Chapter VI: Mutagenesis to Restore Chimera Function

Introduction

Identifying characteristics of functional and nonfunctional chimeras is one way to address the underlying reasons for why chimeric proteins are nonfunctional. Another approach to this question is to determine if nonfunctional chimeras can be rescued through mutagenesis. Given enough of the right mutations, any chimera can be rescued as it returns to a wild-type sequence. However, whether nonfunctional chimeras are only a few or many mutations away from functional sequences is unknown.

There are many low *E* chimeras that are nonfunctional. It is unknown whether these chimeras are nonfunctional due to specific deleterious interactions, general lack of stability, or some other unknown factors. The specific mutations responsible for rescuing chimera function can indicate whether particular broken contacts are critical for function or chimeras are generally destabilized. Specific mutations that only rescue one or a few chimeras likely are responsible for correcting specific broken contacts. Mutations that rescue many chimeras and seem independent of specific sequences are likely global stabilizers (Poteete et al. 1997).

It is also unclear if all nonfunctional chimeras are equally distant in sequence space from functional sequences or if some chimeras are more likely to be rescued through mutagenesis. Chimeras that have low *E* are much more likely to function than chimeras with high *E*. It is possible that nonfunctional low *E* chimeras are closer in sequence space to functional sequences and may be easily rescued using random mutagenesis. To address these questions we randomly mutated nonfunctional chimeras to

examine which ones regain function, and what mutations are responsible for restoring function. The known TEM-1 stablizing mutation M182T was identified in half of the rescued chimeras. More thermostable proteins are more robust to random mutagenesis (Bloom et al. 2005b). To investigate whether this was true for mutations introduced by recombination, we introduced M182T into randomly selected chimeras and estimated the proportion of the library that might retain function if a more thermostable parent had been used.

Results

Random Mutagenesis Rescues Lactamase Chimera Function

To determine if chimeras could be rescued by random mutagenesis, DNA from all of the nonfunctional chimeras identified in Chapter IV and listed in Appendix II was combined. To ensure that no DNA from active chimeras was present, the collected DNA was transformed into *E. coli* and the transformants selected on ampicillin to verify that no colonies were produced. Following this verification, the DNA was PCR amplified under error-prone conditions as a single pool. The PCR products were cloned back into the expression vector and selected on ampicillin. Many ampicillin resistant clones were identified. However, sequencing these clones revealed that they were either known functional sequences, or functional sequences that had not been previously characterized. Apparently, during the mutagenic PCR, recombination similar to DNA shuffling occurred and scrambled the chimeras, making this strategy of mutating the whole set of nonfunctional at once unusable.

As an alternative, 10 inactive chimeras were subjected to error-prone PCR individually. The chimeras were chosen based on their hypothesized likelihood of being rescued: they all have low *E,* and the N- and C-termini originate from the same parents (Table VI-1A). Of the 10 chimeras, 8 were rescued by at least one single mutation (Table VI-2). There are 132 chimeras with *E* less than 30, of which 46 are nonfunctional. The ease with which the selected chimeras were rescued indicates that it is likely that many low *E* chimeras can be rescued similarly. For one of the two chimeras not rescued, none of the mutations identified in other chimeras can be incorporated because they are not found in the chimera due to differing parent blocks. For the other chimera only the M182T mutation is possible.

Table VI-1. Randomly Mutated Chimeras

A									E	m
S	T	S	P	T	S	S	S	*	15	12
T	S	T	S	S	T	T	T	*	15	13
T	T	T	S	S	T	P	T	*	20	29
P	T	T	T	P	T	T	P		20	71
S	T	T	T	T	T	S	S	*	20	64
S	P	T	T	P	S	T	S	*	22	81
T	P	S	P	T	T	T	T		22	55
S	P	P	T	T	S	T	S	*	28	76
P	S	T	T	S	T	T	P	*	29	71
S	S	T	T	T	P	T	S	*	30	80

B								E	m
P	T	S	T	P	S	T	T	45	86
P	S	T	T	S	S	T	S	50	89
T	S	T	P	S	T	T	S	53	62
T	P	T	T	T	T	T	S	53	52
P	P	T	T	T	S	T	S	54	85
T	T	T	P	T	T	T	S	55	56
P	T	P	T	T	T	T	S	59	87
P	P	T	S	S	T	P	S	59	103
T	T	T	P	T	P	S	S	61	78
P	S	P	S	T	P	T	S	62	84
T	P	P	P	P	S	P	S	67	84
T	P	S	P	T	T	T	S	71	65

Chimeras rescued by random mutagenesis are marked by an *. The sequence of a chimera is represented by the parents it inherits its blocks from: P for PSE-4, T for TEM-1, and S for SED-1. A: The initial set of chimeras chosen for their low *E* values that were randomly mutated to see if chimera function could be restored. B: The second set of chimeras chosen with higher *E* values to examine whether chimeras of any *E* could be rescued by random mutagenesis.

To explore whether the ease with which chimeras were rescued is a general property of all chimeras or due to the optimized population chosen, an additional 12

chimeras were chosen at high levels of *E* for mutagenesis (Table VI-1B). None of these chimeras were rescued. To ensure that this result was not due to sparse sampling of the possible mutants, the libraries were over sampled by ~10-fold.

Table VI-2. Mutations that Rescue Nonfunctional Chimera

Block			**Signal Seq**			**1**	**2**	**3**					**4**	**5**		**6**	**7**		**8**	
amino acid residue			8	22	27	63	72	99	100	114	120	147	153	171	174	182	191	193	224	261
Sequence	E	M																		
PSE-4	0	0				N	F	K	A	G	D	G	R	E	L	T	N	F	V	V
SED-1	0	0	Q		H	E	S	K	A	G	A	N	R	T	P	S	R	L	G	L
TEM-1	0	0		F		D	F	Q	N	T	R	E	H	E	P	M	R	L	A	V
STSPTSSS	15	12					**S**													
			L				**S**													
TSTSSTTT	15	13														***T***				
														S		***T***				
TTTSSTPT	20	29														***T***				
								R								***T***			T	
												G	**G**				Y			
																		L		
				L						S								**L**		
PTTTPTTP	20	71														***T***				
SPTTPSTS	22	81													**P**					
															P					**A**
						G									**P**					
SPPTTSTS	28	76	***L***		L								R							
PSTTSTTP	29	71														***T***				
SSTTTPTS	30	80	***L***																	
			L									**G**								
									S											**A**

Only unique sequences are shown, and mutations appearing alone in a chimera or in more than one chimera are shown in bold.

Several Mutations Can Rescue Function

Table VI-2 shows a list of the mutations that rescue each chimera; only unique sequences are shown. About half of the mutations mutate a single amino acid to an amino acid found in one or more of the other parents. This is not surprising for several reasons. First the residues in the other parents are more likely to appear upon random nucleotide mutation due to conservation in the genetic code. Second, changing a residue to match

one found in another parent may be correcting an interaction that was mismatched in the chimera.

The mutations that change the amino acid present to the amino acid present in a different parent sequence are: F72S (block 2 from TEM-1), E147G (block 3 from TEM-1), H153R (block 4 from TEM-1), L174P (block 5 from PSE-4) and M182T (block 6 frm TEM-1). Some of these positions have been previously characterized. H153R and M182T in TEM-1 not only revert to the amino acid found in both PSE-4 and SED-1, but also are known stabilizing mutations frequently identified in extended-spectrum TEM-1 variants (Knox 1995). The remaining residues have not been explicitly characterized, but all of them were found to be variable in a site-saturation study of TEM-1 (Huang et al. 1996). Examining the specific contacts that may be restored by a reversion shows that F72S and F193L both decrease the *E* of the chimeras by 1 or 2 contacts, respectively, and that L174P increases the *E* by 1 contact. From these limited studies it is not clear whether these mutations are likely to rescue many chimeras or are limited to specific sequences. Many of the mutations were isolated in only one chimera. While they are usually possible in at least one other chimera tested, it is unknown whether they rescue function in other chimeras.

There are two mutations which rescue several different chimeras. TEM-1 M182T rescues 4 of the 8 chimeras, and SED-1 Q8L rescues 3 of 8 chimeras. Of the rescued chimeras, M182T is identified in every one that has block 6 from TEM-1. The only chimera tested for which this mutation was possible and not identified was not rescued by any mutation. M182T was also identified in all four rescued chimeras as a single mutation. TEM-1 M182T is a well characterized mutation commonly found in extended-

spectrum TEM-1 variants (Knox 1995). It has been shown to mediate the effects of other deleterious mutations by increasing the stability of the protein by 2.7 kcal/mol (Wang et al. 2002). This mutation most likely has a similar effect on the chimeric proteins, providing them with enough additional stability to fold correctly. While PSE-4 already has a methionine at this position, previous studies and the widespread appearance of M182T here indicates that it is likely a global stabilizing mutation rather than correcting specific broken interactions.

The second mutation isolated from several different chimeras is the SED-1 Q8L mutation. This mutation appears alone in one of the three chimeras. It is accompanied by one additional mutation in one chimera and two additional mutations in the third chimera. Interestingly this mutation is in the signal sequence of the protein and not part of the mature protein. SED-1 is much less well characterized than PSE-4 or TEM-1 and there is no protein sequencing data or crystal structure currently available to give a definitive starting residue for the mature protein. However, the hypothesized start of the mature protein based on multiple sequence alignments is significantly further into the protein sequence than Q8. Why exactly this mutation rescues activity is not currently known. SED-1 was originally cloned from *Citrobacter sedalaki,* and it is possible that the signal sequence presumably optimized for this organism may be less efficient at transporting the protein to the periplasm in *E. coli*. However, transport of wild-type SED-1 to the periplasm appears normal (see Chapter V). A mutation in the signal sequence rescuing activity has not been previously observed in lactamases, and this brings our attention to the potentially key role of intracellular transport in an *in vivo* viability based screen or selection.

TEM-1 M182T Can Increase the Fraction of Folded Chimeras in the Library

Because M182T effectively rescued a high percentage of chimeras, it was introduced into 31 randomly chosen nonfunctional chimeras that have TEM-1 at block 6. Of the 31 randomly chosen chimeras, four were rescued by this single mutation (Table VI-3). Chimeras with low *E* are more likely to be rescued (Figure VI-1). All of the chimeras rescued have $E<35$, and they also all have the N- and C-termini from the same parent.

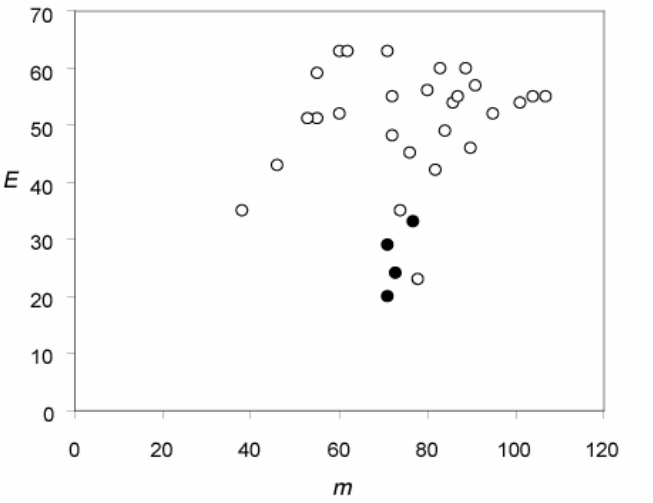

Figure VI-1. Rescued chimeras (solid points) are more likely to have low *E* than chimeras not rescued (open points) by M182T.

To estimate to effect of adding the M182T mutation to all chimeras in with TEM-1 we first calculated the probability of rescuing function P_{rescue} with respect to *E* for the small test set. This was done by fitting the 31 data points to a function of the form

$$P_{rescue} = \frac{1}{c + e^{bE+a}}, \qquad \text{(VI-1)}$$

where a, b, and c are parameters fit with the following constraints: $b \geq 0$, $0 \leq c \leq 1$. This function allows sigmoidal (c=1), exponential (c=0; a=0) and intermediate forms. For the

test set a = -4.8, b = 0.156, c = 1.0, and this fit corresponds well with P_{rescue} calculated for binned data (Figure VI-2). Using this function we calculated the probability of rescuing the remaining nonfunctional chimeras inheriting block 6 from TEM-1. Summing these probabilities shows that approximately 27 of the 442 nonfunctional chimeras (184 with block 6 from TEM-1) are likely to regain function if M182T was present in every chimera in the library. At low *E* nearly all chimeras should fold if the M182T mutation had been incorporated into the library (Figure VI-3A). However, the potentially rescued chimeras are spread over a wide range of *m* levels. Examining the extrapolated effect on fraction functional with respect to *m* shows that there are chimeras with high *m* that would likely function if TEM-1 M182T had been used rather than the wild-type TEM-1 (Figure VI-3B).

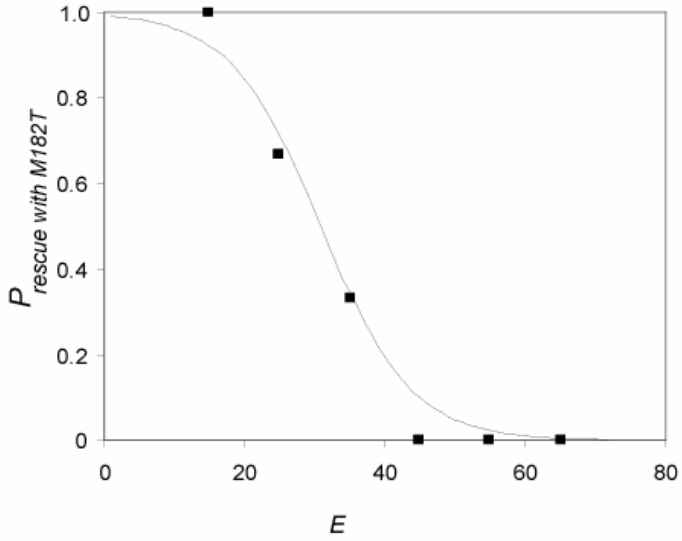

Figure VI-2. Probability of M182T rescuing function with respect to *E*. The points represent the fraction of chimeras rescued by M182T in a bin of 10 *E*. The curve is the fit of the individual data points (Figure VI-1) to Equation (VI-1).

Table VI-3. Randomly Chosen Chimeras M182T was Introduced Into

									E	m
P	T	T	T	P	T	T	P	*	20	71
S	S	T	T	T	T	T	S		23	78
P	T	T	P	P	T	T	P	*	24	73
P	S	T	T	S	T	T	P	*	29	71
P	S	T	P	S	T	T	P	*	33	77
P	T	P	T	S	T	T	T		35	74
P	S	T	T	P	T	T	T		35	38
P	T	S	T	P	T	T	T		42	82
P	T	T	T	T	T	S	T		43	46
P	S	P	S	T	T	T	T		45	76
P	P	T	T	S	T	S	S		46	90
T	P	T	T	S	T	S	S		48	72
P	P	T	T	S	T	T	S		49	84
P	S	T	P	T	T	S	T		51	55
T	S	T	T	T	T	T	S		51	53
P	P	P	P	T	T	P	S		52	60
P	P	T	P	P	T	S	S		52	95
P	S	T	T	T	T	P	S		54	101
P	P	T	S	T	T	S	S		54	86
T	T	T	T	S	T	P	S		55	72
P	P	T	P	T	T	T	S		55	87
P	S	T	P	T	T	P	S		55	107
P	T	T	P	T	T	P	S		55	104
S	T	S	S	P	T	T	T		56	80
P	P	T	S	S	T	T	S		57	91
T	S	S	S	T	T	T	S		59	55
P	S	P	P	T	T	T	S		60	83
P	S	P	T	T	T	T	S		60	89
P	P	S	T	S	T	T	S		63	60
P	S	P	T	P	T	S	T		63	71
P	T	S	T	T	T	T	S		63	62

Randomly chosen chimeras M182T was introduced into, they all inherit block 6 from TEM-1. The sequence of a chimera is represented by the parents it inherits its blocks from: P for PSE-4, T for TEM -1 and S for SED-1. Rescued chimeras are marked with an *.

Overall, the effect of adding M182T to the entire library is significant but not enormous. For the characterized library the increase in overall fraction folded is about 5%. However, because TEM-1 is found more frequently at block 6 than the other parents (Chapter IV) (>60% of characterized chimeras inherit block 6 from TEM-1) this effect is magnified compared to a random population of protein chimeras.

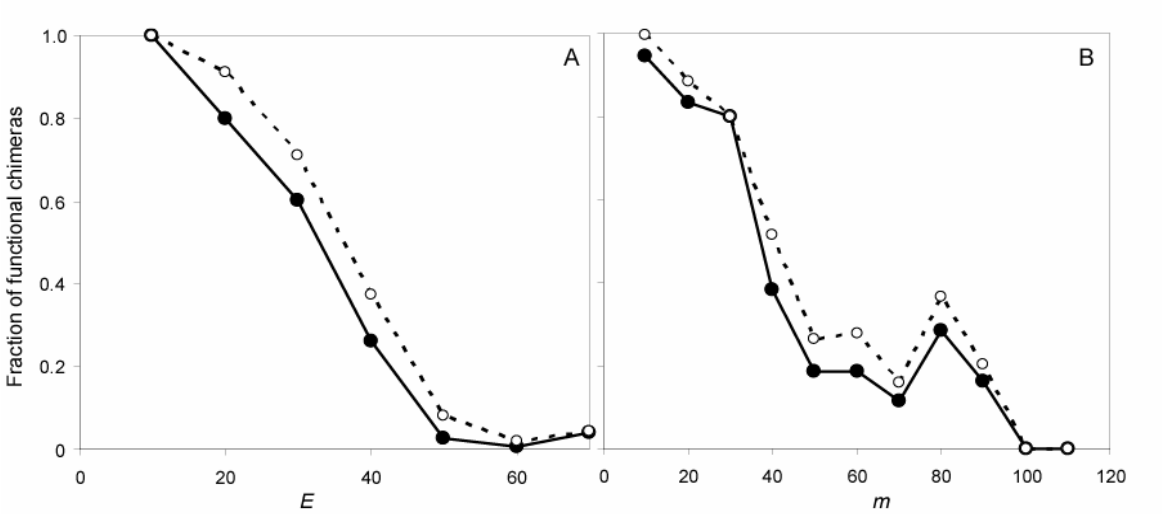

Figure VI-3. A: The fraction of functional chimeras in the library (solid points) and the fraction of the library folded if all possible library members contained M182T (open points) with respect to *E*. B: The fraction of functional chimeras in the library (solid points) and the fraction of the library folded if all possible library members contained M182T (open points) with respect to *m*. See methods for details.

Discussion

Many chimeras can be rescued with a single point mutation, either a global stabilizing mutation or a mutation that may correct specific broken contacts. Chimeras with low *E* are much more likely to be rescued than chimeras at high *E*, even if the chimera has many mutations to the closest parent (high *m*). This indicates that chimeras with low *E* are in an area of sequence space that is densely populated with folded proteins. Chimeras with low *E* are more likely to retain function and fold than chimeras at

higher *E*, and nonfunctional, low *E* chimeras are much closer to sequences that do encode folded functional proteins.

We have recently shown that more thermostable proteins are more tolerant to random mutations (Bloom et al. 2005b) and therefore can have a greater capacity to evolve (Bloom et al. 2006). Most mutations, beneficial or not, are destabilizing. More stable proteins are more likely to withstand a destabilizing mutation to fold correctly so that the phenotypic effects of that mutation are revealed. We have shown here that more thermostable proteins are likely more robust to mutations introduced through recombination as well as to randomly introduced mutations. Identifying mutations that increase thermostability indicates that starting with stabilized parents should increase the fraction of folded chimeras identified. This suggests that another way to increase the final fraction of folded variants in a recombination library is to begin with stabilized parent sequences.

Methods

Random Mutagenesis

DNA for inactive chimeras was sequenced prior to mutagenesis to confirm that no mutations were present in the chimera. Error-prone PCR was performed on each chimera in the following 100 uL reaction: 3 ng template, 1 μM forward and reverse primers listed on Table AII-3, 7mM $MgCl_2$, 75 μM $MnCl_2$, 200 μM dATP and dGTP, 50 μM dTTP and dCTP, 1x Applied Biosystems PCR buffer without $MgCl_2$ and 5 U of Applied Biosystems *taq* polymerase. Reactions were heated to 95 °C for 5 minutes then 14 cycles of 30 seconds at 95 °C, 30 seconds at 55 °C and 1 minute at 72 °C. PCR products were

digested with KpnI and PstI , cloned into pProTet (Clontech) cut with the same enzymes and transformed into XL-1 Blue (Stratagene).

Transformed *E. coli* were plated onto selective medium (35 ug/mL chloramphenocol and 10 ug/mL ampicillin) to identify sequences conferring resistance to ampicillin. A control aliquot was plated onto nonselective medium (35 ug/mL chloramphenicol) in order to assess how many chimeras were present in the selected sample. Colonies present on selective plates after 18 hours of growth at 37 °C were picked and the DNA extracted. The DNA was sequenced to identify mutations and retransformed into *E. coli* to verify plasmid conferred resistance. If no colonies were present on selective plates, 10 colonies were picked from nonselective plates to determine if the insert incorporation frequency, and typically 5 were sequenced to verify successful mutagenesis. A minimum of ~200,000 colonies were examined for each chimera not rescued. For rescued chimeras, typically many positive colonies were identified in much smaller libraries (~20,000 colonies).

Site-directed Mutagenesis

DNA for inactive chimeras was sequenced prior to mutagenesis to confirm that no mutations were present in the chimera. The TEM-1 M182T mutation was introduced using quick-change mutagenesis with the following primer and its reverse complement: 5'-CGT GAC ACC ACG ACC CCT GTA GCA ATG G. The altered codon is underlined. Mutagenesis reactions were transformed into XL-1 Blue (Stratagene) and plated onto selective and nonselective media as described above. Colonies growing on selective media after 18 hours at 37 °C were picked and the DNA extracted for

sequencing. For chimeras for which no colonies appeared on selective plates, 2 colonies were picked from the nonselective plates and the DNA extracted for sequencing to determine if the mutation was properly incorporated.

Extrapolation of Test Set to Library

The probability of M182T rescuing chimera function P_{rescue} was calculated by fitting the 31 data points to a function described by Equation (VI-1). This probability was applied to nonfunctional chimeras that inherited block 6 from TEM-1 (M182T is possible). To construct the figures, functional and nonfunctional chimeras in the naïve library (Appendix III) were counted for bins of 10 *E,* or 10 *m.* The point plotted for the naïve corresponds to the $N_{functional}/N_{total}$ for each bin. The point plotted for M182T added to the naïve library is $(N_{functional} + \Sigma P_{rescue})/N_{total}$, where ΣP_{rescue} is the sum of the probability of rescue for all chimeras within the bin.