Evolution of Protein Quaternary Structure in Response to Selective Pressure for Increased Thermostability

Nicholas J. Fraser[1], Jian-Wei Liu[2], Peter D. Mabbitt[1], Galen Correy[1], Chris W. Coppin[2], Mathilde Lethier[3], Matthew A. Perugini[4], James M. Murphy[5], John G. Oakeshott[2], Martin Weik[3], Colin J. Jackson[1]

Affiliations:

[1]Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[2]Commonwealth Scientific and Industrial Research Organization, Land and Water Flagship, Canberra, ACT 0200, Australia

[3]Institut de Biologie Structurale, Commissariat a l'Energie Atomique, F-38027 Grenoble, France

[4]Department of Biochemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC 3086, Australia

[5]Molecular Medicine Division, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, VIC 3052, Australia

Abstract

Oligomerization has been suggested to be an important mechanism for increasing or maintaining the thermostability of proteins. Although it is evident that protein-protein contacts can result in substantial stabilization in many extant proteins, evidence for evolutionary selection for oligomerization is largely indirect and little is understood of the early steps in the evolution of oligomers. A laboratory-directed evolution experiment that selected for increased thermostability in the αE7 carboxylesterase from the australian sheep blowfly, *Lucilia cuprina,* resulted in a thermostable variant, *Lc*αE7-4a, that displayed increased levels of dimeric and tetrameric quaternary structure. A trade-off between activity and thermostability was made during the evolution of thermostability, with the higher-order oligomeric species displaying the greatest thermostability and lowest catalytic activity. Analysis of monomeric and dimeric *Lc*αE7-4a crystal structures revealed that only one of the oligomerization-inducing mutations was located at a potential protein:protein interface. This work demonstrates that by imposing a selective pressure demanding greater thermostability, mutations can lead to increased oligomerization and stabilization, providing support for the hypothesis that oligomerization is a viable evolutionary strategy for protein stabilization.

Key Words: Oligomerization, Directed Evolution, Thermostability, Carboxylesterase, Interface

Introduction

Oligomerization is one of the most fundamental biophysical interactions in protein chemistry**.** Bioinformatics and pull-down experiments have revealed that a large number of proteins exist as homo-oligomers consisting of two or more identical chains [1–4]. Similarly, many proteins function in the cell as hetero-oligomers composed of non-identical chains [5–9]. Oligomerization have been shown to play important roles in the genetic economy [10], functional gain [11–13], structural stability [14,15]**,** allosteric regulation [16–19] and protection from degradation [17]. Given the biological importance of oligomerization, there is substantial interest in understanding how the evolution of new oligomeric species occurs [20,21], engineering new oligomeric structure [22,23], and developing drugs targeted at complex assembly and disassembly [24,25].

The cellular milieu is a precarious environment for the evolution of oligomeric proteins as the physical forces that drive beneficial protein association are the same as those that drive deleterious aggregation [26,27]. It is well established that changes in protein sequence through non-synonymous point mutations, insertions and deletions can shift the balance of oligomeric states [21,28–30]. Both rational mutagenesis and directed evolution have been exploited to advance our understanding of how oligomers form [31–34], and to design new hetero-oligomers [35,36]. Protein flexibility, shape and symmetry have all been identified as being important for the formation of new oligomeric structure [37,38], since symmetrical and complementary interfaces may form stronger interactions than heterologous surfaces [39–41]. Despite a growing understanding of the nature of oligomers and protein complexes, it is difficult to predict how a new and beneficial protein-protein interface will develop, let alone design one *de novo* [23,42,43]. Moreover, despite intense study, *de novo* evolution of oligomeric structure has not been observed directly, to the best of our knowledge.

The fundamental importance of stability in protein evolution, and the stabilizing contribution of oligomerization (*via* reduction of the surface-to-volume ratio of the complex compared to a free monomer) is well established [44,45,2]. Protein-protein interactions can provide stabilizing polar or hydrophobic contacts (although interfaces in homodimers predominantly involve polar interactions), leading to tighter molecular packing and offering protection from denaturation [2,46]. Indirect evidence for the stabilizing effects of protein-protein interactions comes from thermophilic Archaea and bacteria, where oligomerization has been suggested to be one of the contributing factors to the high thermostability of proteins in these organisms [47,48]. Other examples of oligomerization leading to increased stability come from studies in which the disruption of protein-protein interfaces leads to decreased enzymatic activity [14,49] or the engineered formation of oligomers leads to increased stability [50,51]. However, despite several lines of evidence that oligomerization increases protein stability, most of the evidence for an evolutionary role for oligomerization had been indirect until the recent work of Perica *et al*., who used ancestral protein reconstruction to identify specific mutations directly involved in the process [20]. Notably, many of the mutations were found to be remote from the protein-protein interface.

*Lc*αE7 is a carboxylesterase involved in organophosphate insecticide resistance in the sheep blowfly, *Lucilia cuprina* [52]. We have previously reported an experiment in which we performed directed evolution of *Lc*αE7 in order to stabilize the protein for crystallization. However, there was little analysis of the evolution of thermostability, beyond the observation that after four rounds of evolution a variant (D83A, M364L, I419F, A472T, I505T, K530E, and D554G) was obtained that readily crystallized and displayed enhanced thermostability [53]. In this work, we have focused on the process by which the thermostable variant evolved and the nature of the stabilization. Here, we demonstrate that stabilization of the protein occurred *via* two related routes: stabilization of the monomeric protein through improved side

chain packing, and through the enrichment of more stable dimeric and tetrameric species. Oligomerization as a route to thermostability has been inferred through analysis of extant proteins, and through ancestral reconstruction [20], but seldom has *de novo* evolution of oligomerization been observed. These results allow us to understand the first steps in this process, which will benefit future engineering efforts.

Results

Directed evolution of *Lc*αE7 for increased thermal stability. As reported previously, the wild-type (WT) *Lc*αE7 protein was unstable both during and after purification [53]. Thus, the original rationale behind the design of this directed evolution experiment was to extend the half-life of *Lc*αE7 at temperatures that it might experience during expression and purification to facilitate crystallization. To achieve this, we designed a medium throughput screen in which *Lc*αE7 was heterologously expressed in *Escherichia coli*, replica plated onto filter paper, and assayed colorimetrically for activity using 2-napthyl acetate and fast-red dye. By randomly mutating the *Lc*αE7 gene in an expression vector and incubating the replica-plated *E. coli* transformed with the mutant library on filter paper for up to one-hour at elevated temperatures, we were able to iteratively select more thermostable variants and progressively increase the thermostability of the protein. After one round of random mutagenesis, only marginal improvements in activity over that of the WT enzyme were evident. Accordingly, these variants were not sequenced, but potentially improved variants were pooled for further random mutagenesis (which includes some recombination [54]) ahead of three further rounds of evolution. By Round 2, significant increases in thermostability were apparent, allowing specific mutants to be isolated and sequenced (**Table 1**). In Round 2 four variants (2a, 2b, 2c and 2d) were selected for DNA shuffling and mutagenesis resulting in Round 3 variants. In Round 3 two variants (3a and 3b) were selected for DNA shuffling and mutagenesis resulting in the Round 4a variant. The most stable variant from the experiment was *Lc*αE7-4a (**Figure 1A**), which incorporates three mutations first observed in Round 2 (I419F, A472T, K530E), one mutation from Round 3 (I505T) and three mutations from Round 4 (D83A, M364L, D554G). The mutation D83A was accidently missed in the previously reported structure of this protein (49). Subsequent DNA sequencing and analysis of electron density maps (5IKX, 5GC3) of *Lc*αE7-4a have confirmed that this mutation is present.

The *Lc*αE7-4a variant was then characterized in detail. We found that its half-life (time at which 50% of original activity remains) at 40 °C was indeed much greater than that of the WT enzyme (>60 min *vs*. 1.6 min), indicating that some of the mutations in *Lc*αE7-4a stabilize the protein (**Figure 1A**). Notably, whereas the activity of the WT protein rapidly diminished to background levels (7.5 % of the initial activity) with a half-life of 1.6 min, consistent with a single population of enzymes with similar thermostability, the activity of *Lc*αE7-4a decayed with a half-life of 24 min, but extrapolates to 56% of its original activity when fit to a one-phase exponential decay curve. Thus, *Lc*αE7-4a appears to be heterogeneous, consisting of one species with a half-life of 24 minutes, and other species that are essentially stable at 40 °C.

To investigate the composition of *Lc*αE7-4a in greater detail, we performed size exclusion chromatography (SEC), with kinetic and SDS-PAGE analysis of each fraction (**Figure 1B**). These analyses revealed that, other than minor contamination by a protein of approximately 70 kDa in fractions 12-18, *Lc*αE7-4a comprised the majority of the protein in the sample. However, the protein quaternary structure was not homogeneous. In fact, there were four distinct species present in the sample: a high molecular weight (HMW) species, which eluted near the void volume of the column, and three species relatively close in elution volume which primarily eluted in fractions 22, 24 and 26, respectively.

Characterization of the oligomeric species. The SEC was repeated several times, using a range of protein concentrations (5-20 mg/mL). *Lc*αE7-WT was consistently observed to elute primarily as a monomer, with small amounts of a HMW species (~120 mL) and small shoulders (170 mL, 185 mL) on the primary monomer peak (~215 mL) (**Figure 2A-C**). In contrast, and at all concentrations, *Lc*αE7-4a eluted with a significantly greater proportion of

larger species eluting at ~170 mL and 185 mL. The peak UV absorbance of these fractions relative to that of the monomer in *Lc*αE7-4a were approximately double those found for *Lc*αE7-WT. Because the cytosol of the cell, where *Lc*αE7-WT and *Lc*αE7-4a were expressed and assayed during the laboratory evolutionary selection, is a viscous environment [55], we repeated these experiments in the presence of 10% glycerol. This revealed that in a more viscous solution, the tetrameric and dimeric high molecular weight species comprised approximately half of the total *Lc*αE7-4a protein, whereas for *Lc*αE7-WT, the majority of the protein was monomeric (**SI Figure 1**).

We next performed SEC-multi angle laser light scattering (SEC-MALLS) analysis to obtain more accurate estimates of the absolute molecular weight of the various *Lc*αE7 forms [56]. This analysis indicated that the three lower molecular weight species, which are comprised exclusively of *Lc*αE7 protein (**Figure 1B**), are approximately 60 kDa, 120 kDa and 260 kDa (**Figure 2D**). These molecular masses correspond to the theoretical molecular masses of *Lc*αE7 monomer (66.3 kDa), dimer (132.6 kDa) and tetramer (265.2 kDa). We then performed analytical ultracentrifugation (AUC) equilibrium sedimentation velocity experiments on the complex sample of *Lc*αE7-4a (**Figure 2E**). These results confirm that *Lc*αE7-4a exists as a mixture of monomeric species and high-order oligomeric species, with the majority of the protein sedimenting as monomer (66 kDa *vs*. 66.3 kDa theoretical molecular mass) but a significant amount forming dimers (137 kDa *vs*. 132.6 kDa theoretical molecular mass) and tetramers (230 kDa *vs*. 265.2 kDa theoretical mass).

Whereas the stoichiometry of the ~260, ~136, and ~66 kDa species is readily apparent, the high molecular weight species (~400 kDa) was more difficult to analyze. We performed small angle X-ray scattering on this sample in order to determine, at low resolution, its molecular structure

(**Figure 2F**). These results indicate that this species is very heterogeneous, consistent with HMW protein aggregates that likely formed during recombinant expression and purification. The observation that the HMW peak is present in the WT sample at similar abundance to the *Lc*αE7-4a samples suggests that this species is not directly relevant to the evolutionary stabilization of the protein. Accordingly, it was not considered further in this study.

We then investigated whether the dimeric and tetrameric oligomeric species were representative of the protein in an equilibrium state, or were artifacts of the purification and concentration process. We obtained single 5 mL fractions from the center of the monomer and tetramer peaks from SEC and then performed SEC again after approximately five hours at 4 °C (monomer) or 24 hours at 4 °C (tetramer) (**SI Figure 2**). We found that the tetramer had equilibrated into approximately equal distributions of monomer, dimer and tetramer, while the monomer had also equilibrated to form dimeric and tetrameric species. Thus, the tendency of *Lc*αE7-4a to self-associate is a genuine equilibrium property.

The effect of oligomerization on enzyme activity and thermostability. Kinetic analyses of each species from *Lc*αE7-4a (**Figure 3** and **Table 2)** showed that the monomeric, dimeric and tetrameric forms were all significantly less active than the WT enzyme with the model substrate 4-nitrophenyl butyrate. However, the dimeric and tetrameric species were kinetically distinct from the monomer. Interestingly, the monomeric *Lc*αE7-4a species exhibited significant substrate inhibition (K_i), which was not apparent in either the dimeric or tetrameric species or the monomeric WT species. When the activity of the tetrameric and monomeric species were assayed at 46 °C over a time course (**SI Figure 3**), the tetrameric species retained greater activity, after one hour (the monomeric species was almost fully inactivated). Given that the evolutionary selection was used to identify mutants that could retain activity after one-hour

incubation at elevated temperature, this suggests that the greater thermostability of the higher-order species was selected for and made a significant contribution to the global enhancement of *Lc*αE7 thermostability.

Circular dichroism (CD) spectroscopy was used to investigate the thermostability of the three *Lc*αE7-4a species (**Table 3, SI Figure 4**). None of these species underwent reversible folding, meaning that accurate estimates of the free energy of folding energy could not be obtained. Nevertheless, the thermal melting curves provide an appropriate measurement of the temperature at which the majority of secondary structure in these proteins was lost (unfolding). All three *Lc*αE7-4a species were significantly more stable that the WT protein. As was the case with the catalytic activity of these species, the thermostability was different between the monomer, dimer and tetramer. Given the propensity of the higher-order species to slowly re-equilibrate (**SI Figure 2**), we were unable to obtain completely homogeneous samples, although those sampled were analyzed within two hours of SEC (after buffer exchange). This was apparent in the monomer and dimer samples, where the use of a three-state model showed that separate unfolding transitions corresponding to the monomer (55.5 °C) and dimer (66.2 °C) were apparent (Table 3; SI Figure 4). The tetrameric sample showed a single transition at ~63 °C. Thus, the higher order species are significantly more thermostable, consistent with the results from the activity decay measurement (**SI Figure 3**).

We repeated these thermostability measurements using differential scanning fluorimetry (DSF) [57]. These results showed that, whereas a complex mixture (pre-SEC) of *Lc*αE7-WT appeared to comprise a single dominant species that permitted the use of a 2-state unfolding model, the complex mixture of *Lc*αE7-4a consisted of two species, with one having a significantly higher melting temperature (**SI Figure 5**). When monomeric and tetrameric fractions from SEC (0.3

mg/mL) were immediately analyzed by DSF (the dimeric fraction was found to comprise a mix of both species), we could obtain pure samples with a single transition temperature. This confirmed that the higher-order tetramer fraction was more stable by ~5 °C. After concentration to 3 mg/mL, some re-equilibration occurred and both tetrameric and monomeric samples showed two transitions, one corresponding to monomer and the other corresponding to tetramer (**SI Table 1, SI Figure 6**). Overall, the melting temperatures were slightly lower with DSF than with circular dichroism (although the results were qualitatively identical), which can be explained by local exposure of hydrophobic surface (enabling binding of the fluorescent dye) before complete loss of secondary structure.

Structural comparison between monomeric and dimeric *Lc*αE7-4a.. We have previously reported the structures of monomeric and dimeric *Lc*αE7-4a [53]. Monomeric *Lc*αE7-4a was crystallized in the $P2_1$ space group (5IKX), with two molecules in the asymmetric unit. However, there was no close association reminiscent of a dimer/tetramer interface between the two molecules, or between them and any of the other molecules in the crystal lattice *via* symmetry (**Figure 5**). In confirmation of this, the PISA algorithm did not detect any potential association interfaces nor any potential assemblies [58]. In contrast, the *Lc*αE7-4a protein crystalized in the $C222_1$ space group (5CG3) had a monomer in the asymmetric unit, yet formed extensive, symmetrical, interactions with a neighboring molecule that were visually consistent with a typical oligomeric interface (**Figure 5A**). Analysis using PISA confirmed that this structure is likely to be a dimer under physiological conditions: 2260 $Å^2$ of protein surface areas is buried within the dimer interface. A total of 16 hydrogen bonds and 8 salt bridges were observed between the two-monomer chains (**Figure 5**). The monomeric ($P2_1$) and dimeric ($C222_1$) structures were very similar in terms of bother overall topology (C-α RMSD 0.339 Å)

and the main chain conformation at the dimer interface. Thus, in the case of *Lc*αE7-4a, dimerization does not require any large-scale rearrangement.

Interestingly, none of the mutations observed in *Lc*αE7-4a are located at the dimer interface (**Figure 5A, B**). However, three of the mutations are located on the protein surface near regions that are involved in the dimer interface and display some conformational disorder, as indicated by their atomic B-factors: D83A is located at the rear of the protein, M364L is located at a tight turn between two α-helices that overhang the dimer interface and active site entrance, D554G is located on an α-helix that is adjacent to one of the α-helices that comprise a significant part of the dimer interface. K530E is located at the rear of the protein in another comparatively disordered region.

The effect of amino acid substitutions on oligomerization and thermal stability of *Lc*αE7-4a. To better understand the effects of the four surface mutations (D83A, M364L, K530E and D554G) present in the thermostable *Lc*αE7-4a protein, they were individually reverted to the wild-type state. Purified proteins were subject to SEC in order to investigate the formation of oligomeric species **(Figure 4)**. The A83D and L364M reversion mutations had no discernible effect on the oligomeric equilibrium. The G554D reversion mutation resulted in a loss of dimer, which suggests that it may stabilize the adjacent α-helix that participates in the dimer interface, but had no effect on the proportion of tetrameric species. This observation is consistent with recent work that has shown that mutations can alter the oligomeric equilibrium of proteins through long-range effects on mobile regions (protein dynamics) [20]. The E530K mutation had no effect on the amount of dimer that was formed, but resulted in a significant reduction in the abundance of tetramer. In the X-crystal structure of *Lc*αE7-4a E530 is located in a putative tetramerization interface (**Figure 5C**). A salt bridge was observed between E530 and

K104 from the adjacent chain; this salt bridge would be absent from the WT protein and may act to stabilize the *Lc*αE7-4a tetramer. Relative to other amino acids, a large proportion of lysine residues are present on protein surface of *Lc*αE7. Lysine residues are less frequently found at interfaces of oligomeric proteins [59], and removal of lysine residues has been shown to increase the probability of protein crystallization [60], which is a controlled form of aggregation. Thus, it is likely that the K530E mutation is at least partially responsible for formation of the tetrameric species and the increased propensity of *Lc*αE7-4a to crystallize in the $C222_1$ space-group.

The effects of the surface mutations on the thermostability of *Lc*αE7-4a were measured by circular dichroism (**Table 4**). The four surface mutations in *Lc*αE7-4a were individually returned to the WT state. The charge reversal mutation, E530K, was the only mutation to significantly affect the thermostability of the *Lc*αE7-4a monomer, while the other three surface mutations A83D, L364M and G554D had no significant effect on the thermostability of *Lc*αE7-4a. This suggests that the K530E mutation along with at least one of the three internal mutations (I419F, T472A and I505T) play a role in the stabilization of monomeric *Lc*αE7-4a protein, relative to *Lc*αE7. We also investigated the A285S and F478L mutations that occurred in Round 2 mutants 2a and 2c, but did not fixate during the evolution. In terms of both oligomerization (**Figure 4**) and thermostability (**SI Figure 7**), these mutations were selectively neutral, explaining their disappearance by Round 4.

Stabilizing and oligomerization-inducing mutations are found in fly species closely related to *L. cuprina*. To determine if the amino acid substitutions we observed in this experiment also occur in orthologs of *Lc*αE7, we performed a multiple sequence alignment of α-Esterase 7 genes from four closely related species of flies: *Calliphora stygia* (New Zealand

brown blowfly), *Cochliomyia hominivorax* (New World screw-worm fly), *Musca domestica* (house fly) and *Haematobia irritans* (horn fly) (**SI Figure 8**). Interestingly, many of the mutations that fixated during this laboratory evolution experiment to produce *Lc*αE7-4a are commonly occurring amino acid differences amongst these sequences. Three of them, I419F, T472A and I505T, occur amongst the five species' sequences exactly as we find them in *Lc*αE7-4a and variation is also found at three of the other sites, albeit not to exactly the same amino acid D83(E), M364(A/E/H) and D554(E) in those cases. Only K530 is strongly conserved across the five species. Thus, several of the mutations that we observed in our directed evolution experiment are also present in orthologous enzymes.

Discussion

In this work, we have evolved an enzyme with a greater propensity to adopt higher-order molecular structure. In practical terms, the evolution of higher molecular weight species could be advantageous to the possible application of *Lc*αE7 as a bioscavenger to treat individuals poisoned by organophosphates, since it is known that proteins less than ~60 kDa are cleared through glomerular filtration [52,61]. The use of protein engineering or directed evolution to produce proteins with enhanced thermostability is perhaps one of the most common applications of these techniques: several high-profile studies have shed substantial light on the mechanisms by which thermostability can evolve [44,45,62–64]. The increase in the thermostability of monomeric *Lc*αE7-4a over monomeric *Lc*αE7 conforms to our expectations of how monomers can be stabilized, and is likely due to improved packing of hydrophobic residues in the core of *Lc*αE7, as has we have discussed previously [53]. However, the oligomeric states of these stabilized proteins are seldom characterized; enzymes are most often assayed either in crude lysate or after affinity purification. It is therefore possible that our observation that the increased thermostability of *Lc*αE7-4a can be partially attributed to enrichment of higher-order oligomeric species could be a more common process than currently thought.

Our results exemplify a trade-off between activity and stability that has been demonstrated in a number of studies of other enzymes [44,45,65]. Specifically, we observe the monomeric species to be more active than the dimeric and tetrameric species. Furthermore, we find that the monomer displays substrate inhibition, whereas the other species do not. This suggests that the monomeric species is able to adopt conformations in which multiple substrate molecules can bind at one time, inhibiting the reaction, whereas these, presumably more open, conformations are not available to the dimeric and tetrameric species. This is consistent with

the location of the dimer interface: directly at the entrance to (but not occluding) the active site. Thus, we see that oligomerization can have significant stabilizing effects on certain protein regions [37] that can indirectly affect activity [66].

Recent work has shown that oligomers are more tolerant to mutation, and thus more evolvable, primarily owing to their stability [64]. Indeed, numerous studies have implicated oligomerization in protein thermostability, through observation of oligomers in thermophiles [47,48], or engineered disruption/formation of protein:protein interfaces [14,49]. Despite this, there have been few, if any, direct observations (to the best of our knowledge), of proteins spontaneously evolving thermostability through the gain in additional oligomeric structure in laboratory settings in response to a thermal challenge. Because of our inability to replicate the stepwise evolution of oligomers in the laboratory and characterize evolutionary intermediates, our understanding of the sequence changes and structural mechanisms that underlie protein self-association has been limited. Only recently has the use of ancestral protein reconstruction allowed an evolutionary process of oligomerization to be recapitulated, yielding valuable new insights into the mechanism of oligomerization and highlighting the role of remote mutations and protein dynamics [20]. The results presented here show that the evolution of thermostability in *Lc*αE7 has occurred in parallel with enrichment of higher-order oligomeric states that display greater thermostability than the monomeric species, that point mutations can lead to a gradual shift in the oligomeric equilibrium of states in a protein population and that these mutations can alter the propensity for monomers to oligomerize either *via* the formation of new bonds, or indirectly *via* the stabilization of mobile regions of the protein.

Methods

Directed Evolution. The directed evolution experiment was performed as described previously [53]. Selection of improved variants was carried out through library screening (approximately 100,000 colonies). Cells were spread onto Lysogeny Broth (LB)-agar plates supplemented with 100 $\mu g.mL^{-1}$ ampicillin at a cell density which resulted in ~200 colonies per plate. Colonies were blotted onto Whatman grade 3 filter paper (GE Healthcare) and incubated for 1 h at 50 °C (increased to 55 °C, 60 °C, and 70 °C for rounds 2, 3 and 4). The residual esterase activity was assayed by spraying heat-treated colonies with substrate solution (0.8% w/v Fast Red, 10 mM β-naphthyl acetate, and 100 mM Tris, pH 7.0). The most active mutants after thermal challenge were identified as those colonies that produced the most intense red color. A secondary screen of the best variants was carried out, in which the best variants from the plate screen were picked and grown in 96-deep well plate formats. The overnight cultures were heat stressed for 1 h at the same temperatures used in the primary screen, and 25 μL overnight cultures were assayed using a Molecular Devices plate reader at 490 nm in the presence of 0.5 mM β-naphthyl acetate, 0.5 mM Fast Red dye, and 100 mM Tris (pH 7.0). The best 5–10 variants of each generation were carried forward to the next generation, and the final product *Lc*αE7-4a and several variants from rounds 2 and 3 were sequenced at the Micromon Sequencing Facility, Melbourne, Australia.

Site-Directed Mutagenesis. Surface mutations of *Lc*αE7-4a were individually reverted back to the wild-type state (A83D, L364M, E530K, or G554D) by the DNA-fragment assembly method described by Gibson and coworkers [67]. Pairs of primers were designed to introduce the required codon change in the *Lc*αE7 gene located in a pETMCSIII vector. The resulting vectors were sequenced at the Biomolecular Resource Facility, Australian National University, Australia.

Protein Expression and Purification. *Lc*αE7-WT and *Lc*αE7-4a were expressed in *E. coli* BL21 (DE3) cells and grown in Lysogeny Broth supplemented with 100 μg/mL of ampicillin in the leaky expression vector pETMCSIII [68]. Cultures were grown at 25 ˚C for 18 hours, without induction. The cells were resuspended in buffer A (10 mM imidazole, 50 mM HEPES-NaOH (pH 7.5), 300 mM NaCl) and lysed by sonication using the Sonic Ruptor 400 (Omni International). Clarified lysate was applied to a 5 mL Ni-NTA column (Qiagen) for nickel-affinity chromatography and the protein was eluted with buffer A supplemented with 300 mM imidazole. Protein purified by affinity chromatography was tested for esterase activity with 4-nitrophenyl butyrate and analyzed by SDS-PAGE to identify pure fractions.

Size Exclusion Chromatography-Multi Angle Laser Light Scattering. Pure *Lc*αE7 fractions were pooled and further purified by size exclusion chromatography using a Hiload 26/600 Superdex-200 column (GE Healthcare), equilibrated with buffer B (200mM NaCl, 20mM HEPES pH 7.5). The effect of viscosity on *Lc*αE7-WT and *Lc*αE7-4a was investigated by size exclusion chromatography as described above, with the addition of 10% glycerol to the running buffer. The total protein loaded (estimated by total area area under the curve) for the *Lc*αE7-WT samples was 24.5, 12.1, 11.5 and 10.8 mg (highest to lowest) and for *Lc*αE7-4a 16.6, 15.3, 13.2 and 13 mg (highest to lowest). For SEC-MALLS purified *Lc*αE7-4a (3.0 mg/mL) was applied at a flow rate of 1 mL/min to a WTC-030S5-column (Wyatt Technology) connected to a DAWN8+ multiple angle laser light scattering (MALLS) detector and tREX refractive index detector (Wyatt Technology). The column was equilibrated with buffer B. Data analysis was performed using ASTRA (Wyatt Technology).

Small-angle X-ray Scattering. Data were collected in 200mM NaCl, 20mM HEPES pH 7.5, 5% glycerol using an inline S200, 5/150GL (3.2 mL; GE Healthcare) column at 0.2 mL/min

flow rate. Data were collected serially in 2 sec quanta. The radius of gyration was unable to be estimated for any of the eluted protein owing to scatter patterns synonymous with the presence of high molecular weight aggregates. Data were collected using a Pilatus 1M detector at a distance of 3.3m allowing collection of data in the Q range: 0.00486-0.25 nm^{-1}.

Analytical ultracentrifugation. Sedimentation velocity experiments were conducted in a Beckman model XL-A analytical ultracentrifuge at a temperature of 20 °C. *Lc*αE7 and *Lc*αE7-4a were solubilized in 20mM HEPES, 500mM NaCl, pH 7.5 was analyzed at an initial concentrations of 3.0 mg/mL. The sample and reference solution were loaded into conventional double sector quartz cells and mounted in a Beckman 4-hole An-60 Ti rotor. 380 μl of sample and 400 μl of reference solution were centrifuged at a rotor speed of 40,000 rpm, and the data were collected at a single wavelength (295 nm) in continuous mode, using a step-size of 0.003 cm without averaging. The partial specific volume (0.736 mL/g at 20 °C for both wild-type and variant proteins), buffer density (1.0202 g/mL) and buffer viscosity (1.0635 cp) were computed using the program SEDNTERP [69]**.** Sedimentation velocity data at multiple time points were fitted to a continuous sedimentation coefficient [*c(s)*] distribution and a continuous mass [*c(M)*] distribution model [70–72] using the program SEDFIT, which is available at www.analyticalultracentrifugation.com.

Circular Dichroism. Purified *Lc*αE7 at 0.3 $mg.mL^{-1}$ was exchanged into 20 mM NaH_2PO_4 20 mM NaCl buffer pH 7.5. Circular dichroism data was collected with a Chirascan circular dichroism spectrometer (Applied Photophysics) in a 1 mm quartz cuvette. The protein was heated from 25 °C to 90 °C at a rate of 1 °C min^{-1} while ellipticity was monitored at 208 nm. The thermally induced unfolding of *Lc*αE7 was not reversible. Circular dichroism data were fitted using the by non-linear regression to a two-state model (equation 1) using GraphPad Prism

6.00 (GraphPad Software). $y_{obs} = y_F \frac{(y_F - y_U)}{1+\exp\left(\frac{TT_{50}-T}{c}\right)}$ **Equation 1.** Where, y_{obs} is the observed elipticity, y_F_nand y_U are the elipticity values observed for the native and unfolded states, respectively. TT_{50} is the temperature at which the population of unfolded protein is 50 % and c is the slope. Curves were fit by non-linear regression using GraphPad Prism 6.00 (GraphPad Software, USA). A three state model was used for curve fitting with the monomer and dimer data using the software CDpal [73].

Thermal Shift Fluorescence assay. The thermal shift of LcαE7 and LcαE7-4a was measured using the ViiA7 real-time PCR system (Life Technologies) in MicroAmp EnduraPlate Optical 384 well plates (Life Technologies) with a final volume of 20 μl per well. The plates were covered with optical seal and shaken before denaturation. Each sample was measured in duplicate using the SYPRO Orange dye to measure protein unfolding. Protein was diluted to concentrations between 0.3 mg/mL and 3 mg/mL in buffer B. The 5000x SYPRO Orange dye stock was diluted in buffer B to a final concentration of 20x in the plate. Thermal denaturation was measured by increasing the temperature from 20 °C to 90 °C at a rate of 0.017 °C/s and the plate was measured with wavelengths of excitation at 470nm and emission at 580 nm. The TT_{50} was calculated by nonlinear regression with the Boltzmann sigmoidal equation in Graphpad Prism 6.00 (GraphPad Software, USA). A three state model was used for curve fitting with the *Lc*αE7-4a data using the software CDpal [73].

Enzyme Kinetic Assays. Purified *Lc*αE7-WT and *Lc*αE7-4a were assayed against the ester substrate 4-nitrophenyl butyrate in buffer B (200mM NaCl, 20mM HEPES pH 7.5) at 25 °C. Formation of the product, 4-nitrophenolate was monitored at 405 nm using an Epoch microplate spectrophotometer (BioTeK Instruments). Velocities were obtained from the initial linear portion of the reaction progress curves, and product concentration was determined with

the extinction coefficient calculated from a standard curve of 4-nitrophenol. Assays were done in triplicate and the protein concentration used in the range of 0.01 to 0.05 μM as final concentration. The Michaelis constant and the inhibition constant were determined by non-linear regression of the initial velocities to the Michaelis-Menton equation or substrate inhibition equation using GraphPad Prism 6.00 (GraphPad Software). For thermal stability assays, purified *Lc*αE7-WT and *Lc*αE7-4a were incubated at 40 °C, aliquots were removed at set time points and specific activity was assayed as described above with 200 □M 4-nitrophenyl butyrate in triplicate. To measure the decay of *Lc*αE7-4a monomer and tetramer activity over a time course, the proteins were incubated at 46 °C in a buffer consisting of 20 mM HEPES, 200 mM NaCl and 10% glycerol. Aliquots removed at various time points and subsequently assayed with 200 μM 4-nitrophenyl butyrate in triplicate as described above. The data were fit by non-linear regression to a one-phase exponential decay curve using GraphPad Prism 6.00 (GraphPad Software, USA).

Acknowledgements

CJJ thanks the Australian Research Council for a Future Fellowship. PDM thanks the Science and Industry Endowment Fund for a John Stocker Fellowship. This work was supported by a grant from the USA Defense Threat Reduction Agency HDTRA1-11-C-0047.

Abbreviations used:

AUC, analytical ultracentrifugation; DSF, differential scanning fluorimetry; CD, Circular Dichroism; HMW, high molecular weight; LB, Lysogeny Broth; SEC, size exclusion chromatography; SEC-MALLS, size exclusion chromatography-multi angle laser light scattering; WT, wild-type.

References

[1] S. Kühner, V. van Noort, M.J. Betts, A. Leo-Macias, C. Batisse, M. Rode, et al., Proteome organization in a genome-reduced bacterium, Science. 326 (2009) 1235–1240. doi:10.1126/science.1176343.

[2] D.S. Goodsell, A.J. Olson, Structural symmetry and protein function, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 105–153. doi:10.1146/annurev.biophys.29.1.105.

[3] E.D. Levy, J.B. Pereira-Leal, C. Chothia, S.A. Teichmann, 3D complex: A structural classification of protein complexes, PLoS Comput. Biol. 2 (2006) 1395–1406. doi:10.1371/journal.pcbi.0020155.

[4] H. Nishi, K. Hashimoto, T. Madej, A.R. Panchenko, Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization, Prog. Mol. Biol. Transl. Sci. 117 (2013) 3–24. doi:10.1016/B978-0-12-386931-9.00001-5.

[5] P.C. Havugimana, G.T. Hart, T. Nepusz, H. Yang, A.L. Turinsky, Z. Li, et al., A census of human soluble protein complexes, Cell. 150 (2012) 1068–1081. doi:10.1016/j.cell.2012.08.011.

[6] K. Tarassov, V. Messier, C.R. Landry, S. Radinovic, M.M. Serna Molina, I. Shames, et al., An in vivo map of the yeast protein interactome, Science. 320 (2008) 1465–1470. doi:10.1126/science.1153878.

[7] C. V Robinson, A. Sali, W. Baumeister, The molecular sociology of the cell, Nature. 450 (2007) 973–982. doi:10.1038/nature06523.

[8] N.J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, et al., Global landscape of protein complexes in the yeast Saccharomyces cerevisiae, Nature. 440 (2006) 637–643. doi:10.1038/nature04670.

[9] C. Wan, B. Borgeson, S. Phanse, F. Tu, K. Drew, G. Clark, et al., Panorama of ancient metazoan macromolecular complexes, Nature. 525 (2015) 339–344. http://dx.doi.org/10.1038/nature14877.

[10] F.H. Crick, J.D. Watson, Structure of small viruses, Nature. 177 (1956) 473–475. doi:10.1038/177473a0.

[11] R.C.J. Dobson, K. Valegård, J.A. Gerrard, The crystal structure of three site-directed mutants of Escherichia coli dihydrodipicolinate synthase: Further evidence for a catalytic triad, J. Mol. Biol. 338 (2004) 329–339. doi:10.1016/j.jmb.2004.02.060.

[12] M.A. Navia, P.M. Fitzgerald, B.M. McKeever, C.T. Leu, J.C. Heimbach, W.K. Herber, et al., Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, Nature. 337 (1989) 615–620. doi:10.1038/337615a0.

[13] C.H. Heldin, Dimerization of cell surface receptors in signal transduction, Cell. 80 (1995) 213–223. doi:10.1016/0092-8674(95)90404-2.

[14] A.D. Malay, S.L. Procious, D.R. Tolan, The temperature dependence of activity and structure for the most prevalent mutant aldolase B associated with hereditary fructose intolerance, Arch. Biochem. Biophys. 408 (2002) 295–304. doi:http://dx.doi.org/10.1016/S0003-9861(02)00546-5.

[15] E.J. Loveridge, R.J. Rodriguez, R.S. Swanwick, R.K. Allemann, Effect of Dimerization on the Stability and Catalytic Activity of Dihydrofolate Reductase from the Hyperthermophile Thermotoga maritima, Biochemistry. 48 (2009) 5922–5933. doi:10.1021/bi900411a.

[16] J. Monod, J. Wyman, J.P. Changeux, On the Nature of Allosteric Transitions: a Plausible Model, J. Mol. Biol. 12 (1965) 88–118. doi:10.1016/S0022-2836(65)80285-6.

[17] T. Perica, J.A. Marsh, F.L. Sousa, E. Natan, L.J. Colwell, S.E. Ahnert, et al., The emergence of protein complexes: quaternary structure, dynamics and allostery. Colworth Medal Lecture, Biochem. Soc. Trans. 40 (2012) 475–91. doi:10.1042/BST20120056.

[18] K. Gunasekaran, B. Ma, R. Nussinov, Is allostery an intrinsic property of all dynamic proteins?, Proteins Struct. Funct. Genet. 57 (2004) 433–443. doi:10.1002/prot.20232.

[19] D.E. Koshland, G. Némethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry. 5 (1966) 365–385. doi:10.1021/bi00865a047.

[20] T. Perica, Y. Kondo, S.P. Tiwari, S.H. McLaughlin, K.R. Kemplen, X. Zhang, et al., Evolution of oligomeric state through allosteric pathways that mimic ligand binding, Science (80-.). 346 (2014) 1254346–1254346. doi:10.1126/science.1254346.

[21] J.A. Marsh, S.A. Teichmann, Structure, Dynamics, Assembly, and Evolution of Protein Complexes, Annu. Rev. Biochem. 84 (2014) 551–575. doi:10.1146/annurev-biochem-060614-034142.

[22] S.J. Fleishman, J.E. Corn, E.-M. Strauch, T.A. Whitehead, J. Karanicolas, D. Baker, Hotspot-Centric De Novo Design of Protein Binders, J. Mol. Biol. 413 (2011) 1047–1062. doi:10.1016/j.jmb.2011.09.001.

[23] J. Karanicolas, J.E. Corn, I. Chen, L. a. Joachimiak, O. Dym, S.H. Peck, et al., A De Novo Protein Binding Pair By Computational Design and Directed Evolution, Mol. Cell. 42 (2011) 250–260. doi:10.1016/j.molcel.2011.03.010.

[24] B. Kim, A. Eggel, S.S. Tarchevskaya, M. Vogel, H. Prinz, T.S. Jardetzky, Accelerated disassembly of IgE–receptor complexes by a disruptive macromolecular inhibitor, Nature. 491 (2012) 613–617. doi:10.1038/nature11546.

[25] H. Jubb, A.P. Higueruelo, A. Winter, T.L. Blundell, Structural biology and drug discovery for protein-protein interactions, Trends Pharmacol. Sci. 33 (2012) 241–248. doi:10.1016/j.tips.2012.03.006.

[26] T. Laue, B. Demeler, A postreductionist framework for protein biochemistry, Nat. Chem. Biol. 7 (2011) 331–334. doi:10.1038/nchembio.575.

[27] A. Gershenson, L.M. Gierasch, A. Pastore, S.E. Radford, Energy landscapes of functional proteins are inherently risky, Nat. Chem. Biol. 10 (2014) 884–891. doi:10.1038/nchembio.1670.

[28] L.N. Kinch, N. V Grishin, Evolution of protein structures and functions, Curr. Opin. Struct. Biol. 12 (2002) 400–408. doi:http://dx.doi.org/10.1016/S0959-440X(02)00338-X.

[29] E. Akiva, Z. Itzhaki, H. Margalit, Built-in loops allow versatility in domain-domain interactions: Lessons from self-interacting domains, Proc. Natl. Acad. Sci. 105 (2008) 13292–13297. doi:10.1073/pnas.0801207105.

[30] K. Hashimoto, A.R. Panchenko, Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states, Proc. Natl. Acad. Sci. 107 (2010) 20352–20357. doi:10.1073/pnas.1012999107.

[31] M.C. Mossing, R.T. Sauer, Stable, monomeric variants of lambda Cro obtained by insertion of a designed beta-hairpin sequence, Science. 250 (1990) 1712–1715. doi:10.1126/science.2148648.

[32] R.R. Dickason, D.P. Huston, Creation of a biologically active interleukin-5 monomer, Nature. 379 (1996) 652–655. doi:10.1038/379652a0.

[33] G. MacBeath, P. Kast, D. Hilvert, Probing enzyme quaternary structure by combinatorial mutagenesis and selection, Protein Sci. 7 (1998) 1757–1767. doi:10.1002/pro.5560070810.

[34] G. MacBeath, P. Kast, D. Hilvert, Redesigning enzyme topology by directed evolution, Science. 279 (1998) 1958–1961. doi:10.1126/science.279.5358.1958.

[35] Y.-T. Lai, D. Cascio, T.O. Yeates, Structure of a 16-nm Cage Designed by Using Protein Oligomers, Science (80-.). 336 (2012) 1129–1129. doi:10.1126/science.1219351.

[36] Y.-T. Lai, N.P. King, T.O. Yeates, Principles for designing ordered protein assemblies, Trends Cell Biol. 22 (2012) 653–61. doi:10.1016/j.tcb.2012.08.004.

[37] J.A. Marsh, S.A. Teichmann, Protein flexibility facilitates quaternary structure assembly and evolution, PLoS Biol. 12 (2014) e1001870. doi:10.1371/journal.pbio.1001870.

[38] J.A. Marsh, H.A. Rees, S.E. Ahnert, S.A. Teichmann, Structural and evolutionary versatility in protein complexes with uneven stoichiometry, Nat. Commun. 6 (2015) 6394. doi:10.1038/ncomms7394.

[39] X. Zhang, T. Perica, S.A. Teichmann, Evolution of protein structures and interactions from the perspective of residue contact networks, Curr. Opin. Struct. Biol. 23 (2013) 954–63. doi:10.1016/j.sbi.2013.07.004.

[40] T.L. Blundell, N. Srinivasan, Symmetry, stability, and dynamics of multidomain and multicomponent protein systems, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 14243–14248. doi:10.1073/pnas.93.25.14243.

[41] I. André, C.E.M. Strauss, D.B. Kaplan, P. Bradley, D. Baker, Emergence of symmetry in homooligomeric biological assemblies, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 16148–16152. doi:10.1073/pnas.0807576105.

[42] S.J. Fleishman, T.A. Whitehead, D.C. Ekiert, C. Dreyfus, J.E. Corn, E. Strauch, et al., Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin, Science (80-.). 332 (2011) 816–821. doi:10.1126/science.1202617.

[43] D. Grueninger, N. Treiber, M.O.P. Ziegler, J.W.A. Koetter, M.-S. Schulze, G.E. Schulz, Designed Protein-Protein Association, Sci. . 319 (2008) 206–209. doi:10.1126/science.1150421.

[44] J.D. Bloom, S.T. Labthavikul, C.R. Otey, F.H. Arnold, Protein stability promotes evolvability, Proc. Natl. Acad. Sci. 103 (2006) 5869–5874. doi:10.1073/pnas.0510098103.

[45] S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D.S. Tawfik, Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein, Nature. 444 (2006) 929–932. doi:10.1038/nature05385.

[46] T.A. Larsen, A.J. Olson, D.S. Goodsell, Morphology of protein-protein interfaces., Structure. 6 (1998) 421–427.

[47] H. Walden, G.S. Bell, R.J. Russell, B. Siebers, R. Hensel, G.L. Taylor, Tiny TIM: a small, tetrameric, hyperthermostable triosephosphate isomerase, J. Mol. Biol. 306 (2001) 745–57. doi:10.1006/jmbi.2000.4433.

[48] M. Robinson-Rechavi, A. Alibés, A. Godzik, Contribution of Electrostatic Interactions, Compactness and Quaternary Structure to Protein Thermostability: Lessons from Structural Genomics of Thermotoga maritima, J. Mol. Biol. 356 (2006) 547–557. doi:http://dx.doi.org/10.1016/j.jmb.2005.11.065.

[49] R. Thoma, M. Hennig, R. Sterner, K. Kirschner, Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from Thermotoga maritima, Structure. 8 (2000) 265–276. doi:http://dx.doi.org/10.1016/S0969-2126(00)00106-4.

[50] B. Kuhlman, J.W. O'Neill, D.E. Kim, K.Y. Zhang, D. Baker, Conversion of monomeric protein L to an obligate dimer by computational protein design, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10687–91. doi:10.1073/pnas.181354398.

[51] M. Kirsten Frank, F. Dyda, A. Dobrodumov, A.M. Gronenborn, Core mutations switch monomeric protein GB1 into an intertwined tetramer, Nat. Struct. Biol. 9 (2002) 877–885. doi:10.1038/nsb854.

[52] R.D. Newcomb, P.M. Campbell, D.L. Ollis, E. Cheah, R.J. Russell, J.G. Oakeshott, A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly, Proc Natl Acad Sci U S A. 94 (1997) 7464–7468.

[53] C.J. Jackson, J.-W. Liu, P.D. Carr, F. Younus, C. Coppin, T. Meirelles, et al., Structure and function of an insect α-carboxylesterase (αEsterase7) associated with insecticide resistance, Proc. Natl. Acad. Sci. U. S. A. (2013). doi:10.1073/pnas.1304097110.

[54] R.H. Brakenhoff, J.G. Schoenmakers, N.H. Lubsen, Chimeric cDNA clones: a novel PCR artifact., Nucleic Acids Res. 19 (1991) 1949.

[55] A.D. Keith, W. Snipes, Viscosity of cellular protoplasm., Science. 183 (1974) 666–668.

[56] P.J. Wyatt, Multiangle Light Scattering: The Basic Tool for Macromolecular Characterization, Instrum. Sci. Technol. 25 (1997) 1–18. doi:10.1080/10739149709351443.

[57] F.H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability., Nat. Protoc. 2 (2007) 2212–2221. doi:10.1038/nprot.2007.321.

[58] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state., J. Mol. Biol. 372 (2007) 774–797. doi:10.1016/j.jmb.2007.05.022.

[59] L. Lo Conte, C. Chothia, J. Janin, The atomic structure of protein-protein recognition sites, J. Mol. Biol. 285 (1999) 2177–2198. doi:10.1006/jmbi.1998.2439.

[60] Z.S. Derewenda, P.G. Vekilov, Entropy and surface engineering in protein crystallization, Acta Crystallogr. Sect. D Biol. Crystallogr. 62 (2006) 116–124. doi:10.1107/S0907444905035237.

[61] B. Meibohm, H. Zhou, Characterizing the impact of renal impairment on the clinical pharmacology of biologics, J. Clin. Pharmacol. 52 (2012) 54S–62S. doi:10.1177/0091270011413894.

[62] C. Mitchinson, J.A. Wells, Protein engineering of disulfide bonds in subtilisin BPN'., Biochemistry. 28 (1989) 4807–4815. doi:10.1021/bi00437a043.

[63] L. Giver, A. Gershenson, P.O. Freskgard, F.H. Arnold, Directed evolution of a thermostable esterase., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 12809–12813.

[64] S. Bershtein, W. Mu, E.I. Shakhnovich, Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations, Proc. Natl. Acad. Sci. . 109 (2012) 4857–4862. doi:10.1073/pnas.1118157109.

[65] B.M. Beadle, B.K. Shoichet, Structural bases of stability-function tradeoffs in enzymes., J. Mol. Biol. 321 (2002) 285–296.

[66] S.R. Devenish, J.A. Gerrard, The role of quaternary structure in (beta/alpha)(8)-barrel proteins: evolutionary happenstance or a higher level of structure-function relationships?, Org. Biomol. Chem. 7 (2009) 833–9. doi:10.1039/b818251p.

[67] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. Methods. 6 (2009) 343–345. doi:10.1038/nmeth.1318.

[68] C. Neylon, S.E. Brown, A. V Kralicek, C.S. Miles, C.A. Love, N.E. Dixon, Interaction of the Escherichia coli Replication Terminator Protein (Tus) with DNA: A Model Derived from DNA-Binding Studies of Mutant Proteins by Surface Plasmon Resonance, Biochemistry. 39 (2000) 11989–11999. doi:10.1021/bi001174w.

[69] T.M. Laue, B.D. Shah, T.M. Ridgeway, S.L. Pelletier, Computer-Aided Interpretation of Analytical Sedimentation Data for Proteins, in: S.E. Harding, A.J. Rowe, J.C. Horton (Eds.), Anal. Ultracentrifugation Biochem. Polym. Sci., Royal Society of Chemistry, Cambridge, U.K., 1992: pp. 90–125.

[70] M.A. Perugini, P. Schuck, G.J. Howlett, Self-association of Human Apolipoprotein E3 and E4 in the Presence and Absence of Phospholipid, J. Biol. Chem. 275 (2000) 36758–36765. doi:10.1074/jbc.M005565200.

[71] P. Schuck, Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling, Biophys. J. 78 (2000) 1606–1619. doi:10.1016/S0006-3495(00)76713-0.

[72] P. Schuck, M.A. Perugini, N.R. Gonzales, G.J. Howlett, D. Schubert, Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems, Biophys. J. 82 (2002) 1096–1111. doi:10.1016/S0006-3495(02)75469-6.

[73] M. Niklasson, C. Andresen, S. Helander, M.G.L. Roth, A. Zimdahl Kahlin, M. Lindqvist Appell, et al., Robust and convenient analysis of protein thermal and chemical stability, Protein Sci. 24 (2015) 2055–2062. doi:10.1002/pro.2809.

TABLES

	Round of evolution and mutation						
	2a	2b	2c	2d	3a	3b	4a
Residue							
Asp83							Ala
Ala285			Ser				
Met364							Leu
Ile419				Phe	Phe		Phe
Ala472		Thr	Thr	Thr	Thr	Thr	Thr
Phe478	Leu						
Ile505						Thr	Thr
Lys530		Glu				Glu	Glu
Asp554							Gly

Table 1: Mutations incorporated into *Lc*αE7 during laboratory-directed evolution for thermostability. Mutations were not recorded in the first round of mutagenesis. In Round 2 four variants (2a, 2b, 2c and 2d) were selected for DNA shuffling and mutagenesis resulting in Round 3 variants. In Round 3 two variants (3a and 3b) were selected for DNA shuffling and mutagenesis resulting in the Round 4a variant.

	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)	K_i (mM)
*Lc*αE7	68.6 ±12.3	40.6 ± 2.2	$6.0 \times 10^5 \pm 1.3 \times 10^5$	-
*Lc*αE7-4 Monomer	38.1 ± 8.3	14.5 ± 1.1	$3.8 \times 10^5 \pm 1.6 \times 10^5$	2.6 ± 0.8
*Lc*αE7-4 Dimer	30.4 ±4.8	4.5 ± 0.2	$1.5 \times 10^5 \pm 0.2 \times 10^5$	-
*Lc*αE7-4 Tetramer	35.6 ± 4.9	5.7 ± 0.3	$1.6 \times 10^5 \pm 0.3 \times 10^5$	-

Table 2. Kinetic parameters for variants of *Lc*αE7 with the model carboxylesterase substrate 4-nitrophenyl butyrate. Kinetic data for *Lc*αE7, *Lc*αE7-4 Dimer and *Lc*αE7-4 Tetramer were fit to a model for standard Michaelis-Menten kinetics ($v=(k_{cat}.[S])/(K_M + [S])$), whereas kinetic data for *Lc*αE7-4 Monomer were fit to a Michaelis-Menten kinetic model incorporating substrate inhibition ($v=(k_{cat}.[S])/((K_M + [S]).(1 + [S]/K_{si}))$).

Protein	TT_{50} (° C)
LcαE7 Monomer	48.8 ± 0.2
LcαE7-4a Monomer	55.5 ± 0.5, 66.2 ± 0.6
LcαE7-4a Dimer	56.3 ± 0.6, 61.7 ± 0.9
LcαE7-4a Tetramer	65.2 ± 2.0

Table 3. Calculated Transition Temperature (TT50) for the LcαE7 proteins as measured by temperature-ramp circular dichroism. Ellipticity was recorded at 208 nm and fitted by non-linear regression to two-state model (LcαE7 monomer and LcαE7-4a tetramer) and a three-state model (LcαE7-4a monomer and LcαE7-4a tetramer. The best fitting model was chosen based on the sum of residuals. Raw data is plotted in SI Figure 2.

Protein	TT_{50} (° C)
LcαE7-4a	55.5 ± 0.5
LcαE7-4a A83D	55.7 ± 0.4
LcαE7-4a L364M	56.5 ± 0.6
LcαE7-4a E530K	53.9 ± 0.2
LcαE7-4a D554G	56.3 ± 0.8

Table 4. Calculated Transition Temperature (TT_{50}) for the *Lc*αE7-4a proteins variants with reversed surface mutations for thermostability as measured by temperature-ramp circular dichroism. Spectra were recorded at 208 nm and fitted to a non-linear regression to a three-state model (equation 1).

FIGURES

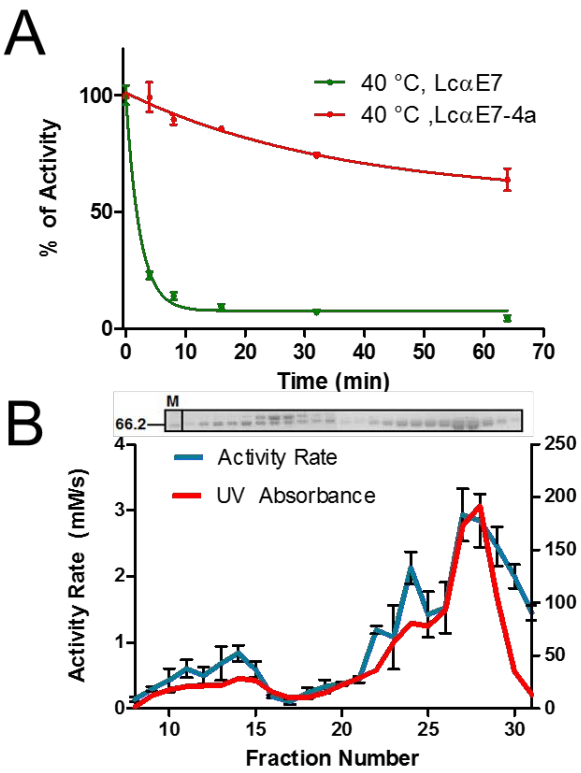


Figure 1. Directed evolution of *Lc*αE7 for enhanced thermostability. (A) When incubated at 40 °C over a time-course, there was a one-phase exponential decay curve of activity in the wild-type protein, and of the *Lc*αE7-4a protein (B) Size exclusion chromatography detected the presence of *Lc*αE7 in four main peaks, fraction 12, fraction 22, fraction 24 and fraction 28. A contaminating band was present, represented by a peak in fraction 14.

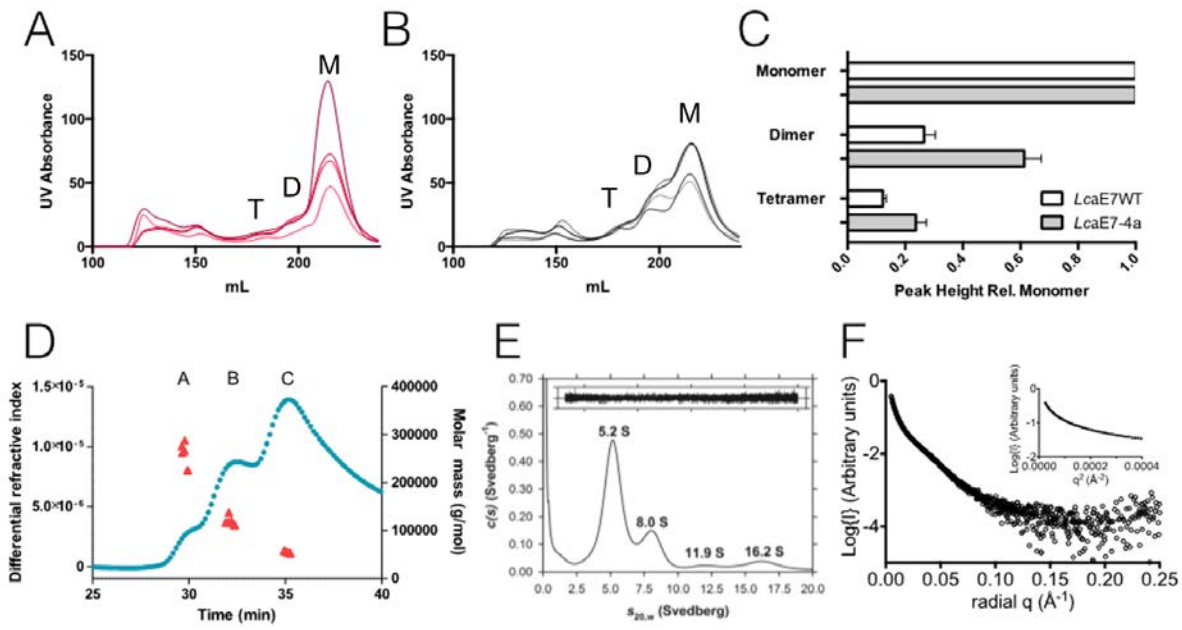


Figure 2. Characterization of the oligomeric species observed in this work. (A, B) Comparison between size exclusion chromatographic spectra of wild-type *Lc*αE7 (A) and *Lc*αE7-4a (B) reveals that directed evolution has resulted in substantial increases in the amounts of dimeric and tetrameric species (C). Monomer (M), dimer (D) and tetramer (T) peaks are labeled. Each line in the graphs (A,B) correspond to an independent purification. (D) Size exclusion chromatography/multi-angle laser light scatter (SEC-MALLS) analysis reveals the three species of *Lc*αE7-4a correspond to monomer, dimer and tetramer stoichiometry. (E) Analytical ultracentrifugation (AUC) sedimentation velocity analysis of *Lc*αE7-4a at an initial concentration of 3.0 mg/mL. The continuous sedimentation coefficient [*c(s)*] distribution is plotted as a function of the standardized sedimentation coefficient. Analysis was performed using a resolution of 200 with a sedimentation coefficient range of 0-20 S and a P-value of 0.95. The nonlinear least squares fit shown resulted in an rmsd of 0.00585 and a runs test Z of 4.61. Inset - residuals from the *c(s)* distribution best fit plotted as a function of radial distance from the axis of rotation. (F) Small angle X-ray scattering (SAXS) analysis of the high molecular weight species reveals it to be a heterogeneous aggregate.

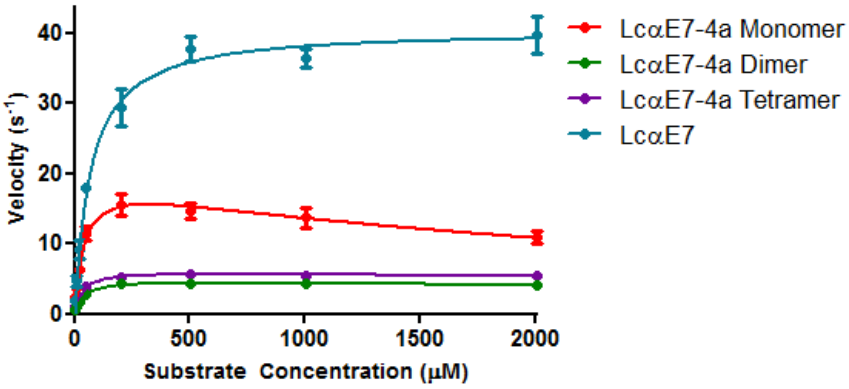


Figure 3. Characterization of the oligomeric species. Rate *vs.* concentration plots for the catalysis of 4-nitrophenyl butyrate hydrolysis by wild-type *Lc*αE7 and *Lc*αE7-4a. All curves were fit to the Michaelis-Menton equation $v=(k_{cat}.[S])/(K_M + [S])$, with the exception of *Lc*αE7-4a monomer, which was fitted to the equation modified for substrate inhibition $v=(k_{cat}.[S])/((K_M + [S]).(1 + [S]/K_{si}))$.

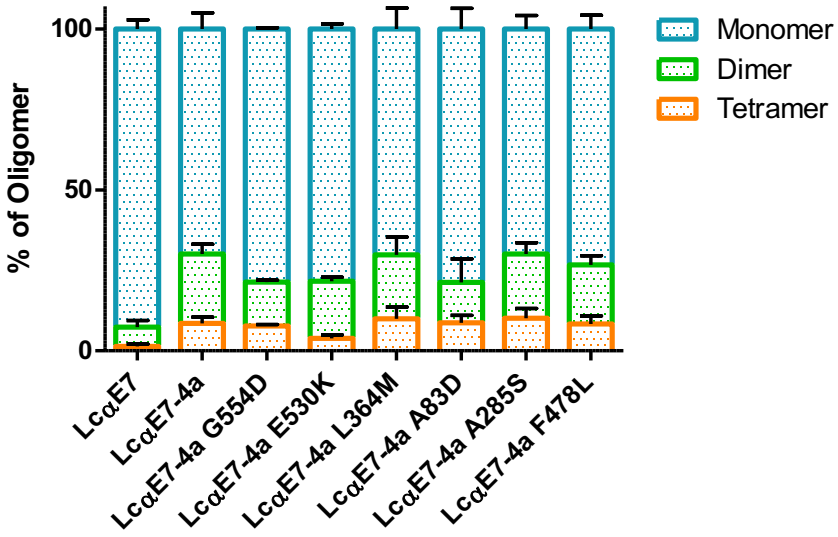


Figure 4. Proportion of oligomeric species in the reverse mutations *Lc*αE7-4a (A83D *Lc*αE7-4a, L364M *Lc*αE7-4a, E530K *Lc*αE7-4a, G554D *Lc*αE7-4a) and two mutations found in round two of the directed evolution (A285S *Lc*αE7-4a and F478L *Lc*αE7-4a) as estimated by peak heights obtained by size exclusion chromatography. Standard errors are shown for each species.

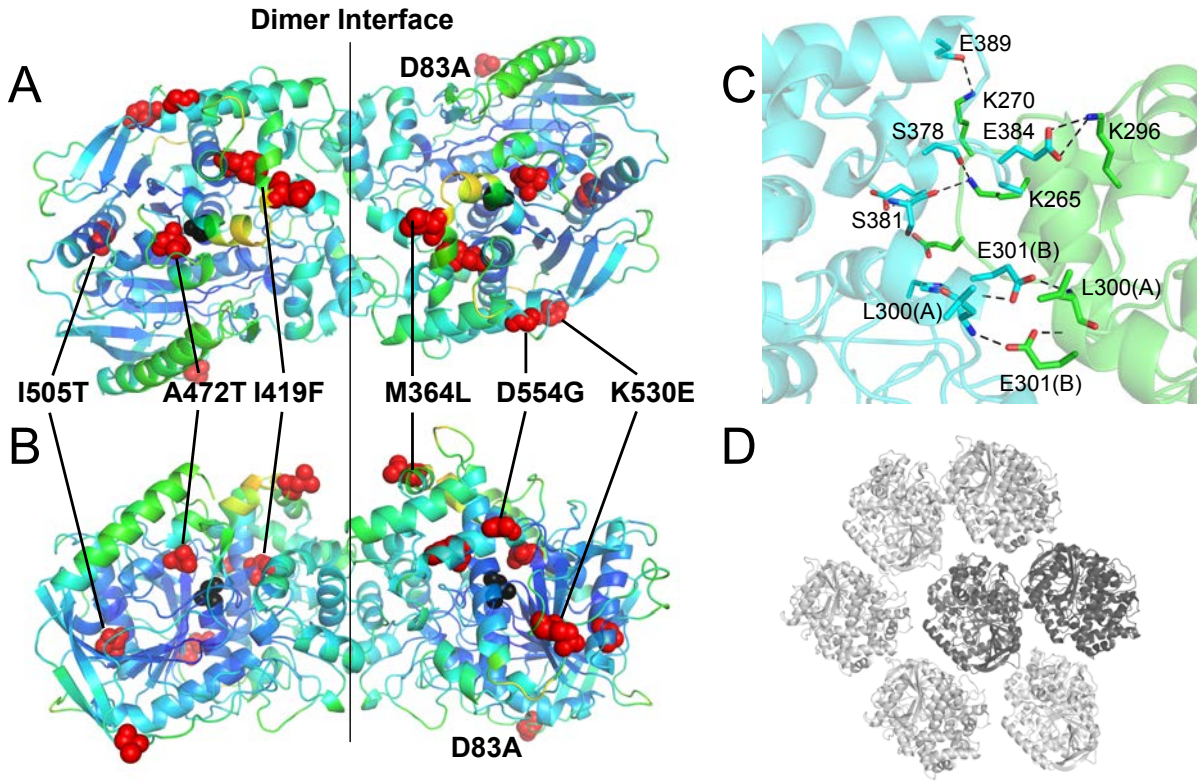
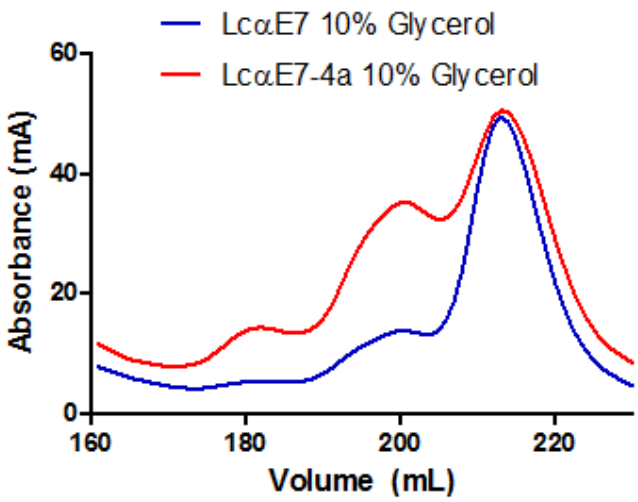
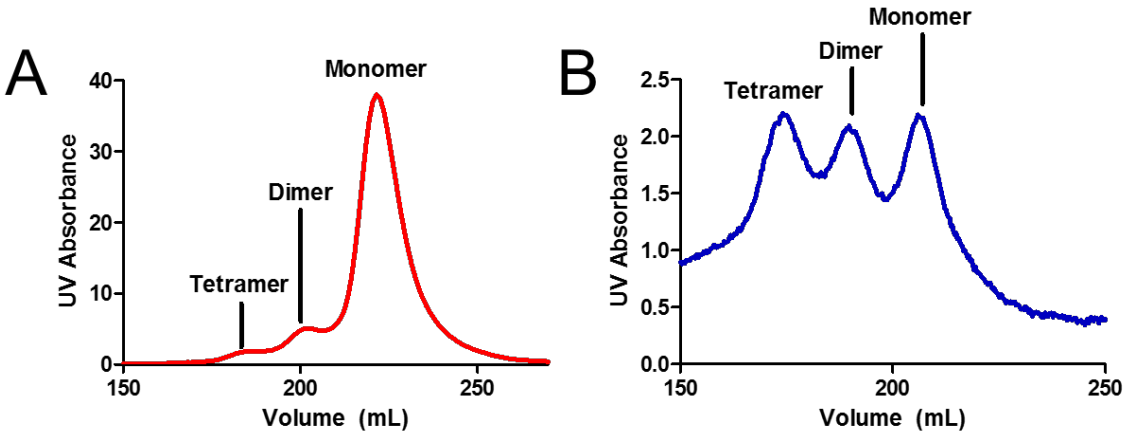


Figure 5. (A and B) The structure of dimeric *Lc*αE7-4a, showing the location of the dimer interface and the location of the mutations. The surface mutations D83A, M364L, K530E and D554G are shown on the right, while the three internal mutations I505T, A472T and I419F are shown to the left. The catalytic serine (S218) is shown in black to indicate the location of the active site. B is rotated 90° relative to A**.** (C) A number of salt bridges and hydrogen bonds are formed at the dimer interface. Monomer A is shown in cyan and monomer B in green. (D) The comparatively loose crystal packing interactions of the $P2_1$ crystal form of *Lc*αE7-4a. The two molecules in the asymmetric unit are colored dark grey, symmetry mates light grey. Other symmetry-related molecules are omitted for clarity.

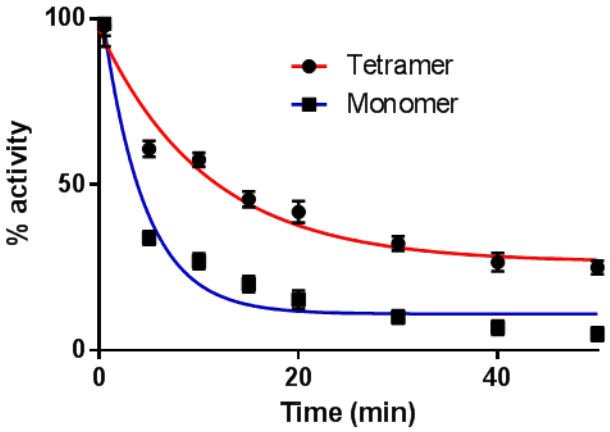
SUPPLEMENTARY INFORMATION



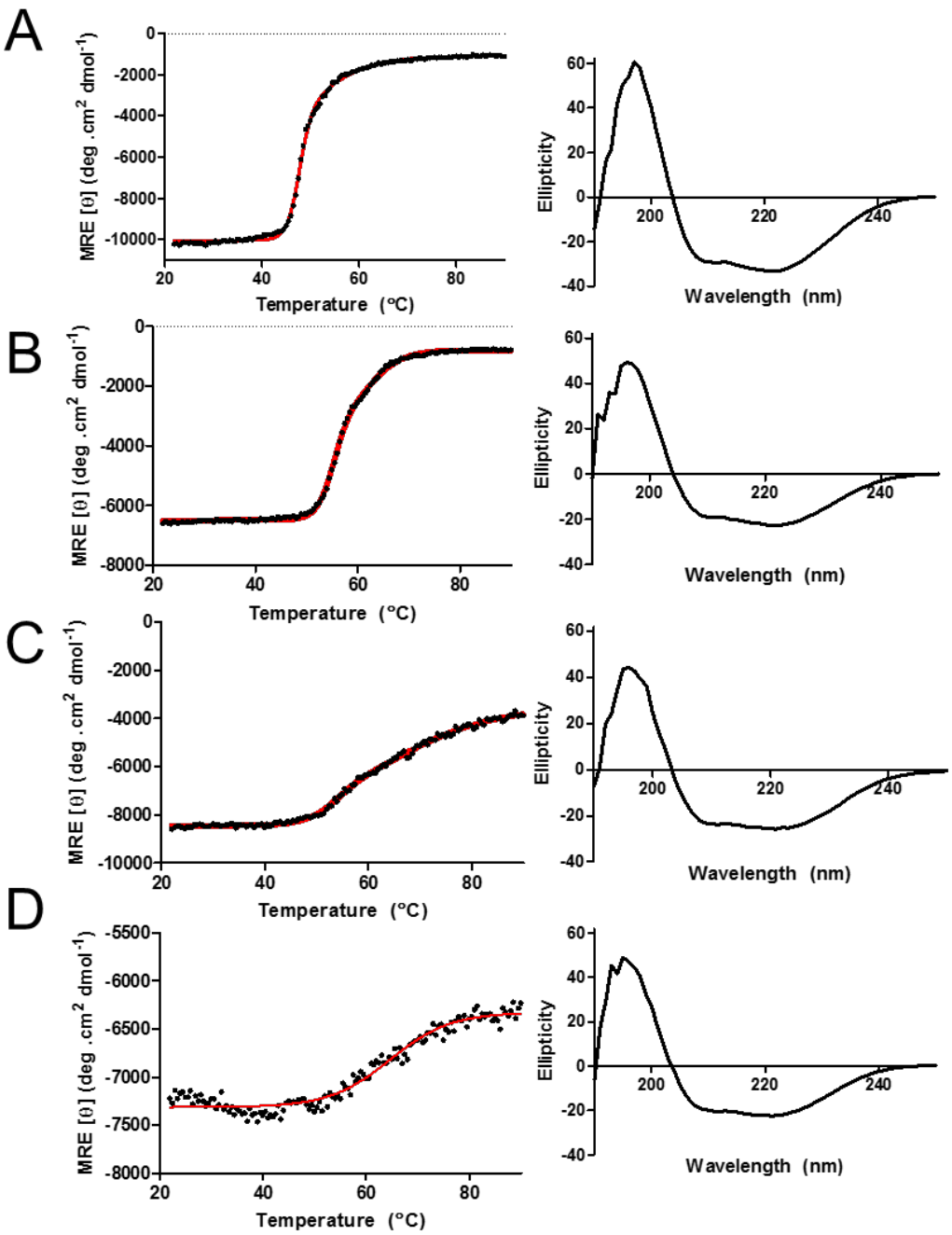
SI Figure 1. The effect of viscosity on oligomerization. Size exclusion chromatograms of *Lc*αE7-4a and *Lc*αE7-4a WT in the presence of 10% glycerol shows that for *Lc*αE7-4a, the tetrameric and dimeric fractions (peak heights 15 mAU and 35 AU) comprise approximately 50% of the total protein (monomer peak height 51 mAU). In contrast, for *Lc*αE7-4a WT, the higher molecular weight species (tetramer and dimer peak heights 6 mAU and 15 mAU *vs.* 50 for monomer) comprise less than 30% of the total protein.



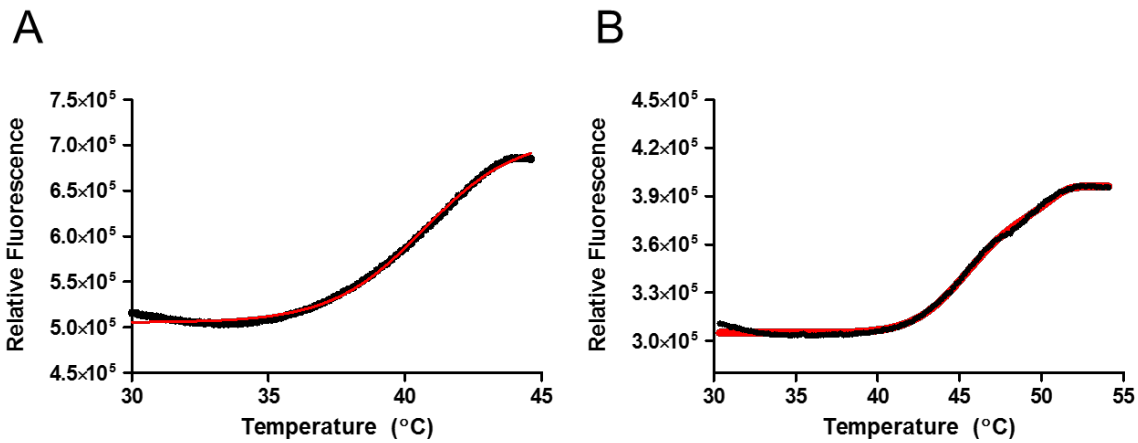
SI Figure 2. (A) Re-equilibration of monomeric fraction (single 5 mL fraction from middle of monomeric peak) into dimeric and tetrameric peaks after 5 hours at 4 °C. (B) Re-equilibration of tetrameric fraction (single 5 mL fraction from middle of tetrameric peak) into dimeric and monomeric species after 24 hours at 4 °C.



SI Figure 3. Activity decay over time at 46 °C for monomeric and tetrameric species of *Lc*αE7-4a. Aliquots were taken over a series of time points when tetrameric and monomeric species were incubated at 46 °C. Specific activity was measured using 4-nitrophenyl butyrate, 200mM NaCl, 20mM HEPES pH 7.5.



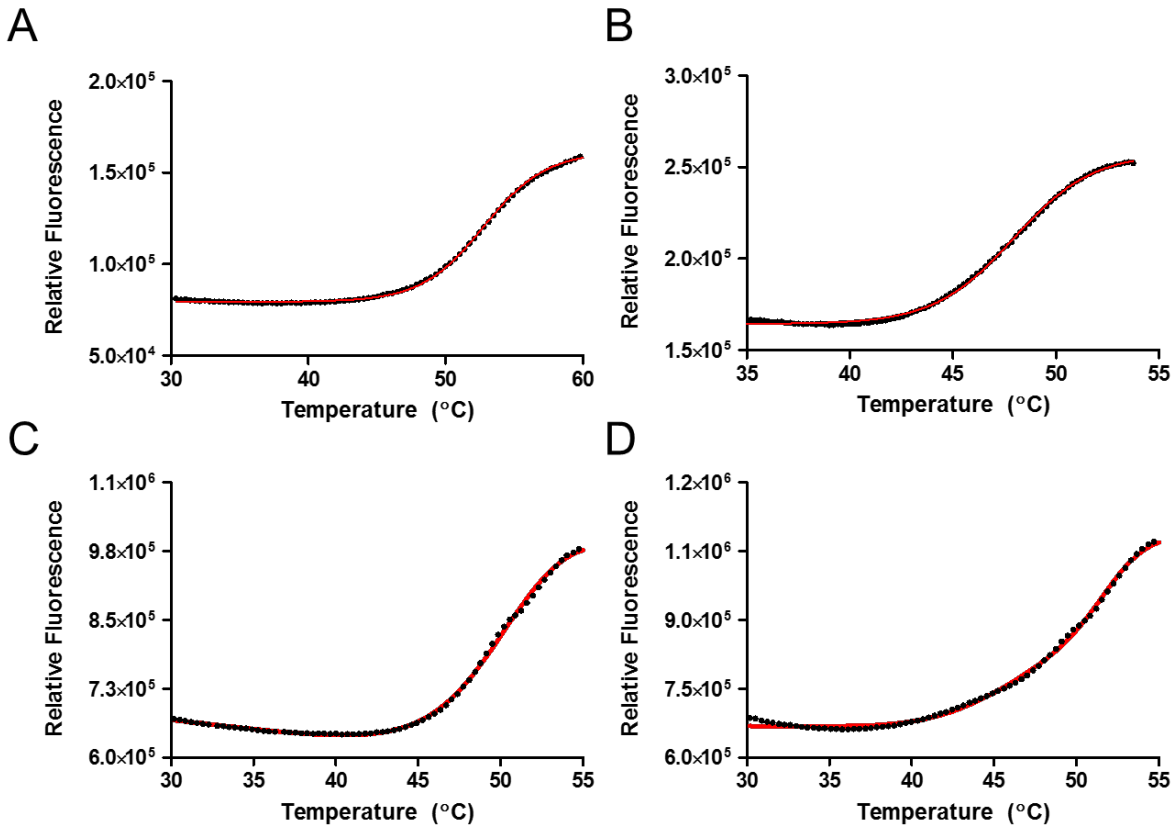
SI Figure 4. Thermally induced unfolding of *Lc*αE7 and the *Lc*αE7-4a oligomeric species. Ellipticity as a function of temperature was monitored at 208 nm and curves were fit by non-linear regression with equation 1. For the *Lc*αE7-4a monomer and dimer data, a three-state model was used to fit the data. Spectra of the protein before denaturation are shown next to the denaturation curves. (A) *Lc*αE7 (B) *Lc*αE7-4a Monomer (C) *Lc*αE7-4a Dimer (D) *Lc*αE7-4a Tetramer.



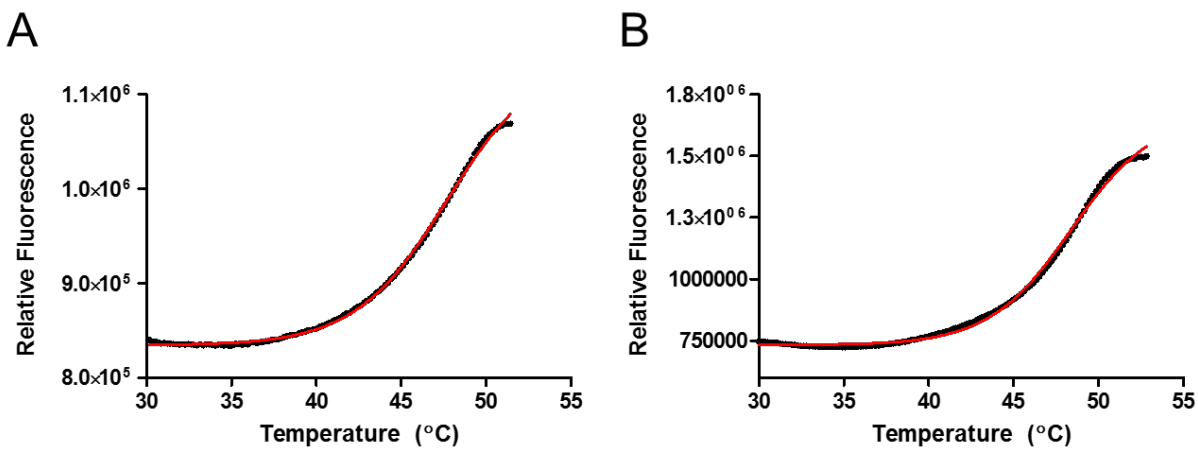
SI Figure 5. Differential scanning fluorimetry (DSF) analysis of pre-SEC samples of *Lc*αE7 (2 mg/mL). (A) Wild-type *Lc*αE7 undergoes a single transition, with an apparent melting point of 40.6 ± 0.6 °C, whereas *Lc*αE7-4a is comprised of one species with an apparent melting point of 44.8 ± 0.5 °C and a second species with an apparent melting point of 50.9 ± 0.9 °C.

Protein	Concentration	TT_{50} (° C)
LcαE7-4a Monomer	0.3 mg/mL	47.92 ± 0.2
	3mg/mL	46.0 ± 0.9, 52.5 ± 0.3
LcαE7-4a Tetramer	0.3 mg/mL	52.8 ± 0.3
	3mg/mL	48.0 ± 0.4, 53.4 ± 0.7

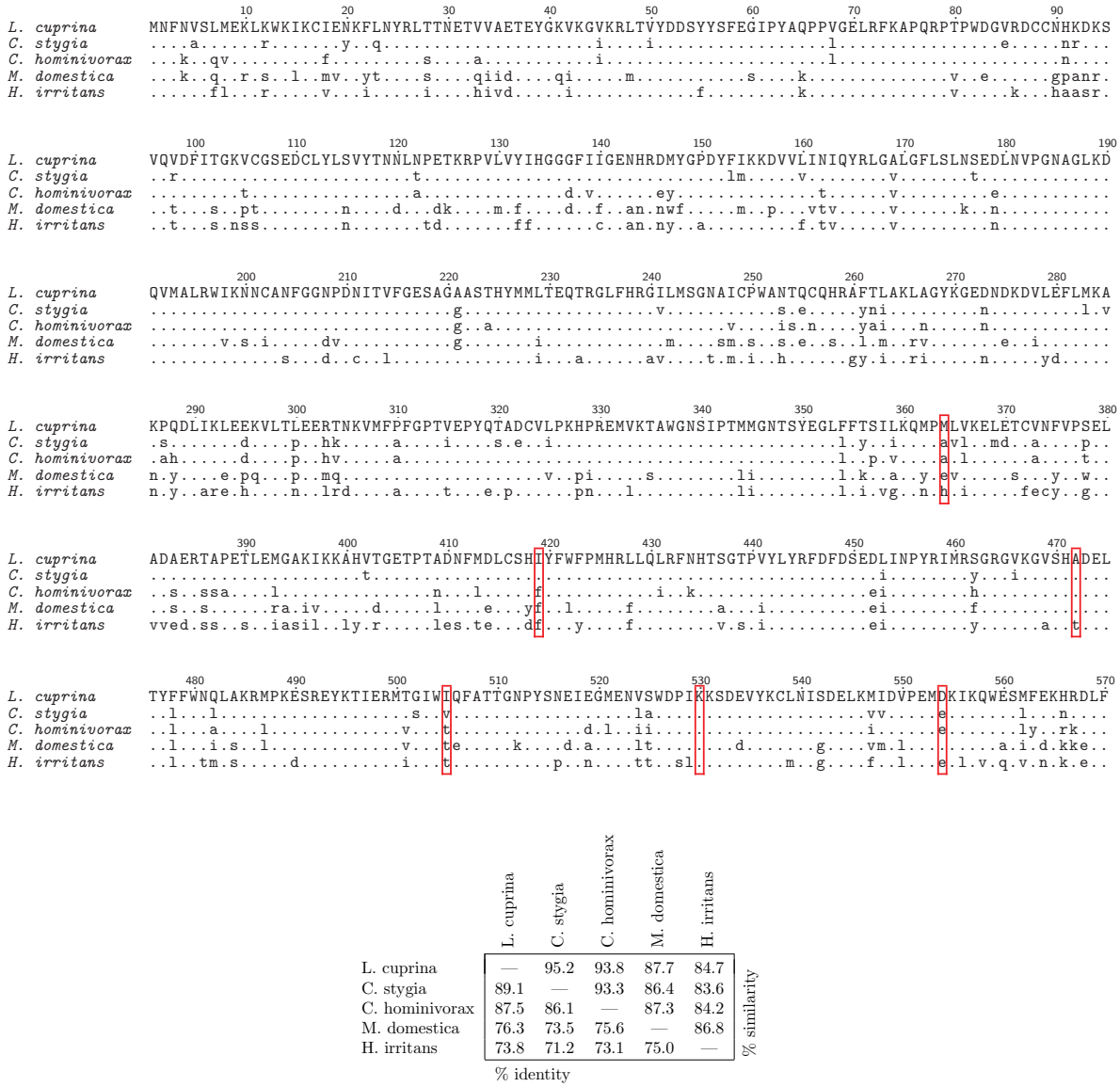
SI Table 1. Calculated Transition Temperature (TT50) for the LcαE7 proteins as measured by DSF. The 3 mg/mL sampled were concentrated prior to analysis.



SI Figure 6. Differential scanning fluorimetry (DSF) analysis of monomeric and tetrameric fractions of *Lc*αE7-4a. (A) Tetrameric fraction of *Lc*αE7-4a at 0.3 mg/mL shows a single apparent melting point at 52.8 °C, whereas monomeric fraction of *Lc*αE7-4a at 0.3 mg/mL shows a single transition at 47.9 °C. After concentration to 3 mg/mL, some re-equilibration occurred and both tetrameric and monomeric samples showed two transitions, one corresponding to monomer (~47 °C) and the other corresponding to tetramer (~53 °C).



SI Figure 7. Thermostability effects of the A285S and F476L mutations. Both mutations were introduced in to *Lc*αE7-4a. Thermostability was estimated by DSF. (A) Monomeric fraction of *Lc*αE7-4a(A285S) at 0.3 mg/mL shows a single apparent melting point at 47.6 ± 1.3 °C, within error of *Lc*αE7-4a (47.9 ± 0.2 °C). (A) Monomeric fraction of *Lc*αE7-4a(F478L) at 0.3 mg/mL shows a two apparent melting points: a major species at 48.2 ± 0.9 °C, within error of *Lc*αE7-4a (47.9 ± 0.2 °C), and a minor species at 51.2 ± 0.2 °C, which was not higher than the higher-order species in *Lc*αE7-4a (52.5 ± 0.3 °C). Thus, neither A285S nor F476L are both likely to be selectively neutral in terms of thermostability to *Lc*αE7.



SI Figure 8. Multiple sequence alignment of *Lc*αE7 with αE7 genes from other fly species (*Calliphora stygia*, *Cochliomyia hominivorax*, *Musca domestica* and *Haematobia irritans*) reveals significant variation at the positions that were mutated during the directed evolution experiment. Specifically, the I419F, A472T, and I505T sequence differences already exist across the five species compared.