:1680–1685.
- 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.
- 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.
- Lesk AM, Chothia C. 1980. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. *J Mol Biol* 136:225–270.
- Miller JH. 1972. *Experiments in molecular genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Miller JH, Albertini AM. 1983. Effects of surrounding sequence on the suppression of nonsense codons. *J Mol Biol* 164:59–71.
- Miller JH, Coulondre C, Hofer M, Schmeissner U, Sommer H, Schmitz A. 1979. Genetic studies of the *lac* repressor. IX. Generation of altered proteins by the suppression of nonsense mutations. *J Mol Biol* 131:191–222.
- Muñoz V, Serrano L. 1994. Intrinsic secondary structure propensities of the amino acids, using statistical ϕ - ψ matrices: Comparison with experimental scales. *Proteins Struct Funct Genet* 20:301–311.
- Nagi AD, Regan L. 1997. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* 2:67–75.
- O'Neil KT, DeGrado WF. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250:246–250.
- Perutz MF, Kendrew JC, Watson HC. 1965. Structure and function of haemoglobin. II. Some relations between polypeptide chain configuration and amino acid sequence. *J Mol Biol* 13:669–678.
- Predki PF, Agrawal V, Brünger AT, Regan L. 1996. Amino-acid substitutions in a surface turn modulate protein stability. *Nature Struct Biol* 3:54–58.
- Predki PF, Regan L. 1995. Redesigning the topology of a four-helix-bundle protein: Monomeric Rop. *Biochemistry* 34:9834–9839.
- Reidhaar-Olson JF, Sauer RT. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241:53–57.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Plainview, New York: 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.

- Shortle D. 1989. Probing the determinants of protein folding and stability with amino acid substitutions. *J Biol Chem* 264:5315–5318.
- Stewart J, Wilson DB, Ganem B. 1990. A genetically engineered monofunctional chorismate mutase. *J Am Chem Soc* 112:4582–4584.
- Vlassi M, Steif C, Weber P, Tsemoglou D, Wilson KS, Hinz HJ, Kokkinidis M. 1994. Restored heptad pattern continuity does not alter the folding of a four- α -helix bundle. *Nature Struct Biol* 1:706–716.
- West MW, Hecht MH. 1995. Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Sci* 4:2032–2039.
- Xiong H, Buckwalter BL, Shieh HM, Hecht MH. 1995. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proc Natl Acad Sci USA* 92:6349–6353.
- Ybe JA, Hecht MH. 1996. Sequence replacement in the central β -turn of plastocyanin. *Protein Sci* 5:814–824.
- Zhang S, Kongsaree 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.
- Zhou HX, Hoess RH, DeGrado WF. 1996. In vitro evolution of thermodynamically stable turns. *Nature Struct Biol* 3:446–451.