Design of highly stable functional GroEL minichaperones

QINGHUA WANG, ASHLEY M. BUCKLE, NICHOLAS W. FOSTER,
CHRISTOPHER M. JOHNSON, AND ALAN R. FERSHT

Cambridge Centre for Protein Engineering and Cambridge University Chemical Laboratory, MRC Centre, Hills Road,
Cambridge CB2 2QH, United Kingdom

(RECEIVED May 3, 1999; ACCEPTED July 16, 1999)

Abstract

GroEL minichaperones have potential in the biotechnology industry for the refolding of recombinant proteins. With the aim of enhancing and widening their use, we have created two highly stable functional variants of minichaperone GroEL(193–345). A sequence alignment of 130 members of the chaperonin 60 (Cpn60) family was used to design 37 single mutations. Two small-to-large mutations, A223T, A223V and one similar-size mutation, M233L, all located in the hydrophobic core were found to stabilize the protein by more than 1 kcal mol⁻¹ each. Six stabilizing mutations were combined, yielding two multiple mutants that were 6.99 and 6.15 kcal mol⁻¹ more stable than wild-type protein. Even though some of the substituted residue pairs are close to each other in the protein structure, the energetic effects of mutation are approximately additive. In particular, the stabilizing substitution A223T is unexpected and would have been missed by purely structural analysis. In the light of previously reported successes employing similar methods with several other proteins, our results show that a homology based approach is a simple and efficient method of increasing the stability of a protein.

Keywords: protein engineering; protein stability; refolding chromatography; sequence homology

The molecular chaperone GroEL assists the folding of many newly synthesized proteins in *Escherichia coli* (Gething & Sambrook, 1992). The active site of GroEL, which is responsible for the binding of polypeptide substrates, is located on two helices H8 and H9 as well as an adjacent loop (residues 199–204) of the apical domain (residues 191–376) (Fenton et al., 1994; Buckle et al., 1997). Polypeptides corresponding to the apical domain of GroEL and fragments thereof, termed minichaperones, are active in vitro, facilitating the refolding of rhodanese and cyclophilin A (CyPA) and catalyzing the unfolding of native barnase (Zahn et al., 1996), and in vivo complementing temperature sensitive mutants of GroEL (Chatellier et al., 1998). Thus, the essential chaperone activity does not require the allosteric properties and central cavity provided by intact GroEL, which are refinements for optimal activity (Ben-Zvi et al., 1998).

When monodispersed on a solid support, minichaperones are highly active in assisting the renaturation of other proteins (Altamirano et al., 1997). “Refolding chromatography” based on this technique has immense potential in the biotechnology and pharmaceutical industries, in particular, when combined with immobilized protein disulfide isomerases (Altamirano et al., 1999). As

part of an overall effort to widen and enhance the activity of minichaperones, we have undertaken a mutagenesis study with the aim of creating a more stable protein. The stability of GroEL (193–345) is relatively low (ΔG of 6.6 kcal mol⁻¹); a more stable minichaperone could be used over a wide range of more stringent conditions, such as high denaturant concentration or high temperature.

Protein stability is not generally optimized during evolution. Thus, enhancement in protein stability can be achieved by designed amino acid substitutions. The rational approach to stabilization is extremely difficult since it is not generally possible to predict the energetic and structural response to mutation. The statistics of isolated helices and parts of sheets are predictable to various extents (Regan et al., 1996), but effects in proteins are often strongly dependent on the structural context. However, an alternative method, which takes advantage of naturally occurring variations among a family of homologous proteins, has provided several successful examples (Serrano et al., 1993; Steipe et al., 1994; Shih & Kirsch, 1995). The strategy is essentially as follows. First, the sequences of family members are aligned. Residues at positions of low conservation within the family are replaced by the representative residues as suggested by the alignment. The stability of the resulting set of single mutants is then measured. Stabilizing mutations are then combined with the aim of creating a “super-stable” variant, since the stabilizing effects of mutations are approximately additive when the residues concerned are not in contact. This procedure has been used to stabilize the tumor suppressor p53 DNA binding domain (Nikolova et al., 1998).

Reprint requests to: Alan R. Fersht, Cambridge University Chemical Laboratory, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, United Kingdom; e-mail: arf10@cam.ac.uk.

Abbreviations: CD, circular dichroism; Cpn60, chaperonin 60; CyPA, cyclophilin A; D, denatured state; DSC, differential scanning calorimetry; DTT, dithiothreitol; N, native state; PCR, polymerase chain reaction.

There are 130 sequences of homologous Cpn60 proteins available in the sequence data bank. Sequence alignment of their apical domains shows that whereas the overall sequence homology is rather good, the conservation at particular positions is relatively low. The minichaperone corresponding to the apical domain of GroEL is thus ideally suited to a homology based mutagenesis study. In this study, we have systematically substituted poorly conserved residues with the most representative ones within the family and determined the changes in stability upon mutation. By combining the stabilizing mutations, we have created two multiple mutants that have significantly increased stability while retaining full chaperone activity.

Results and discussion

Design of minichaperone mutants

Sequence comparison of *E. coli* GroEL(193–345) with the apical domains of 129 Cpn60 proteins reveals that 31 of the *E. coli* GroEL(193–345) residues occur with a frequency of less than 35%. These residues were selected to be individually replaced by the dominant residues at the respective positions and are listed in Table 1. The selected residues are scattered over the entire domain and have side-chain solvent accessibility ranging from completely buried (M233) to more than 90% exposure (R231). No preference of their location in structure is apparent. Residues involved in substrate binding (Fenton et al., 1994; Buckle et al., 1997) are highly conserved and remain unchanged.

Stability of minichaperones and their mutants

Two minichaperones were used in this study, GroEL(193–345) and GroEL(193–335), the latter is the minimal minichaperone identified to be active (Chatellier et al., 1998). The intrinsic fluorescence of both GroEL(193–335) and GroEL(193–345) from the two tyrosine residues, Tyr199 and Tyr203, is too weak to allow accurate monitoring of denaturation induced by urea or temperature. Circular dichroism (CD) spectroscopy was, therefore, used instead, which reflects changes in helicity upon denaturation. The values for $[\text{urea}]_{50\%}$ are 2.8 M and 3.1 M for GroEL(193–335) (Table 2) and GroEL(193–345) (Table 3), respectively.

For the first six mutations, GroEL(193–335) was used as parent protein (Table 2). However, its high tendency to aggregate presented severe difficulties in purifying some mutant proteins. The larger construct, GroEL(193–345), which does not aggregate to such a degree, was thus used as parent protein for all the subsequent mutations. Except for three mutations D316M, Q319T, and R322K, for which no proteins can be obtained, the thermodynamic parameters of each single mutation were determined by urea-induced equilibrium denaturation (Tables 2, 3). The m_{D-N} values of both constructs and their respective mutants are fairly stable, with an average of 2.23 ± 0.05 for GroEL(193–335) and 2.15 ± 0.03 for GroEL(193–345). The standard errors in $\Delta\Delta G_{D-N}^{\text{H}_2\text{O}}$ are 0.5 to 0.8 kcal mol⁻¹ for GroEL(193–335) and 0.3 to 0.4 kcal mol⁻¹ for GroEL(193–345), whereas the errors in $\Delta\Delta G_{D-N}^{\text{urea}}_{50\%}$ are much smaller, 0.06 to 0.12 kcal mol⁻¹ for GroEL(193–335) and 0.03 to 0.04 kcal mol⁻¹ for GroEL(193–345). The inherent inaccuracy of $\Delta\Delta G_{D-N}^{\text{H}_2\text{O}}$ has been suggested to originate from the long linear extrapolation to the absence of denaturant (Jackson et al., 1993). Therefore, $\Delta\Delta G_{D-N}^{\text{urea}}_{50\%}$ was

Table 1. Properties of residues^a selected for mutation

Residue ^b position	Location	Residue frequency ^c (%)	Side-chain solvent accessibility ^d (%)
N206	Loop between S6 and S7	25	26
K207	Loop between S6 and S7	12	51
P208	Loop between S6 and S7	35	82
T210	Loop between S6 and S7	22	65
G211	Loop between S6 and S7	7	
A212	Loop between S6 and S7	5	9
V213	S7	33	1
S217	S7	8	74
F219	S8	26	9
A223	S8	7	1
R231	H8	27	91
E232	H8	28	16
M233	H8	20	0
A239	H8	22	21
K242	H8	35	77
A243	H8	26	13
T266	H9	10	14
M267	H9	18	47
I270	Loop between H9 and S10	31	48
V271	Loop between H9 and S10	19	4
T294	H10	8	17
T299	Loop between H10 and S11	30	52
I305	Loop between H10 and S11	15	65
M307	Loop between H10 and S11	24	25
E308	Loop between H10 and S11	12	64
D316	Loop between H10 and S11	19	29
Q319	S11	21	46
R322	S11	33	45
N326	S11	12	53
T329	S12	16	27
I333	S12	35	18

^aUsing the crystal structure of minichaperone GroEL(191–376) (Buckle et al., 1997).

^bThe numbers refer to residue positions (Hemmingsen et al., 1988), while the letters represent amino acid residues using one-letter codes.

^cThe relative frequency of occurrence of GroEL(193–345) residues as calculated from the sequence alignment of 130 Cpn60 proteins.

^dPercentage of residue accessibility relative to that in an extended Gly-X-Gly tri-peptide (Miller et al., 1987), calculated with a 1.4 Å probe and the program AREAIMOL (CCP4, 1994).

used as an accurate measure of the changes in stability upon mutation. The effect on stability of each single mutation ranges from +1.55 to -1.78 kcal mol⁻¹, relative to the respective wild-type minichaperones (Tables 2, 3).

(i) Stabilizing mutations

Thirteen mutations stabilize the protein by more than 0.10 kcal mol⁻¹. Six occur in loops (K207N, A212E, V271L, I305L, E308K, and E308S), four are in β -strands (S217D, A223T, A223V, and N326T), while three in α -helices (M233L, M267L, and T294R). Mutations A212E, A223T, A223V, M233L, and N326T each stabilize the protein by more than 1 kcal mol⁻¹. The T_m values of these five mutants were analyzed by differential scanning calorimetry (DSC), the increases range from 1.5 to 6.8 °C (Table 4). The increases in chemical stability ($\Delta\Delta G_{D-N}^{\text{urea}}_{50\%}$) and thermostability

Table 2. Thermodynamic parameters^a of urea equilibrium denaturation for GroEL(193–335) and its mutants

Mutation ^b	m_{D-N} (kcal mol ⁻¹ M ⁻¹)	[urea] _{50%} (M)	$\Delta G_{D-N}^{H_2O}$ (kcal mol ⁻¹)	$\Delta\Delta G_{D-N}^{H_2O}$ (kcal mol ⁻¹)	$\Delta\Delta G_{D-N}^{[urea]_{50\%}}$ (kcal mol ⁻¹)
GroEL(193–335)	2.18 ± 0.10	2.81 ± 0.02	6.1 ± 0.3		
K207N	2.37 ± 0.27	2.93 ± 0.03	6.9 ± 0.8	−0.8 ± 0.8	−0.27 ± 0.08
P208S	2.03 ± 0.20	2.47 ± 0.05	5.0 ± 0.5	1.1 ± 0.6	0.76 ± 0.12
T210K	2.34 ± 0.14	2.69 ± 0.02	6.3 ± 0.4	−0.2 ± 0.5	0.27 ± 0.06
G211M	2.36 ± 0.14	2.81 ± 0.02	6.6 ± 0.4	−0.5 ± 0.5	0.00 ± 0.06
G211Q	2.16 ± 0.18	2.80 ± 0.03	6.0 ± 0.5	0.1 ± 0.6	0.02 ± 0.08
A212E	2.16 ± 0.13	3.43 ± 0.02	7.4 ± 0.4	−1.3 ± 0.5	−1.38 ± 0.07

^aExperiments were performed as described in Materials and methods. m_{D-N} , slope of the transition region of a denaturation curve that reflects the sensitivity to chemical denaturant; [urea]_{50%}, concentration of urea at which 50% of the protein is denatured; $\Delta G_{D-N}^{H_2O}$, free energy of denaturation (ΔG_{D-N}) in the absence of denaturant, calculated using $\Delta G_{D-N}^{H_2O} = m_{D-N}[\text{urea}]_{50\%}$; $\Delta\Delta G_{D-N}^{H_2O}$, difference in $\Delta G_{D-N}^{H_2O}$ between wild-type GroEL(193–335) and any mutant; $\Delta\Delta G_{D-N}^{[urea]_{50\%}}$, difference in $\Delta G_{D-N}^{[urea]_{50\%}}$ between wild-type GroEL(193–335) and any mutant, calculated using $\Delta\Delta G_{D-N}^{[urea]_{50\%}} = \langle m_{D-N} \rangle ([\text{urea}]_{50\%}^{\text{wt}} - [\text{urea}]_{50\%}^{\text{mut}})$, where $\langle m_{D-N} \rangle$ is the average value 2.23 ± 0.05 for GroEL(193–335) and its mutants. Positive values of $\Delta\Delta G_{D-N}^{H_2O}$ or $\Delta\Delta G_{D-N}^{[urea]_{50\%}}$ indicate decreased stability whereas negative values indicate increased stability.

^bThe numbers refer to residue positions, while the letters represent amino acid residues. For example, K207N means that K at position 207 is mutated to N.

(ΔT_m) correlate very well with each other (correlation coefficient of 0.97).

A212 is located in a loop between β -strands S6 and S7 and is almost completely buried. The side chain of A212 is about 4 Å away from the side chains of K207 and V213, and is surrounded by potential hydrogen-bonding partners from residues T210, G211, V213, I325, and water molecules. The mutation A212E stabilizes the protein by 1.38 kcal mol⁻¹ and 5.6 °C. This could be the result of the enhanced hydrophobic packing due to the introduction of one methylene group and hydrogen bonding with the introduced carboxyl group.

A223 is the C-terminal residue of β -strand S8 and is buried in the hydrophobic core. It occurs with a frequency of 7% in the Cpn60 family. The C β atom of A223 is surrounded by the methyl(ene) groups from residues I227, A251, and I301, and potential hydrogen-bonding groups from residues L222, D224, K225, A250, and I301, all within a radius of 4.5 Å. The mutation A223V stabilizes the protein by 1.03 kcal mol⁻¹ and 1.5 °C. The source of the stabilization is likely to be the increased hydrophobic packing within the protein core. It is surprising that the isosteric substitution A223T is 0.52 kcal mol⁻¹ and 4.1 °C more stable than A223V, given the destabilizing effect of burying a hydroxyl group in protein hydrophobic cores. One explanation is that the introduced hydrogen-bonding capacity of Thr223 more than compensates for the decreased hydrophobicity compared to Val223. A full explanation awaits structural analysis. This is one of the very few observations of the stabilization of the hydrophobic core by the introduction of a hydroxyl group (Eijssink et al., 1992b). Despite numerous studies by both experimental and theoretical approaches, the net contribution of hydrogen bonds to protein stability remains controversial (Dill, 1990; Serrano et al., 1992; Blaber et al., 1993; Chen et al., 1993; Honig & Yang, 1995; Lazaridis et al., 1995; Myers & Pace, 1996; Takano et al., 1999).

The most stabilizing mutation is M233L, a similar-size replacement on the buried face of α -helix H8 that participates in the hydrophobic core. In the crystal structure of GroEL(191–376) solved at 1.7 Å (Buckle et al., 1997), the side chain of M233 displays two conformations, indicating intrinsic local flexibility in the wild-type

protein. Surrounding residues within 4.5 Å, namely L221, I227, I230, V236, L237, L247, I249, and L262, form a small hydrophobic cluster. Among them, I230 and L237 have been postulated to interact with substrates (Fenton et al., 1994; Buckle et al., 1997). The mutation M233L stabilizes the protein by 1.78 kcal mol⁻¹ and 5.9 °C. It has been suggested that the substitution Met → Leu at buried positions stabilizes proteins because of the more favorable solvent transfer free energy (Fauchere & Pliska, 1983) and reduced entropic cost of ordering a leucine side chain on folding. The combined stabilizing effect is about 1.4 kcal mol⁻¹ without steric effects (Gassner et al., 1996).

N326 is the C-terminal residue of β -strand S11, with 53% of its side chain accessible to solvent. Its close neighbors within 4.5 Å are mainly oxygen and nitrogen atoms from surrounding residues I325, K327, D328, and T329. The mutation N326T stabilizes the protein by 1.35 kcal mol⁻¹ and 6.8 °C, where hydrogen-bonding interactions may contribute to the enhanced stability.

T294 is located in α -helix H10 close to its C-terminus and has a solvent accessibility of 17%. The mutation T294R stabilizes the protein by 0.37 kcal mol⁻¹. Stabilization may be the result of the favorable electrostatic interaction between the helical dipole and the introduced R294. I305 and E308 are located in the solvent-exposed segment of a mobile loop between H10 and S11. The region between residues 301 and 308 is the least resolved in the crystal structures of intact GroEL (Braig et al., 1995), which suggests an inherent flexibility of this region. The mutations I305L, E308K, and E308S stabilize the protein by 0.69, 0.34, and 0.47 kcal mol⁻¹, respectively. I305L and E308K are included in two multiple mutants described below. Other mutations K207N, S217D, M267L, and V271L stabilize the protein by 0.27, 0.17, 0.13, and 0.32 kcal mol⁻¹, respectively.

Generally, substituting a small hydrophobic residue in the core of a protein by another of larger size (“small-to-large”) will result in destabilization (Fersht et al., 1987). Protein cores are densely packed; to accommodate a substituted larger side chain, a reorganization of the local or even global structure is always required, which, in most cases, destabilizes the protein. However, in our study, three core mutations A223T, A223V, and V236L (0.58 kcal mol⁻¹,

Table 3. Thermodynamic parameters of urea equilibrium denaturation for GroEL(193–345)^a and its mutants

Mutation	m_{D-N} (kcal mol ⁻¹ M ⁻¹)	[urea] _{50%} (M)	$\Delta G_{D-N}^{H_2O}$ (kcal mol ⁻¹)	$\Delta\Delta G_{D-N}^{H_2O}$ (kcal mol ⁻¹)	$\Delta\Delta G_{D-N}^{[urea]_{50\%}}$ (kcal mol ⁻¹)
GroEL(193–345)	2.12 ± 0.09	3.12 ± 0.01	6.6 ± 0.3		
N206T	2.28 ± 0.05	2.79 ± 0.01	6.4 ± 0.1	0.2 ± 0.3	0.71 ± 0.03
V213A	2.26 ± 0.06	2.40 ± 0.01	5.4 ± 0.1	1.2 ± 0.3	1.55 ± 0.04
S217D	2.14 ± 0.04	3.20 ± 0.01	6.8 ± 0.1	−0.2 ± 0.3	−0.17 ± 0.03
F219Y	2.10 ± 0.05	3.15 ± 0.01	6.6 ± 0.2	0.0 ± 0.3	−0.06 ± 0.03
A223T	1.96 ± 0.05	3.84 ± 0.01	7.5 ± 0.2	−0.9 ± 0.3	−1.55 ± 0.04
A223V	2.05 ± 0.08	3.60 ± 0.01	7.4 ± 0.3	−0.8 ± 0.4	−1.03 ± 0.03
R231Q	2.28 ± 0.08	2.95 ± 0.01	6.7 ± 0.2	−0.1 ± 0.4	0.37 ± 0.03
E232D	2.17 ± 0.04	2.79 ± 0.01	6.0 ± 0.1	0.6 ± 0.3	0.71 ± 0.03
M233L	1.80 ± 0.04	3.95 ± 0.01	7.1 ± 0.2	−0.5 ± 0.3	−1.78 ± 0.04
A239Q	2.45 ± 0.08	3.00 ± 0.01	7.3 ± 0.2	−0.7 ± 0.4	0.26 ± 0.03
K242Q	2.43 ± 0.07	2.91 ± 0.01	7.1 ± 0.2	−0.5 ± 0.3	0.45 ± 0.03
A243S	2.14 ± 0.05	3.02 ± 0.01	6.5 ± 0.1	0.1 ± 0.3	0.21 ± 0.03
T266K	2.30 ± 0.05	2.97 ± 0.01	6.8 ± 0.1	−0.2 ± 0.3	0.32 ± 0.03
M267L	2.49 ± 0.08	3.18 ± 0.01	7.9 ± 0.2	−1.3 ± 0.4	−0.13 ± 0.03
I270G	2.20 ± 0.04	3.01 ± 0.01	6.6 ± 0.1	0.0 ± 0.3	0.24 ± 0.03
I270T	2.15 ± 0.03	2.90 ± 0.01	6.2 ± 0.1	0.4 ± 0.3	0.47 ± 0.03
V271L	2.01 ± 0.04	3.27 ± 0.01	6.6 ± 0.1	0.0 ± 0.3	−0.32 ± 0.03
T294R	2.40 ± 0.09	3.29 ± 0.01	7.9 ± 0.3	−1.3 ± 0.4	−0.37 ± 0.03
T294I	2.24 ± 0.06	3.10 ± 0.01	6.9 ± 0.2	−0.3 ± 0.3	0.04 ± 0.03
T299Q	2.26 ± 0.04	2.64 ± 0.01	6.0 ± 0.1	0.6 ± 0.3	1.03 ± 0.03
I305L	2.02 ± 0.05	3.44 ± 0.01	6.9 ± 0.2	−0.3 ± 0.3	−0.69 ± 0.03
M307L	2.18 ± 0.04	2.86 ± 0.01	6.2 ± 0.1	0.4 ± 0.3	0.56 ± 0.03
E308K	2.08 ± 0.08	3.28 ± 0.01	6.8 ± 0.3	−0.2 ± 0.4	−0.34 ± 0.03
E308S	1.99 ± 0.04	3.34 ± 0.01	6.6 ± 0.1	0.0 ± 0.3	−0.47 ± 0.03
Q319K	2.15 ± 0.07	3.14 ± 0.01	6.7 ± 0.2	−0.1 ± 0.4	−0.04 ± 0.03
N326T	2.12 ± 0.06	3.75 ± 0.01	8.0 ± 0.2	−1.3 ± 0.4	−1.35 ± 0.03
T329N	2.29 ± 0.06	2.78 ± 0.01	6.4 ± 0.2	0.2 ± 0.3	0.73 ± 0.03
I333V	2.15 ± 0.06	2.68 ± 0.01	5.8 ± 0.2	0.8 ± 0.3	0.95 ± 0.03

^aAll as explained in Table 2, except that the wild-type here is GroEL(193–345) and $\langle m_{D-N} \rangle$ is the average value 2.15 ± 0.03 for GroEL(193–345) and its mutants.

Q. Wang, A.M. Buckle, & A.R. Fersht, unpubl. result), one buried boundary mutation A212E, all are small-to-large stabilizing mutations. Another stabilizing mutation occurring nearby is the similar-size substitution M233L. It has been shown that while repacking a

protein interior is possible, the energetic cost could be very high (Kellis et al., 1988; Lim & Sauer, 1989; Sandberg & Terwilliger, 1989, 1991; Lim et al., 1992; Baldwin et al., 1993; Munson et al., 1994), unless mutations were designed to fill an existing cavity in protein cores (Eijsink et al., 1992a; Ishikawa et al., 1993). Thus the unexpected small-to-large stabilizing mutations reported here highlight the unique advantage of a homology based approach in protein stabilization.

Table 4. Midpoints (T_m) for thermal denaturation of GroEL(193–345) and representative stabilizing mutants determined by DSC^a

Protein	T_m (°C)	ΔT_m (°C)
GroEL(193–345)	67.1	
A212E	72.7	5.6
A223T	72.7	5.6
A223V	68.6	1.5
M233L	73.0	5.9
N326T	73.9	6.8

^aExperiments were performed as described in Materials and methods. T_m is the midpoint of thermal denaturation, ΔT_m is the change in T_m relative to wild-type GroEL(193–345), calculated as $\Delta T_m = T_m^{mut} - T_m^{wt}$.

(ii) Destabilizing mutations

Sixteen of the 34 single mutations destabilize the protein by more than 0.10 kcal mol⁻¹. Seven are in loops (N206T, P208S, T210K, I270G, I270T, T299Q, and M307L), three in β -strands (V213A, T329N, and I333V), and six in α -helices (R231Q, E232D, A239Q, K242Q, A243S, and T266K). The largest loss of stability is caused by the truncation of a buried hydrophobic residue in the core of the protein to a smaller one (V213A). The destabilizing effect is 1.55 kcal mol⁻¹. Similarly, another truncated mutation I333V at a largely buried position destabilizes the protein by 0.95 kcal mol⁻¹. These “large-to-small” mutations in the interior of a protein could destabilize the protein by the reduction in hydrophobicity and the introduction of one or more cavities of varying sizes (Lee, 1993; Buckle et al., 1996; Matthews, 1996).

(iii) Neutral mutations

There are five neutral mutations with $\Delta\Delta G_{D-N}^{[\text{urea}]_{50\%}}$ within the experimental error ($0.10 \text{ kcal mol}^{-1}$). These are G211M, G211Q (loop between β -strands S6 and S7), F219Y (β -strand S8), T294I (α -helix H10), and Q319K (β -strand S11).

Ultimately, the correlation of energetic and structural responses to mutation awaits the elucidation of the three-dimensional structures of the mutant minichaperones.

Semi-rational design of highly stable multiple mutants of GroEL(193–345)

Two multiple variants (M1: A212E/A223T/M233L/I305L/E308K/N326T; M2: A212E/A223V/M233L/I305L/E308K/N326T) were constructed, each encompassing six stabilizing mutations (Fig. 1). The only difference between M1 and M2 is the residue at position 223; T223 in M1, whereas V223 in M2. The values of $[\text{urea}]_{50\%}$ are 6.37 and 5.98 M for M1 and M2, respectively (Fig. 2), which correspond to 6.99 and 6.15 kcal mol^{-1} more stable than wild-type protein, in terms of $\Delta\Delta G_{D-N}^{[\text{urea}]_{50\%}}$. The sum of stabilizing effects of the constituent mutations is roughly equal to the increased stability of M1 and M2 (Table 5). Despite the close proximity of some residue pairs in the three-dimensional structure, such as A212 and N326, A223 and M233, I305 and M308, the stabilizing effects are additive rather than cooperative. The T_m values measured by DSC are 85.7 °C for M1 and 81.3 °C for M2, 18.6 and 14.2 °C higher than that of wild-type, respectively (Table 5).

Activity of minichaperone multiple mutants

CyPA is a peptidyl-prolyl *cis-trans* isomerase with a molecular mass of 18 kDa. The refolding of denatured CyPA requires the assistance of GroEL (Zahn et al., 1994) or minichaperones (Zahn et al., 1996) to achieve a high recovery of activity. The highly stable minichaperone variants M1 and M2 demonstrate high ac-

Fig. 1. Three-dimensional structure of minichaperone GroEL(191–345) (Buckle et al., 1997). Residues included in the multiple mutants M1 and M2 (A212, A223, M233, I305, E308, and N326) are indicated. Drawn with MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt & Murphy, 1994).

Fig. 2. Urea denaturation at 25 °C of GroEL(193–345) wild-type and two multiple mutants M1 and M2. The fraction of protein remaining folded for wild-type (Δ), M1 ($\diamond$), and M2 ($\circ$) is represented as a function of urea concentration. The protein concentration is $\sim 10 \mu\text{M}$. The best fitting of the experimental data to a two-state transition model using the program Kaleidagraph™ is indicated by solid lines.

tivity in assisting the refolding of CyPA, at the same level as full-length GroEL and wild-type GroEL(193–345) (Fig. 3). Spontaneous refolding of CyPA has been reported to be about 30% in the absence of chaperones. However, when human CyPA was subjected to denaturing condition (100 mM potassium phosphate, 10 mM dithiothreitol (DTT), pH 7.0 and 8 M urea) for at least 16 h at 30 °C, no significant spontaneous refolding upon removal of urea was observed. Therefore, the modified method described here renders the measured chaperone activity much more reliable than previously reported.

Validity of the approach

Out of the 34 single mutants made in this study, 13 are stabilizing, 16 are destabilizing, while 5 are neutral. It seems somewhat disappointing that the chance for obtaining stabilizing mutants is only about 38%. However, lowering the threshold of sequence identity from the previous 35 to 20% when selecting target residues, results in an increased success rate. Among a set of 18 mutations included, there are 13 stabilizing mutations, 2 neutral and only 3 destabilizing, as the majority of destabilizing mutations are of sequence identity between 20 to 35%. Thus, the success rate is increased to 72%, which undoubtedly demonstrates the validity of this approach. Nevertheless, qualitative correlation between the occurrence of a certain residue and its effect on stability (Steipe et al., 1994) has not been observed.

There are so far five Cpn60 proteins isolated from habitats of extremely high temperature that are closely related to GroEL in primary sequence. The overall homology pattern of sequence alignment among the apical domain of GroEL and these thermophilic Cpn60 proteins is quite similar to that among 130 Cpn60 proteins. From the whole mutant pool, a subset of 20 mutations was selected whereas 3 out of 5 sequences suggest the mutations. Among them, 8 mutations are stabilizing, 1 is neutral, while 11 are destabilizing. Not surprisingly, the majority of the stabilizing mutations, espe-

Table 5. Thermodynamic parameters^a for two highly stable GroEL(193–345) mutants

Protein ^b	m_{D-N} (kcal mol ⁻¹ M ⁻¹)	[urea] _{50%} (M)	$\Delta G_{D-N}^{H_2O}$ (kcal mol ⁻¹)	$\Delta\Delta G_{D-N}^{H_2O}$ (kcal mol ⁻¹)	$\Delta\Delta G_{D-N}^{[urea]_{50\%}}$ (kcal mol ⁻¹)	$\Sigma\Delta\Delta G_{D-N}^{[urea]_{50\%}}$ ^c (kcal mol ⁻¹)	T_m (°C)	ΔT_m (°C)
M1	1.84 ± 0.04	6.37 ± 0.01	11.7 ± 0.2	-5.1 ± 0.4	-6.99 ± 0.10	-7.09	85.7	18.6
M2	1.78 ± 0.04	5.98 ± 0.01	10.6 ± 0.2	-4.0 ± 0.4	-6.15 ± 0.09	-6.57	81.3	14.2

^aAll as defined in Tables 2 and 4.^bGroEL(193–345) M1 includes six mutations: A212E, A223T, M233L, I305L, E308K, and N326T, while GroEL(193–345) M2 includes six mutations: A212E, A223V, M233L, I305L, E308K, and N326T.^c $\Sigma\Delta\Delta G_{D-N}^{[urea]_{50\%}}$ is the sum of the effects on protein stability ($\Delta\Delta G_{D-N}^{[urea]_{50\%}}$) of all constituent single mutations.

cially all the five most stabilizing ones (A212E, A223T, A223V, M233L, and N326T), are indicated by this “small scale” sequence alignment. Therefore, in good agreement with previous studies (Shih & Kirsch, 1995), sequence alignment of thermolabile proteins with thermostable proteins is an efficient way of designing substitutions to stabilize proteins.

In summary, our results show that a homology-based approach is a simple and efficient method of increasing the stability of proteins, given that we are still far from a comprehensive understanding of the general principles governing protein stability to ensure success in rational design. In particular, some unexpected stabilizing substitutions picked up by this approach would have been missed by purely structural analysis. Elucidation of the structural response to mutations found by this approach will further our understanding of the principles of protein stabilization.

Materials and methods

Site-directed mutagenesis and production of proteins

Standard molecular biology methods were employed (Sambrook et al., 1989). The plasmids pRSETA GroEL(193–335) and pRSETA

GroEL(193–345) are previously described (Chatellier et al., 1998). Site-directed mutagenesis was made by polymerase chain reaction (PCR) with two complementary primers following the instruction of Quickchange™ (Stratagene, La Jolla, California). Mutations were confirmed by dideoxynucleotide sequencing using PRISM dye terminators (PE-ABI, Foster City, California). Multiple mutants were obtained by multiple cycles of PCR. The overexpression and purification of minichaperones and their mutants were carried out essentially as described previously (Zahn et al., 1996). Production of GroEL was as described (Corrales & Fersht, 1996).

Denaturation experiments

The chemical denaturation was performed using urea as denaturant in 50 mM Tris·HCl, 150 mM NaCl, pH 8.2 at 25 °C, monitored at 218 nm using a Jasco J-720 spectropolarimeter with cuvettes of pathlength of 0.1 cm, essentially as described previously (Golbik et al., 1998). The samples were usually incubated at 25 °C for 4–5 h to allow equilibrium to be reached before the measurements were performed. The plots of CD signals at 218 nm vs. urea concentrations were fitted to a two-state equation (Clarke & Fersht, 1993) using the program KaleidaGraph™. Two-state model is the simplest assumption that fits our experimental data very well.

$$F = \frac{(a_N + b_N[\text{urea}]) + (a_D + b_D[\text{urea}])\exp[m_{D-N}([\text{urea}] - [\text{urea}]_{50\%})/RT]}{1 + \exp[m_{D-N}([\text{urea}] - [\text{urea}]_{50\%})/RT]}$$

Here F is the recorded CD signal, a_N and a_D are the intercepts, b_N and b_D are the slopes of the pre-transition and post-transition regions where proteins are in native (N) and denatured (D) states, $[\text{urea}]_{50\%}$ is the concentration of urea at which 50% of the proteins is denatured, m_{D-N} is the slope of the transition region that reflects the sensitivity to chemical denaturant (urea), R is the gas constant, and T is the absolute temperature.

DSC experiments were performed essentially as described previously (Zahn et al., 1996). The protein concentration was ~60 μM in 25 mM sodium phosphate buffer, 100 mM sodium chloride, pH 7.8. The data were fitted to a non-two-state model and values for T_m , the midpoint of thermal denaturation, were obtained. The estimated error for T_m is ±0.1 °C. The thermal denaturation is not completely reversible as judged from the area of endotherms of rescanned samples; however, higher reversibility (>80%) was obtained by stopping the scan immediately after the transition as suggested by Zahn et al. (1996).

Refolding experiments

Human CyPA was denatured by incubating the protein (94 μM) in 100 mM potassium phosphate, 10 mM DTT, pH 7.0 and 8 M urea

Fig. 3. In vitro refolding of human CyPA in the presence of various chaperones. They are no chaperone (Spont.), 0.94 μM GroEL, 0.94 μM GroEL(193–345), 0.94 μM M1, and 0.94 μM M2. The final concentration of CyPA is 0.94 μM. One hundred percent activity was obtained with native CyPA. Standard error bars are shown.

at 30°C for at least 16 h. Refolding of CyPA was initiated by diluting the urea-denatured protein 100-fold to a final concentration of 0.94 μM in 100 mM potassium phosphate, 10 mM DTT, pH 7.0 supplemented with 0.94 μM of various chaperones. The concentration of GroEL is in terms of monomers. Refolding was allowed to proceed for 2 h at 25°C. Regain of CyPA activity was assayed spectrophotometrically using a modification of a method first described by Fischer et al. (1984). Briefly, immediately before the assay, 50 μL of the refolding samples were diluted into 987.5 μL of assay mixture (100 mM potassium phosphate, 10 mM DTT, pH 7.0 containing 0.3 mg mL⁻¹ α -chymotrypsin) at 10°C. 12.5 μL of peptide substrate N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (dissolved in trifluoroethanol/0.45 M LiCl at 6 mM) was then added and absorption at 390 nm was recorded as a function of time. One hundred percent activity was obtained for native CyPA, while the spontaneous refolding in the absence of chaperone is close to zero.

Acknowledgments

The authors would like to thank Dr. J. Chatellier for making the plasmids pRSETA GroEL(193–335) and pRSETA GroEL(193–345) available prior to publication and for helpful discussions, Drs. M.M. Altamirano, N. Kobayashi, and N. Tanaka for discussions and help. Q.W. is funded by studentships from UK Overseas Research Students Awards, Cambridge Overseas Trust, and the Herchel Smith Endowment from Department of Chemistry, University of Cambridge.

References

- Altamirano MM, Garcia C, Possani LD, Fersht AR. 1999. Oxidative refolding chromatography: Folding of the scorpion toxin Cn5. *Nature Biotechnol* 17:187–191.
- Altamirano MM, Golbik R, Zahn R, Buckle AM, Fersht AR. 1997. Refolding chromatography with immobilized mini-chaperones. *Proc Natl Acad Sci USA* 94:3576–3578.
- Baldwin EP, Hajiseyediavadi O, Baase WA, Matthews BW. 1993. The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. *Science* 262:1715–1718.
- Ben-Zvi AP, Chatellier J, Fersht AR, Goloubinoff P. 1998. Minimal and optimal mechanisms for GroE-mediated protein folding. *Proc Natl Acad Sci USA* 95:15275–15280.
- Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, Matthews BW. 1993. Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala $\rightarrow$ Ser and Val $\rightarrow$ Thr substitutions in T4 lysozyme. *Biochemistry* 32:11363–11373.
- Braig K, Adams PD, Brünger AT. 1995. Conformational variability in the refined structure of the chaperonin GroEL at 2.8 Å resolution. *Nature Struct Biol* 2:1083–1094.
- Buckle AM, Cramer P, Fersht AR. 1996. Structural and energetic responses to cavity-creating mutations in hydrophobic cores: observation of a buried water molecule and the hydrophilic nature of such hydrophobic cavities. *Biochemistry* 35:4298–4305.
- Buckle AM, Zahn R, Fersht AR. 1997. A structural model for GroEL-polypeptide recognition. *Proc Natl Acad Sci USA* 94:3571–3575.
- CCP4 (Collaborative Computational Project, Number 4). 1994. The CCP4 suite: Programs for protein crystallography. *Acta Cryst D50*:760–763.
- Chatellier J, Hill F, Lund PA, Fersht AR. 1998. In vivo activities of GroEL minichaperones. *Proc Natl Acad Sci USA* 95:9861–9866.
- Chen YW, Fersht AR, Henrick K. 1993. Contribution of buried hydrogen bonds to protein stability. The crystal structures of two barnase mutants. *J Mol Biol* 234:1158–1170.
- Clarke J, Fersht AR. 1993. Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* 32:4322–4329.
- Corrales FJ, Fersht AR. 1996. Kinetic significance of GroEL14·(GroES₇)₂ complexes in molecular chaperone activity. *Folding Design* 1:265–273.
- Dill KA. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
- Eijsink VG, Dijkstra BW, Vriend G, van der Zee JR, Veltman OR, van der Vinne B, van den Burg B, Kempe S, Venema G. 1992a. The effect of cavity-filling mutations on the thermostability of *Bacillus stearothermophilus* neutral protease. *Protein Eng* 5:421–426.
- Eijsink VG, Vriend G, Van der Zee JR, Van den Burg B, Venema G. 1992b. Increasing the thermostability of the neutral proteinase of *Bacillus stearothermophilus* by improvement of internal hydrogen-bonding. *Biochem J* 285:625–628.
- Fauchere J-L, Pliska V. 1983. Hydrophobic parameters II of amino acid side-chains from the partitioning of N-acetyl-amino-acid amides. *Eur J Med Chem* 18:369–375.
- Fenton WA, Kashi Y, Furtak K, Horwich AL. 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* 371:614–619.
- Fersht AR, Leatherbarrow RJ, Wells TN. 1987. Structure-activity relationships in engineered proteins: Analysis of use of binding energy by linear free energy relationships. *Biochemistry* 26:6030–6038.
- Fischer G, Bang H, Berger E, Schellenberger A. 1984. Conformational specificity of chymotrypsin toward proline-containing substrates. *Biochim Biophys Acta* 791:87–97.
- Gassner NC, Baase WA, Matthews BW. 1996. A test of the “jigsaw puzzle” model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc Natl Acad Sci USA* 93:12155–12158.
- Gething MJ, Sambrook J. 1992. Protein folding in the cell. *Nature* 355:33–45.
- Golbik R, Zahn R, Harding SE, Fersht AR. 1998. Thermodynamic stability and folding of GroEL minichaperones. *J Mol Biol* 276:505–515.
- Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, Georgopoulos CP, Hendrix RW, Ellis RJ. 1988. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* 333:330–334.
- Honig B, Yang AS. 1995. Free energy balance in protein folding. *Adv Protein Chem* 46:27–58.
- Ishikawa K, Nakamura H, Morikawa K, Kanaya S. 1993. Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry* 32:6171–6178.
- Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR. 1993. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* 32:11259–11269.
- Kellis JT Jr, Nyberg K, Sali D, Fersht AR. 1988. Contribution of hydrophobic interactions to protein stability. *Nature* 333:784–786.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structure. *J Appl Crystallogr* 24:946–950.
- Lazaridis T, Archontis G, Karplus M. 1995. Enthalpic contribution to protein stability: Insights from atom-based calculations and statistical mechanics. *Adv Protein Chem* 47:231–306.
- Lee B. 1993. Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of smaller size. *Protein Sci* 2:733–738.
- Lim WA, Farruggio DC, Sauer RT. 1992. Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* 31:4324–4333.
- Lim WA, Sauer RT. 1989. Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* 339:31–36.
- Matthews BW. 1996. Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB J* 10:35–41.
- Merritt EA, Murphy MEP. 1994. Raster3D version 2.0. A program for photo-realistic molecular graphics. *Acta Crystallogr D50*:869–873.
- Miller S, Janin J, Lesk AM, Chothia C. 1987. Interior and surface of monomeric proteins. *J Mol Biol* 196:641–656.
- Munson M, O'Brien R, Sturtevant JM, Regan L. 1994. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci* 3:2015–2022.
- Myers JK, Pace CN. 1996. Hydrogen bonding stabilizes globular proteins. *Biophys J* 71:2033–2039.
- Nikolova PV, Henckel J, Lane DP, Fersht AR. 1998. Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc Natl Acad Sci USA* 95:14675–14680.
- Regan L, Serrano L, Sali A, Horovitz A, Wilson C. 1996. Paper alert. *Folding Design* 1:R115–119.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: Laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.
- Sandberg WS, Terwilliger TC. 1989. Influence of interior packing and hydrophobicity on the stability of a protein. *Science* 245:54–57.
- Sandberg WS, Terwilliger TC. 1991. Energetics of repacking a protein interior. *Proc Natl Acad Sci USA* 88:1706–1710.
- Serrano L, Day AG, Fersht AR. 1993. Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J Mol Biol* 233:305–312.
- Serrano L, Kellis JT Jr, Cann P, Matouschek A, Fersht AR. 1992. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* 224:783–804.
- Shih P, Kirsch JF. 1995. Design and structural analysis of an engineered thermostable chicken lysozyme. *Protein Sci* 4:2063–2072.

- Steipe B, Schiller B, Pluckthun A, Steinbacher S. 1994. Sequence statistics reliably predict stabilising mutations in a protein domain. *J Mol Biol* 240:188–192.
- Takano K, Yamagata Y, Kubota M, Funahashi J, Fujii S, Yutani K. 1999. Contribution of hydrogen bonds to the conformational stability of human lysozyme: Calorimetry and X-ray analysis of six Ser → Ala mutants. *Biochemistry* 38:6623–6629.
- Zahn R, Buckle AM, Perrett S, Johnson CM, Corrales FJ, Golbik R, Fersht AR. 1996. Chaperone activity and structure of monomeric polypeptide binding domains of GroEL. *Proc Natl Acad Sci USA* 93:15024–15029.
- Zahn R, Spitzfaden C, Ottiger M, Wüthrich K, Pluckthun A. 1994. Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature* 368:261–265.