Chapter II: Library Analysis of SCHEMA-Guided Recombination

Portions of this chapter are reproduced from Meyer et al. 2003 "Library Analysis of SCHEMA-Guided Recombination" *Protein Science* **12**: 1686-1693.

Introduction

Recombination is an effective strategy for exploring protein sequences that differ significantly from those found in nature but maintain folded and functional structures. However, as the sequence identity between the proteins to be recombined decreases, the fraction of folded variants created also decreases (Ostermeier 2003). Several computational energy functions have been developed to predict which chimeras are likely to fold and function (Voigt et al. 2002; Moore and Maranas 2003; Saraf and Maranas 2003; Saraf et al. 2004; Hernandez and LeMaster 2005). These scoring functions examine potential pairwise clashes between amino acids introduced from different parents. The residue-residue interactions are predicted to be the dominant contributors to whether a chimera retains the parental structure (Drummond et al. 2005). However, most energy functions are typically tested using small and incompletely characterized data sets, making it difficult to determine how well the energy function is performing.

In this work we examine the pairwise scoring function SCHEMA that predicts which fragments of homologous proteins can be recombined without disturbing the integrity of the structure (Voigt et al. 2002). This is by far the simplest scoring function described and makes few assumptions. Based on a three-dimensional structure of a parent protein, SCHEMA identifies pairs of amino acids that are interacting, defined as those residues within a cutoff distance of 4.5Å, and determines the net number of interactions

broken when a chimeric protein inherits portions of its sequence from different parents (defined as *E*). A pair of residues whose identities do not change upon recombination cannot be broken by the recombination event.

Because calculating *E* (see methods) for all possible combinations of recombination sites, or crossovers, is computationally intractable, it is difficult to identify optimal crossovers that yield folded chimeras. The SCHEMA profile proposed by Voigt et al. (2002) circumvents this computational difficulty by finding compact, contiguous polypeptides with the largest number of intra-block interactions – these polypeptides correspond to fragments which, in theory, can be swapped with minimal cost. This is achieved by scanning the protein sequence with a window of defined size to create a disruption profile whose minima are predicted to represent crossover locations that preserve more interactions. It was proposed that the resulting fragments, or schemata, could be recombined using available laboratory recombination methods (Horton et al. 1989; Solaiman et al. 2000; Gibbs et al. 2001; O'Maille et al. 2002) to generate novel mosaic sequences that retain the parental structure.

A strong correlation exists between SCHEMA profiles and existing experimental data on chimeras from site-directed recombination and DNA-shuffling experiments. In particular, the vast majority of the crossovers found in functional chimeras containing 1 or 2 crossovers appear in or near the minima of their calculated disruption profiles (Voigt et al. 2002), suggesting that crossovers at other locations (e.g., profile maxima) are unfavorable. Furthermore, functional analysis of twelve lactamase chimeras revealed that proteins tolerate a limited level of *E*; only those with $E \leq 26$ were functional (Voigt et al. 2002). However, the small numbers of functional and nonfunctional chimeras analyzed in

these studies and the small number of crossovers incorporated make it difficult to determine just how SCHEMA predictions correlate with functional and structural disruption. We would like to know whether chimeras with low *E* have a higher probability of retaining parental function than those with the same effective level of mutation but chosen at random. We would also like to know whether the minima in the profile still correspond to the best recombination sites when multiple crossovers are allowed.

To address these questions we created a large library of chimeras with a broad range of *E* and examined which recombination events conserved function. For this test we chose to recombine two class A β-lactamases, TEM-1 and PSE-4, that share 40% sequence identity. The class A β-lactamases represent an ideal model system because functional chimeras are easily identified using antibiotic selection. Additionally, due to their medical significance there is a great deal of structural and sequence information available for class A β-lactamases. There are hundreds of β-lactamase sequences available in the database sharing between 99% and 15% sequence identity (Bateman et al. 2004) and twelve class A β-lactamases have been crystallized. The crystallized proteins share between 70% and 23% identity, and despite highly diverged sequences they have nearly identical structures with no more than 3.5 Å RMSD over all backbone atoms (Figure II-1) (Dideberg et al. 1987; Herzberg 1991; Knox and Moews 1991; Swarent et al. 1998; Ibuka et al. 1999; Kuzin et al. 1999; Tranier et al. 2000).

Sequence identity (%)

	1BSG	1BTL	1BUE	1BZA	1DY6	1E25	1G68	1MFO	1SHV	3BLM	4BLM
1BSG		37	41	41	40	23	30	42	40	31	41
1BTL	2.3		33	36	34	24	41	41	67	33	37
1BUE	1.8	2.2		47	74	23	34	38	35	34	42
1BZA	1.2	2.4	1.3		47	25	37	44	38	34	40
1DY6	1.6	2.0	0.5	1.2		26	36	36	39	36	44
1E25	3.2	3.0	2.9	3.1	2.9		23	26	25	21	24
1G68	2.5	1.4	1.9	1.9	1.9	2.7		34	45	36	28
1MFO	2.0	2.4	1.6	1.5	1.6	3.5	2.1		39	32	39
1SHV	2.9	1.3	2.5	2.8	2.4	2.7	1.6	2.7		30	34
3BLM	2.1	2.5	2.2	2.2	2.1	3.9	2.5	2.5	2.6		43
4BLM	1.4	2.3	1.6	1.2	1.6	3.3	1.7	1.7	2.8	1.4	

RMSD over backbone atoms (Å)

Figure II-1. Pairwise sequence identity and RMS deviation over the backbone atoms of all distinct class A β-lactamase structures designated by their protein data bank (pdb) code. Despite the highly diverged sequences, all lactamases crystallized have very similar structures.

Results

Library Design and Characterization

The SCHEMA-calculated profile shown in Figure II-2 was used to guide the creation of a diverse library of lactamase chimeras exhibiting a broad range of disruption. Eight major peaks in the profile correspond to eight polypeptides with the largest number of intra-block interactions. We allowed recombination at seven minima and six maxima of the disruption profile, yielding a library containing 2^{14} (16,384) possible unique chimeras. By calculating the exact disruption (E) of every sequence, we determined that the library contains chimeras with disruption values ranging from 7 to 113. Additionally, the chimeras display a broad range of effective mutations (m), from 7 to 75 amino acid substitutions relative to the closest parent.

Twenty-eight gene modules were synthesized chemically or by PCR (fourteen for each parent). Gene modules encoding structurally related elements contained identical unique 5' overhangs, but the sequences of the overhangs at each module boundary were

distinct and nonpalindromic. Each parental gene was assembled to confirm that no mutations were present in the modules and to validate that full-length genes could be created. Because ligation efficiency decreased as the number of fragments increased, we used a serial assembly protocol. Two or three adjacent gene fragments were ligated and purified using an agarose gel to create six distinct sets of products. This process was repeated using the ligated products until the full-length genes were assembled.

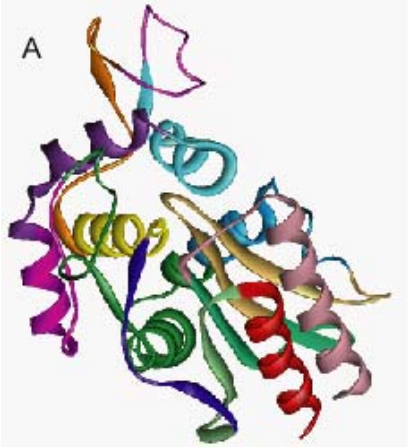
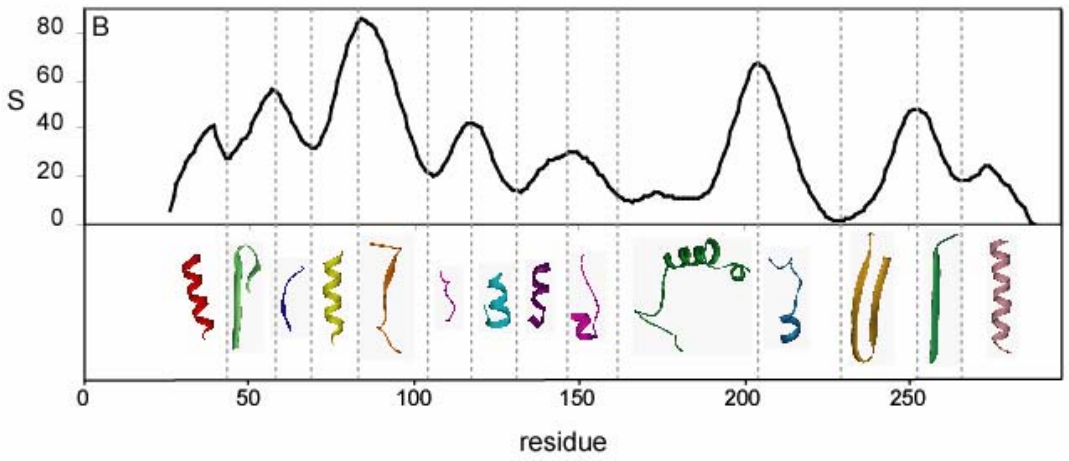


Figure II-2. Polypeptides recombined between TEM-1 and PSE-4. A: Polypeptide modules swapped between lactamases are mapped onto the structure of TEM-1. B: Profile disruption S was calculated for recombination of TEM-1 and PSE-4 using the crystal structure coordinates for TEM-1 and a window size of 14 (see methods). Residues are numbered based on the sequence of TEM-1. Vertical dashed lines represent crossover sites. (This figure is reproduced from Meyer et al. 2003 *Protein Science* **12**: 1686-1693).



To create the library of chimeric lactamases, equimolar mixtures of modules from each parent were mixed and ligated using a procedure similar to that for assembling the parental genes. *E. coli* were transformed with this library, and thousands of variants were plated on nonselective medium, i.e., LB-agar plates containing kanamycin. To determine

if the library contained any significant sequence biases, we measured the distribution of *pse-4* and *tem-1* modules in 79 randomly chosen chimeras using oligonucleotide probe hybridization (Meinhold et al. 2003). Figure II-3 shows the incorporation of the different parental sequences at seven positions throughout the genes and the frequency of crossovers between the modules probed, i.e., how often adjacent probed positions had sequence from the same parent. All chimeras exhibited a near-random crossover frequency between the probed modules (46 ± 5%), and the average frequency of observing the rarer of the two parents at each position was 40 ± 6%. Sequencing of unselected chimeras shows that up to 25% of clones may contain a single basepair deletion incorporated in the oligonucleotides used for construction. However, these deletions occur throughout both parental sequences so it is likely that the sampled portion of the library is reflective of the entire library. Assuming the (small) sequence bias arises from systematic errors in the assembly process, the average module bias can be used to calculate the probabilities of finding each chimera in our library. This type of analysis indicates that >90% of the unique chimeras occur with a probability $\geq 5.3 \times 10^{-6}$ at a confidence of 90%, and suggests that a sample of 150,000 unselected variants contains ≥65% of the unique sequences.

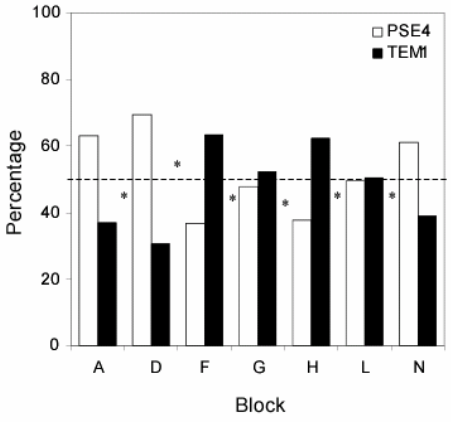


Figure II-3. Incorporation of *tem-1* and *pse-4* at different sequence positions in the unselected library. The presence of sequence from *tem-1* and *pse-4* at seven different module positions in 79 randomly picked unselected chimeras was determined using oligonucleotide probe hybridization. Asterisks represent the percentage of chimeras with crossovers occurring between adjacent probed positions. The dashed line represents the expected percentage of genes and crossovers in an unbiased library (50%).

Functional Chimera Characterization

Approximately 200,000 variants were plated on selective medium, LB-agar containing kanamycin and 20 μg/mL ampicillin. More than 100 colonies were observed, and sequencing these clones identified thirty unique functional lactamase chimeras, in addition to PSE-4 and TEM-1. Identification of the parental clones is consistent with predictions from hybridization results that suggest more than half of the chimeras were analyzed. Despite the PCR steps involved in library construction, the selected library displays a low point mutagenesis rate (0.005%). Only one chimera, the third sequence shown in Figure II-4, has amino acid substitutions. In this chimera, PSE-4 residues 265 and 266 are mutated from glutamine to histidine and threonine to serine, respectively. Examination of the TEM-1 and PSE-4 crystal structures reveal that these residues are both on the surface of the protein, and neither is in the active site (Jelsch et al. 1993; Lim et al. 2001a).

As shown in Figure II-4, the functional chimeras are highly mosaic, with 1, 2, 3, 4, 5, 6, or 7 modules swapped, and have between 7 and 67 effective mutations per chimera; the maximum possible in the library is 75. Furthermore, selected chimeras exhibited an average of 3.8 ± 2.0 crossovers, significantly lower than that expected from a random library (6.5 ± 1.8), and all chimeras have an even number of crossovers (2, 4, or 6), i.e., each functional chimera derives the A and N modules from the same parent. Modules A and N modules are derived from different parents in 41% of the clones in the unselected library.

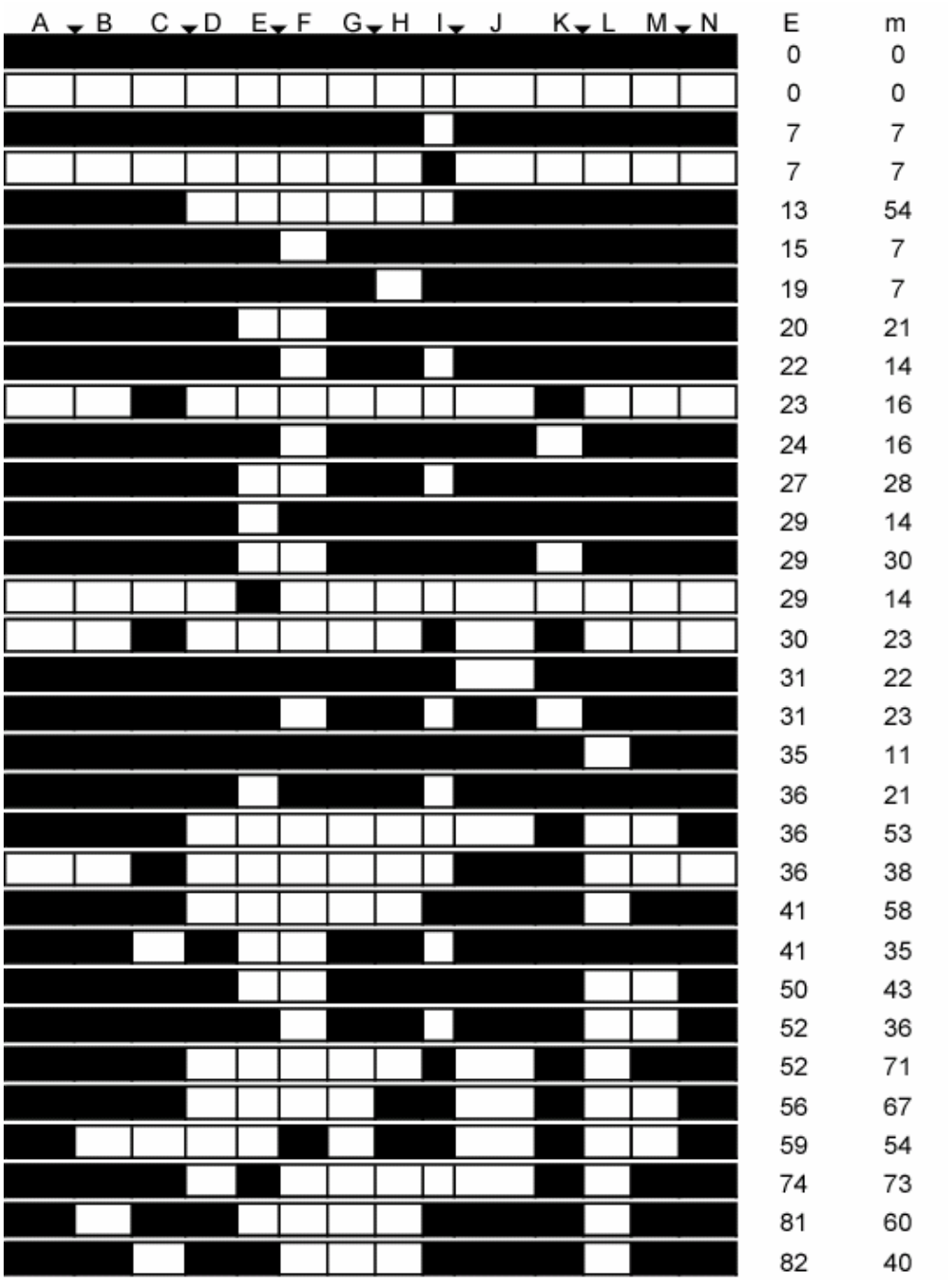


Figure II-4. Sequences, calculated disruption, and effective level of mutation of functional lactamases. Closed triangles indicate profile minima, and filled and open blocks represent TEM-1 and PSE-4 sequences, respectively. The calculated disruption *E* represents the number of interactions broken by recombination. Effective level of mutation (*m*) is the minimum number of mutations required to convert a chimera into one of its parents at only those residues recognized by SCHEMA, i.e., residues whose coordinates are defined in the TEM-1 structure (Jelsch et al. 1993).

Of the functional lactamases, only six derived both terminal fragments from TEM-1, five chimeras and one TEM-1. This indicates that chimeras which derive sequence from opposite parents at each position (chimera mirrors) are not functionally equivalent, even though SCHEMA does not distinguish them. Sequence analysis of randomly picked clones from the unselected library showed that 34% of the clones which acquire the A and N modules from the same parent contain TEM-1 at these positions. This small bias in the library does not account for the low level of TEM-1 terminal modules in functional chimeras (18%). The enrichment of one chimera from a mirror pair may arise because functional chimeras with TEM-1 terminal modules exhibit lower activity than those with PSE-4 at those positions. In fact, functional chimeras with TEM-1 terminal modules exhibit a significantly lower average MIC (250 μg/mL) than those with PSE-4 termini (1,400 μg/mL).

To determine if conservation of function corresponds to low *E*, we compared the distribution of *E* for the functional sequences with every theoretically possible unique chimera in our library. Figure II-5A shows the distributions of disruption for all chimeras in the selected and theoretical unselected libraries. The average *E* observed for functional clones (34 ± 21) is significantly lower than that calculated for the entire library (72 ± 16), indicating a strong association of low levels of disruption with maintenance of function. More than 85% of the functional chimeras have $E \leq 54$, while only 14% of the chimeras in the theoretical library fall below this threshold. We quantified the fraction of functional chimeras at each *E* in Figure II-5A by dividing the number of different functional sequences by the number of different sequences in the unselected library at

each *E* (Figure II-5B). This analysis reveals that the fraction of chimeras that retain lactamase activity decreases exponentially with increasing disruption.

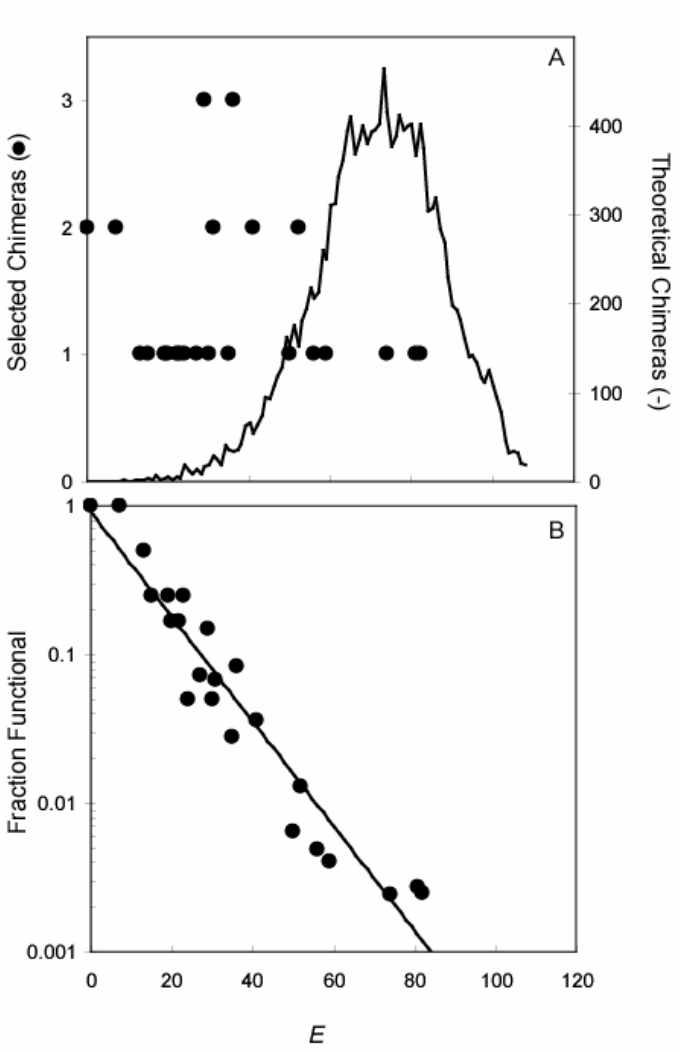


Figure II-5. Relationship between *E* and chimera function. A: The disruption distribution of all possible chimeras (solid line) is compared with those discovered in the selection for activity (●). B: The fraction of theoretical chimeras identified as functional is shown for each *E*. The data were fit to Equation (II-1) using N = 322 to obtain the probability that a disruption leads to a nonfunctional chimera, $f_d = 0.083$.

The fraction of chimeras in our library that retain function also depends on the level of mutation (Figure II-6), which raises the possibility that the low average *E* of functional chimeras could arise because low *E* corresponds to a lower average number of mutations. To investigate this, we calculated the relative difference $(E_{selected} - \langle E \rangle)/\langle E \rangle$ for each functional chimera, where $E_{selected}$ is the disruption of the functional chimera, and $\langle E \rangle$ is the average disruption of all chimeras in the theoretical library with the same

effective level of mutation (Figure II-7). The average relative difference for all functional chimeras in our library is -17.3%, suggesting that functional chimeras have lower disruption than those chosen at random with the same level of mutation. We then applied the Wilcoxon signed-rank test to evaluate the significance of these relative differences (Bernstein and Bernstein 1999). The Wilcoxon analysis yielded a ≥99% probability that the relative difference for all functional chimeras in any library is <0. Thus, chimeras that minimize *E* will have a greater likelihood of exhibiting undisturbed function than those chosen at random with the same level of mutation.

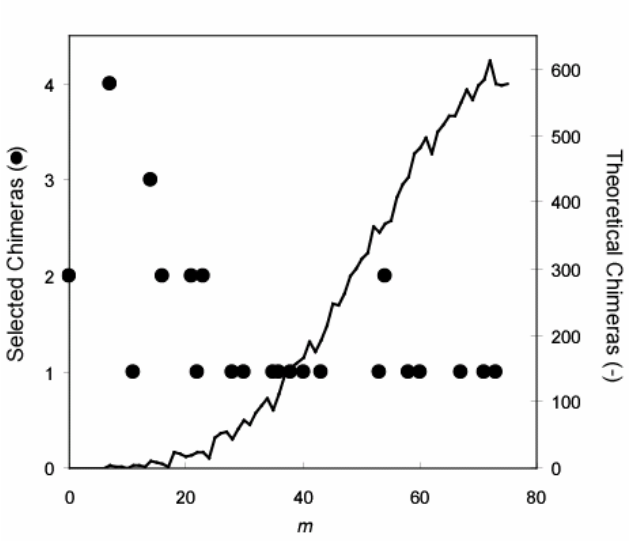


Figure II-6. Relationship between level of mutation and chimera function. The underlying distributions for the number of effective mutations (*m*) of all possible chimeras (solid line) and selected (●) chimeras are shown.

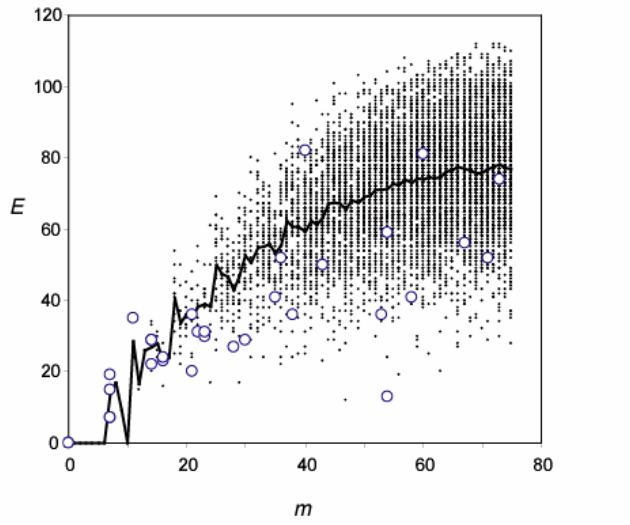


Figure II-7. *E* and *m* for all possible chimeras. At each level of mutation (*m*) where functional chimeras were obtained, the possible *E* values (•), the mean *E* for all possible chimeras (solid line), and the *E* of functional chimeras (○) are shown. Highly mutated chimeras have significantly lower disruption than the mean.

Discussion

Probabilistic Model for Chimera Function

Our results demonstrate that SCHEMA-calculated disruption (*E*) is a good metric for predicting functional conservation upon recombination. Sequence analysis of functional lactamases selected from a large library shows that chimeras with low *E* have a higher probability of retaining function than do chimeras with the same effective level of mutation but chosen at random. Our results also show that functional conservation decreases exponentially as *E* increases. This complements our previous finding, based on a small number of chimeras, that recombination disrupts protein function when it breaks many contacts in the three-dimensional structure (high *E*) (Voigt et al. 2002).

A simple probabilistic model can be invoked to anticipate the likelihood that lactamase chimeras will retain function. Assuming all contacts defined by SCHEMA are statistically independent, the fraction of possible recombinants at each *E* that retain function P_f is

$$P_f = (1 - f_d\, E/\mathrm{N})^{\mathrm{N}}, \qquad \text{(II-1)}$$

where N is the total number of interactions in the parental structures that can be disrupted upon recombination, and f_d is the probability that a disrupted contact yields a nonfunctional chimera. When N is large, as it is for proteins, this model yields a P_f that decays exponentially with *E*. Fitting Equation (II-1) to our data yields $f_d = 0.083$ (Figure II-5B). Because the experiment selected for functional chimeras, it could not uncover nonfunctional proteins that nonetheless retain proper fold. Furthermore, our use of a weak constitutive lactamase promoter to express chimeras in *E. coli* limits our ability to

identify lactamases with very low activity. Therefore, this value for f_d should be considered an upper bound on the probability that a disrupted contact yields unstructured or misfolded proteins, and the value of P_f that we calculate from f_d is therefore a conservative estimate of the probability that a protein structure will not be disrupted by recombination.

Identification of Optimal Crossover Locations

To simplify the identification of chimeras with low disruption, the SCHEMA algorithm generated a disruption profile such as shown in Figure II-2 by calculating the contribution each residue makes to the internal interactions within a fragment covered by a sliding window of a given size. We previously found that nondisruptive crossovers frequently occur in or near minima of SCHEMA profiles in chimeras with 1 or 2 crossovers, suggesting these minima may be a useful guide for generating folded and functional chimeras (Voigt et al. 2002). Interestingly, crossovers in functional lactamase chimeras from our library did not occur predominantly at these minima. Almost half of all crossovers in the functional lactamases occurred at the sites corresponding to profile maxima (Figure II-4). In addition, no functional chimeras were found with an odd number of crossovers: only 2, 4, and 6 crossovers generated functional chimeras. This crossover distribution is similar to that predicted for chimeras with a ≥10% probability of exhibiting undisturbed function ($E \leq 24$; Figure II-5B); of these chimeras, 88% have even numbers of crossovers and almost half of the crossovers (46%) occur at maxima. These findings suggest that interactions between polypeptides distal in the primary sequence, i.e., those not included in the profile calculation, should be considered when choosing

crossover locations. In other words, profile minima become a poor guide for predicting nondisruptive crossover locations when many crossovers can take place.

A better way to identify crossover points that minimize functional disruption is to determine which chimeras have the lowest *E*. But, because crossovers that do not lead to mutation will always minimize *E*, we also have to maintain a desired level of mutation. For chimeras arising from a small number of crossovers, it is easy to enumerate *E* for all possible chimera and identify crossover locations that minimize disruption. However, complete enumeration becomes impossible when multiple crossovers are allowed. For example, it is computationally intractable to calculate *E* for all possible seven crossovers between PSE-4 and TEM-1 and identify which seven-crossover library encodes chimeras with the lowest average *E* values, among libraries encoding chimeras with similar average levels of mutation. However, it is not difficult to evaluate thousands of randomly chosen seven-crossover libraries using SCHEMA to determine which ones encode chimeras with lower than average *E*. We find that this type of analysis is better than using profile minima to choose nondisruptive crossover locations for multiple-crossover libraries. For example, a PSE-4 and TEM-1 recombinant library made by allowing crossovers at the seven profile minima of Figure II-2 is predicted to encode 10 times fewer functional chimeras ($\langle E \rangle = 52 \pm 17$) than the best library found by searching 10,000 randomly generated libraries with seven crossovers ($\langle E \rangle = 33 \pm 10$), even though both libraries encode chimeras with similar levels of mutation.

Methods

Materials

E. coli XL1-Blue was from Stratagene (La Jolla, CA). Enzymes for DNA manipulations were obtained from New England Biolabs (Beverly, MA), Roche Biochemicals (Indianapolis, IN), or United States Biochemical Corp (Cleveland, OH). Synthetic oligonucleotides were obtained from Invitrogen (Carlsbad, CA). DNA purification kits were from Zymo Research (Orange, CA) and Qiagen (Valencia, CA), and other reagents were from Sigma Chemical Co (St Louis, MO) or Fisher Scientific (Pittsburgh, PA).

Calculations

For hybrids in which fragment(s) α and β are inherited from PSE-4 and TEM-1, respectively, the disruption (E) of the hybrid was calculated using Equation (II-2), where $c_{ij} = 1$ if residues are contacting (otherwise $c_{ij} = 0$), and $\Delta_{ij} = 0$ if i or j are identical in PSE-4 and TEM-1 (otherwise $\Delta_{ij} = 1$) (Voigt et al. 2002). Two residues were considered contacting if any atoms in the TEM-1 structure (1BTL) (Jelsch et al. 1993), excluding hydrogens, backbone nitrogens, and backbone oxygens, were within 4.5Å. Software to calculate the SCHEMA disruption *E* of protein chimeras is available on the web at http://www.che.caltech.edu/groups/fha.

$$E = \sum_{i \in \alpha} \sum_{j \in \beta} c_{ij} \Delta_{ij} \qquad \text{(II-2)}$$

To calculate the SCHEMA profile, a window of *w* residues was defined, and the number of intra-window interactions was counted. The profile disruption (S_i) of all residues in this window was incremented by the number of contacts within the window. The window was then slid along the protein sequence, and a profile was generated by incrementing the disruption of each residue (S_i) for all windows in which it resides. The numerical value of the SCHEMA-profile function S at residue i was defined by Equation (II-3); the magnitude of S_i corresponds with the level of predicted structural disruption for a crossover at a residue. A window of 14 residues was used to calculate the profile in Figure II-2.

$$S_i = \left(w^{-1/2}\right) \sum_{j=i-w+1}^{i} \sum_{k=j}^{j+w-2} \sum_{l=k+1}^{j+2-1} c_{kl} \Delta_{kl} \qquad \text{(II-3)}$$

Vectors

Lactamases were cloned into the vector pMon·1A2, which was created by cloning the gene encoding the heme domain of cytochrome P450 1A2 into pMon711 (Sabbagh et al. 1998). This vector was used for all selections. However, since this vector yields high background in oligonucleotide probe hybridization experiments, chimeras were cloned into pBC KS+ (Stratagene; La Jolla, CA) for these studies. *Escherichia coli* XL1-Blue transformed with these vectors were used for all analysis.

Library Construction

Twenty-eight gene modules were created to assemble the lactamase genes (fourteen for each parent). The protein modules correspond to TEM-1 residues 1-39 (A), 40-57 (B), 58-67 (C), 68-84 (D), 85-102 (E), 103-115 (F), 116-131 (G), 132-146 (H),

147-163 (I), 164-204 (J), 205-222 (K), 223-249 (L), 250-264 (M), 265-286 (N) and structurally related residues in PSE-4 identified using a structure-based alignment with Swiss-Pdb Viewer (Guex and Peitsch 1997). All modules used in assembly were double-stranded and contained unique nonpalindromic overhangs that allow for specific sequential ligation without concatamer production. Silent mutations were introduced into both genes at module boundaries (overhangs) to allow for facile assembly.

Chemically synthesized oligonucleotides used to create modules B, C, D, E, F, G, H, I, K, and M were phosphorylated using T4 polynucleotide kinase, and double-stranded modules were created from these by heating a reaction mixture containing 2.5 μM of complementary oligonucleotides, 10 mM Tris pH 8.0, 1 mM EDTA, and 50 mM NaCl at 95 °C for 2 min and subsequently cooling the reaction to room temperature at a rate of 0.1 °C per second. Modules larger than 70 basepairs (A, J, L, and N) were amplified with Vent DNA polymerase using primers containing SapI restriction sites; this allowed for rapid generation of complementary overhangs after amplification. Primers that amplified the terminal modules had a single SacI or HindIII site to allow for subsequent cloning. Amplified modules were purified by agarose gel electrophoresis, each (200 ng) was cut with 10 units of SapI at 37 °C for 24 hours, and digested modules were purified using agarose gel electrophoresis before assembly.

T4 DNA ligase was used to assemble *pse-4*, *tem-1*, and chimeric genes through a sequential process where pairs of adjacent modules were ligated, purified by agarose gel electrophoresis, and subsequently ligated to other assembled modules. Gene fragments composed of modules AB, CD, EFG, HIJ, KL, and MN were created in the first ligation reactions. For reactions in which the chimeric library was assembled, equimolar mixtures

of modules derived from each parent were used in this step. The ligated module dimers and trimers were further assembled, using the ligated fragments which had been purified with an agarose gel, to construct ABCDEFG and HIJKLMN using T4 DNA ligase. Because yields were low, ABCDEFG and HIJKLMN were amplified using Vent DNA polymerase and cleaved by SapI prior to assembly of full-length lactamases in a third ligation step; SapI created complementary overhangs at the G and H termini. Full-length constructs were treated with SacI and HindIII, purified using a Zymo DNA Clean and Concentrator Kit and ligated into pMon·1A2 and pBC KS(+), which were prepared similarly, to create the chimeric library.

Oligonucleotide Probe Hybridization

Sequences of 79 randomly selected chimeras from the unselected library in pBC KS+ were determined for 7 modules (A, D, F, G, H, L, and N) using oligonucleotide probe hybridization (Meinhold et al. 2003). The fraction of unique chimeras (*f*) found in a sample size *m* is

$$f \geq (1 - e^{m*\ln(1-v)}), \qquad \text{(II-4)}$$

where v is the probability of finding each sequence obtained from probe hybridization data.

Functional Chimera Selection

The minimal inhibitory concentration (MIC) of ampicillin for XL1-Blue *E. coli* containing pMon·1A2 is <5 µg/mL on LB-agar medium containing 10 µg/mL kanamycin. Therefore, functional selections using the pMon plasmid were performed

under conditions that gave no background, i.e., 20 μg/mL ampicillin and 10 μg/mL kanamycin. XL1-Blue were transformed with the unselected library using a heat-shock protocol recommended by the supplier, plated on selective medium, and incubated at 37 °C for 24 hours. Plasmid DNA was purified from all functional clones and digested with HindIII and SacI to confirm *pse-4* and *tem-1* length inserts (*ca*. 1 kb) were present. In addition, XL1-Blue were transformed with the purified DNA to verify the purified vectors conferred the ampicillin resistance. A majority of the clones had plasmids with an appropriate-size insert and conferred resistance in a second selection; fifty of these were sequenced.

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a nonparametric technique for investigating hypotheses about the median of a population (Bernstein and Bernstein 1999). While this test has less power than a *t* test for small sample sizes, i.e., is less likely to yield as dramatic a P value, we used this method because it makes no assumptions about the data being sampled from a normal distribution. To calculate the test statistic (W), we ranked the relative differences, $(E_{selected} - \langle E \rangle)/\langle E \rangle$, at each level of mutation according to their absolute magnitude and summed the rank scores according to the sign of the relative difference. This yielded W+ and W- values of 104 and –361, respectively.