æ

14. PROTEIN MANAGEMENT

26 November 2022

Some of the more fundamental kinetic and structural features of proteins were reviewed in the two preceding chapters. We now move on to additional issues relevant to the life histories of proteins, most notably matters associated with folding assistance, post-translational modifications, and protein disposal through degradation. Whereas much attention has been given to transcriptional control of gene expression (Chapter 21), these three processes are also central dimensions to protein management.

As noted in Chapter 12, many small proteins are capable of folding on their own without any external physical assistance. Such proteins must be endowed with amino-acid sequences carrying all of the "information" essential to acquiring proper three-dimensional structures. However, in all organisms, numerous proteins (particularly large ones) require some form of folding assistance from helper proteins called chaperones. Even in the presence of chaperones, some proteins fail to ever achieve their native states, and these must be disposed of to avoid misinteractions with other proteins and potential cellular malfunctions. In addition, some proteins, such as those involved in the cell cycle, need to be conditionally active and efficiently eliminated after completing their missions. Such selective protein removal often relies upon particular markings directing their delivery to the cellular degradation machinery. Still other post-translational markings on proteins confer particular subcellular functions, e.g., in signal-transduction pathways (Chapter 22).

In addition to outlining general aspects of protein management, this chapter also provides numerous examples that reinforce the principles regarding the evolutionary aspects of cellular features outlined in preceding and subsequent chapters. For example, much of the machinery associated with protein-folding assistance and selective degradation consists of higher-order multimers that have frequently changed with respect to subunit number and type (Chapter 13). The coevolution of chaperones and their client-gene products raise issues of how a cellular feature with multiple substrates might become constrained by a "jack of all trades, master of none" syndrome (Chapter 21). Finally, the sites of post-translational markings appear to be nearly free to wander evolutionarily over protein surfaces, providing a means for the effectively neutral re-wiring of regulatory mechanisms (Chapters 6, 10, and 21).

Chaperone Assistance

A common mechanism by which proper folding of proteins is achieved involves chaperone provisioning of a protective environment for confining and restricting the ways

in which captured protein molecules can move. This lowers the energetic barrier necessary to achieve a stable folding configuration by a newly formed protein, while also minimizing the potential for harmful interactions of misfolded proteins with others in the cell.

The widespread use of molecular chaperones across the Tree of Life motivates numerous evolutionary questions. First, what are the mechanisms by which chaperones recognize their cognate client proteins? Second, do certain classes of chaperones coevolve with individual client proteins in ways that make them less effective with other potential clients? Third, once a protein becomes reliant on chaperone assistance for proper folding, does this act as an evolutionary trap by further relaxing selection on features essential to unassisted self-folding, thereby facilitating the effectively neutral accumulation of otherwise deleterious mutations? Fourth, does chaperone dependence facilitate the evolution of adaptations that would otherwise not be possible because of their negative effects on self-folding? Fifth, given that chaperones themselves consume ATP in the folding process, what is the energetic cost to the cell of producing and relying upon chaperones?

Phylogenetic diversity of chaperones. Because orthologs of some chaperones are found in all three domains of life, they were likely present in LUCA (Rebeaud et al. 2021), paving the way to the establishment of proteins too large to self-fold. However, the substantial diversity of chaperone types within organismal lineages also leads to the conclusion that these helper molecules originated more than once, often converging on similar molecular structures and mechanisms (Schilke et al. 2006; Stirling et al. 2006). Further functional diversification of chaperones followed gene duplication and sub/neofunctionalization on multiple occasions in eukaryotes (Abascal et al. 2013; Carretero-Paulet et al. 2013) and prokaryotes (Bittner et al. 2007; Wang et al. 2013; Weissenbach et al. 2017), possibly driven by adaptive conflicts imposed by alternative client protein pools. Moreover, as will be outlined below, striking examples exist of evolutionary transitions between homomeric and heteromeric chaperone structures.

The following paragraphs provide a brief overview of chaperone diversity in the bacterial, archaeal, and eukaryotic domains, although only the most well-studied chaperone families will be introduced here. Unfortunately, even with such a distillation, chaperone-family nomenclature is difficult to navigate, as the labeling of orthologous genes is often inconsistent among organismal lineages. To avoid this notational morass, an attempt is made below to simplify discussion via a slight abuse of taxon-specific notation. Many chaperones are referred to as heat-shock (or heat-stress) proteins, owing to their induced overexpression at extreme temperatures, and such labels are often post-scripted by a number referring to the approximate size in kiloDaltons (a measure of mass, with one kDa $\simeq 7.5$ amino acids), a notation that will be adhered to in a number of cases below. Adding to the complexity of classification, not all heat-shock proteins are exclusively involved in protein folding, with some being more associated with protein degradation and/or disaggregation.

Within the bacteria, there are three major classes of chaperones: 1) Trigger Factor; 2) a consortium of Hsp40, Hsp70, and a Nucleotide Exchange Factor (NEF); and 3) GroEL/GroES. All three classes have divergent molecular architectures and are deployed in substantially different ways. Trigger Factor is a monomeric protein

that binds to nascent peptides as they emerge from the ribosome, effectively producing a preliminary folding space without requiring ATP for function. Hsp40 (a tweezers-like dimeric protein) acts as a cochaperone, binding exposed hydrophobic patches on unfolded proteins and targeting them to Hsp70 (a monomeric protein), which stabilizes the client protein in an ATP-dependent manner. Hsp70 can also operate as an "unfoldase," consuming ~ 5 ATPs per protein in the process, and has many other housekeeping roles, including guidance in delivering proteins to their organelle destinations and uncoating of endocytic-membrane vesicles in eukaryotes (Sharma et al. 2010; Rosenzweig et al. 2019). NEF plays a regulatory role in these processes. Although this Hsp40/Hsp70/NEF system is widespread among bacteria, at least one lineage appears to have lost it (Warnecke 2012).

The best studied bacterial chaperone is GroEL (more generally known as chaperonin 60, with the name GroEL being used for the *E. coli* protein). GroEL has a large barrel-like structure, consisting of two heptameric rings stacked back to back (with all 14 subunits encoded by the same gene) (Figure 14.1). Each ring comprises a separate chamber within which the folding of individual client proteins proceeds after closure of the GroEL cavity by a cochaperone lid (heptameric GroES). Substrate proteins are captured via interactions with their hydrophobic residues and then stretched and remodeled within the folding cage. The mechanics of GroEL/ES involve a form of allostery, with cycles of enclosure and release – the binding of ATPs to one ring results in the release of the GroES cap from the other. Each round of turnover of a protein requires ~ 11 seconds and consumes 7 ATPs, one ATP for each of the subunits of the ring (Keskin et al. 2002; Ueno et al. 2004; Horwich et al. 2009). However, in *E. coli,* the half-time for completion of assisted folding is ~ 45 sec (Kerner et al. 2005), suggesting that an average client protein engages in $45/11 \simeq 4$ GroEL-assisted folding attempts before success is achieved, and consuming ~ 28 ATPs in doing so. Although some *E. coli* proteins require an average of ~ 40 cycles (~ 280 ATPs) to achieve proper folding (Santra et al. 2017), this is still a relatively small price to pay, as the biosynthetic cost of a single amino acid is ~ 30 ATPs (Chapter 17), and proteins typically consist of many dozens of proteins.

GroEL is present throughout the entire bacterial phylogeny, with some species harboring multiple variants that are likely subfunctionalized with respect to client proteins (Lund 2009; Henderson et al. 2013). However, a few bacterial species (e.g., *Mycoplasma* and *Ureaplasma;* Wong and Houry 2004) seem to have lost GroEL. Random mutagenesis studies indicate that *Ureaplasma* proteins are just one or two mutations removed from GroEL dependence (Ishimoto et al. 2014), further evidence for the point made in Chapter 12 that proteins commonly evolve to be just beyond the margin of stability. There is also some evidence that not all bacterial GroELs follow the *E. coli* model of oligomeric structure, with dimeric or tetrameric structures likely present in some taxa. It is difficult to see how such reduced structures could serve as chaperones, and they may have entirely different functions, as GroEL is known to engage in different activities in a number of species, e.g., adhesion to host cells, secretion, DNA binding, cell-cell communication, and even toxicity (Henderson et al. 2013).

Like bacteria, many archaea deploy chaperones in the Hsp40/Hsp70/NEF group, suggesting that this particular family dates back to LUCA. However, Hsp40 and Hsp70 are apparently absent from the most thermophilic archaea, which is surprising

given the negative effects of high temperature on folding stability. Those archaeal species containing Hsp40 and Hsp70 appear to have acquired them by lateral transfer from bacterial lineages (Macario et al. 2006). Archaea do not harbor Trigger Factor (Laksanalamai et al. 2004), although there is an apparently unrelated mechanism for stabilizing nascent proteins emerging from ribosomes (Spreter 2005).

Although GroEL/GroES is absent from archaea except in rare cases of horizontal transfer (Hirtreiter et al. 2009), there is a chaperonin (CCT, also known as the thermosome) with substantial structural similarity (Foundations 14.1). As with bacterial GroEL, the archaeal CCT forms a double-ringed barrel structure, but instead of there being a separate GroES-like cap, each monomeric subunit contains a built-in apical loop. These apical lids close like a camera iris, leaving a small opening. This may enable proteins too large to enter the chamber in their entirety to experience progressive folding by threading (Rüßmann et al. 2012). Despite the similarity of the double-barrel architecture of CCT to the form of GroEL, it remains unclear whether the two are derived from a common ancestor, as there is only $\sim 20\%$ sequence similarity. In addition, the GroEL ring contains seven subunits, whereas CCT contains eight or nine (Archibald et al. 1999). Finally, archaeal prefoldin is a heterohexameric cochaperone, comprised of two monomeric subunit types, that serves to transfer proteins to CCT.

Eukaryotes deploy several chaperones to assist in protein folding, the major ones being: a ribosome-associated complex, consisting of Hsp40 and Hsp70 partners; Hsp90 (a dimer involved in both folding and aggregation suppression); prefoldin; and CCT. The eukaryotic prefoldin hexamer consists of six subunit types, as opposed to two in archaea, and all of the monomeric subunits of eukaryotic CCT are also encoded by different genes (Foundations 14.1). Moreover, unlike the situation with bacterial GroEL, where proteins are processed in a chamber with hydrophilic walls, CCT folding assistance involves binding of the substrate to the apical domains of the internal chamber. Eukaryotic organelles (mitochondria and chloroplasts) utilize bacterial-derived orthologs of GroEL and GroES, called Hsp60 and Hsp10 respectively, but unlike bacterial GroEL, mitochondrial Hsp60 operates as a single rather than a double ring.

As in bacteria, the eukaryotic Hsp70 proteins are monomers, containing one domain for protein binding and another for ATPase activity. Hsp70 has commonly diversified into a dozen or more copies in various eukaryotes, and Hsp40 even more so (Craig and Marszalek 2011; Bogumil et al. 2014). Moreover, the ancestry of these gene families is mixed, with some members showing greater phylogenetic affinities to bacteria and others to archaea. As in bacteria, the eukaryotic system initiates with a Hsp40 protein recruiting a client protein and then stimulating Hsp70 ATPase activity to assist in folding of the client protein. Eukaryotic Hsp70s commonly operate with several different Hsp40 proteins, but specialization also occurs.

As with bacterial GroEL, eukaryotic CCTs often have accessory functions (Henderson et al. 2013). For example, the specific system operating as a chaperone in the mitochondrion is also involved in mitochondrial genome maintenance and protein import. Hsp90 proteins are found throughout eukaryotes, with separate families operating in the cytoplasm, the endoplasmic reticulum, the mitochondrion, and the plastid (in plants), and these interact with a diversity of cochaperones, with numerous secondary functions, patchily distributed among various branches of the

eukaryotic tree (Johnson and Brown 2009; Taipale et al. 2010; Johnson 2012).

If any generality emerges from this morass of complexity, it is that nearly all of cellular life depends on protein complexes specifically assigned to protein folding. Despite the increased proteome complexity in eukaryotes, there has been no major expansion in the core types of chaperones, although the numbers of gene copies and assisting cochaperones did expand (Powers and Balch 2013; Rebeaud et al. 2021), as did the heteromeric complexity of the individual types. Such alterations must have been accomplished in such a way that the basic folding capacities of cells remained uncompromised during the transitions. There is no evidence that the systems established in any particular lineage are fundamentally superior in any ways to those in others, but as discussed below, each system must be specifically tuned to its resident client proteins.

Client-chaperone coevolution. Unlike most enzymes, chaperones typically have a wide variety of client substrates. For example, $\sim$ 20% of the $\sim$ 4000 encoded proteins in *E. coli* are chaperone dependent, and of these, at least 250 appear to rely on GroEL for proper folding, while another 400 or so rely on Hsp40/70, and about 170 are serviceable by both (Kerner et al. 2005; Fujiwara et al. 2010; Niwa et al. 2012). In yeast, $\sim$ 20% of proteins are clients of Hsp90 alone (Taipale et al. 2010).

Such a vast repertoire of substrates raises questions about the degree to which the features of individual chaperone systems are compromised by the numbers of client genes, one issue being that any evolutionary movement toward a better fit of one client may diminish the effectiveness with others (Lynch and Hagner 2015; Figure 14.2). Wang et al. (2002) acquired some insight into this matter by engineering *E. coli* to carry a foreign protein (green fluorescent protein, GFP, encoded on a plasmid) and then imposing a selective challenge on cultures to improve the folding of GFP into functional molecules (readily revealed as fluorescing cells). This resulted in the evolution of a novel GroEL variant with substantially improved GFP folding but a reduced ability to fold normal client proteins, consistent with the chaperone being intrinsically constrained by the need to simultaneously satisfy the needs of multiple interactors.

Another potential example of such a compromise is the reliance of eukaryotic actins and tubulins on chaperones for the production of properly folded monomeric subunits. Together, these two molecules form the cytoskeleton, serve as highways for the transport of various cargoes, and have roles in numerous other eukaryotic cell functions (Chapter 16). However, despite their relatively simple and highly conserved structures, and a history extending back to at least LECA, in no case have the monomers of either protein been found to be capable of self-folding. Instead, both are major clients of CCT, which appears to have evolved specialized features for such processing (Llorca et al. 2001). Although bacterial GroEL will bind eukaryotic actin and tubulin, it is incapable of guiding them to their native conformations (Tian et al. 1995). Given the high intracellular concentrations of actins and tubulins, it is plausible that fine-tuning for processing these key client proteins imposes constraints on the capacity of CCT to service alternative substrates.

A third potential example of chaperones coevolving with client features involves bacterial species experiencing deleterious-mutation accumulation as a consequence

of serial bottlenecks. Notably, GroEL comprises up to 70% of the protein in some insect endosymbiotic bacteria, which are thought to experience increased random genetic drift owing to their vertical transmission from maternal to daughter insect. This has led to the suggestion that the elevated investment in chaperones arose as a mechanism to accommodate the accumulation of mildly deleterious mutations in the endosymbiont's protein-coding genes (Moran 1996; Fares et al. 2004). However, such drift-prone bacterial lineages experience even more accelerated rates of sequence evolution in the chaperones themselves than in other proteins (Herbeck et al. 2003; Warnecke and Rocha 2011). This coincident elevation of amino-acid substitutions in both chaperones and client proteins then raises questions as to whether the force driving these changes is adaptive remodeling of key chaperone motifs in response to mutations in specific client genes, and/or whether chaperone over-expression is an evolutionary compensation for its own reduced catalytic capacity. Whatever the mechanism, it is notable that substantial overexpression of chaperone genes also occurs in nucleomorphs (remnants of the nuclear genomes contained within photosynthetic endosymbionts of some algae) that also exhibit elevated rates of protein evolution (Hirakawa et al. 2014).

Given their large size, high expression levels, and reliance on ATP, chaperones such as GroEL comprise a significant fraction of the energy budget of a cell. Thus, elevated chaperone expression may come at a considerable cost that is only warranted under extreme genetic conditions. Experimental data do suggest that bacterial cells respond physiologically to the genome-wide accumulation of deleterious mutations by up-regulating GroEL expression. For example, Maisnier-Patin et al. (2005) observed such a response in mutation-accumulation lines of *Salmonella,* with additional artificial enhancement of GroEL expression resulting in still further increase in fitness. Similarly, Fares et al. (2002) found that *E. coli* lines allowed to accumulate enough mutations to reduce fitness by $\sim 50\%$ were restored to $\sim 90\%$ fitness following overexpression of GroEL; this type of observation extends to Hsp70 as well (Aguilar-Rodríguez et al. 2016). In contrast, *E. coli* cultures maintained at large population sizes often evolve reduced GroEL expression, possibly as a consequence of selection for mutations that reduce unnecessary energetic expenditure (Sabater-Mũnoz et al. 2015).

This being said, selection for improved client-protein folding is not the only evolutionary determinant of the architectural features of chaperones. Most notably, owing to their roles as safe havens for protein assembly, chaperones are vulnerable to exploitation by pathogens. For example, the genome of bacteriophage T4 (a virus of *E. coli*) encodes for a protein that is a molecular mimic of GroES and uses this feature to assemble its head proteins with GroEL (Keppel et al. 2002). Many other bacteriophage are dependent on host-encoded chaperones for proper development (Nakonechny and Teschke 1998; Karttunen et al. 2015). In fact, it was a serendipitous study of bacteriophage that led to the discovery of GroEL/GroES – the finding of *E. coli* mutants that promoted defective bacteriophage capsid assembly (Georgopoulos 2006). Many eukaryotic viruses also rely on host-cell chaperones to complete their life cycles (Geller et al. 2012). Thus, the degree to which selection to avoid cellular parasites directly conflicts with selection for efficient handling of a cell's endogenous proteins by chaperones merits further study.

Many other open questions remain with respect to the coevolution of chaperones

and their client proteins, including the extent to which clients become evolutionarily addicted to assisted folding once reliance on a chaperone has become initiated. Following the sort of constructive neutral-evolution scenario outlined in Chapter 6, with a reliable mechanism of assisted folding in place, mutations that would otherwise prevent self-folding of a protein might be expected to accumulate. A phylogenetic analysis of the clients of human Hsp90 suggest that this is not the case, with both gains and losses of chaperone dependence being common (Taipale et al. 2012), so this is another area ripe for further investigation.

Much of the uncertainty here is a consequence of the low level of understanding of the precise mechanisms by which chaperones identify their client proteins, the details of which are central to all aspects of coevolutionary engagement and escape. Rousseau et al. (2006) suggest that 10 to 20% of the residues within proteomes across the Tree of Life are contained within segments with a capacity for aggregation if left unfolded, but that such regions tend to be flanked with positively charged amino acids (arginine, lysine, and proline) that are targets of chaperones. Less clear, however, is whether the latter sequences arose in response to the accumulation of aggregation-prone sequences, or appeared first and simply paved the way for the safe accumulation of otherwise adhesive amino-acid residues. If the former is involved, this then raises the challenging question as to why selection should not minimize the accumulation of aggregation features to start with, as opposed to accepting such properties and then making compensatory modifications to minimize their effects.

Chaperone-mediated phenotypic evolution. Given that chaperones modulate protein quantity and quality, the question arises as to whether such activity can influence individual phenotypes in ways that might modify the course of evolution. In the search for adaptive purposes of traits, one particularly extreme view has been promoted – that chaperones facilitate adaptive evolution by buffering the normally deleterious effects of mutant alleles, thereby encouraging the effectively neutral build-up of a load of hidden but latent phenotypic effects. The idea that such variation might be exposed if a chaperone system becomes overwhelmed in a stressful environment lead to the suggestion that chaperones can act as "capacitors" for evolutionary change by promoting the expression of conditionally beneficial effects (Rutherford and Lindquist 1998). Further imagining that stressful environments are the ones where aberrant phenotypes are most likely to have utility lead to speculation that chaperones (in actuality, their liability to becoming overwhelmed) enhance the ability of populations to adapt to extreme selective challenges. If sustained, this might somehow eventually lead to the constitutive expression of the previously suppressed variant, moving the population into an entirely new phenotypic domain.

Multiple arguments shed doubt on the correctness of these ideas (Levy and Siegal 2008; Tomala and Korona 2008; Siegal and Masel 2012; Charlesworth et al. 2017). First, because chaperones service hundreds of client proteins, for adaptive capacitance to work, the exposure of any single transiently beneficial variant must outweigh the consequences of a likely vast array of other exposed deleterious variants. Second, there is the issue of how a variant that is not expressed for considerable periods of time can avoid the neutral accumulation of still more deleterious, condition-dependent mutations, thereby eventually being rendered nonfunctional

when exposed. Third, if some mechanism does exist by which transient exposure could lead to the expression of a novel protein function, then what becomes of the original function? Fourth, the suppression of chaperone activity can lead to the release of mobile-element activity (Specchia et al. 2010) and/or elevated rates of production of aneuploid progeny (Chen et al. 2012), imposing additional negative consequences. Fifth, for the entire scenario to work, chaperone stress must keep the extreme phenotype exposed to selection for a long enough time to enable new mutations to produce a mechanism for constitutive expression, but short enough to avoid population extinction.

Finally, implicit in the argument that compromised chaperone capacity leads to a release of latent variation is the assumption that chaperones do indeed buffer the effects of new mutations. In fact, the empirical evidence suggests otherwise. In yeast, the effects of standing variation are muted by chaperone activity, but the phenotypic effects of *de novo* mutations are actually magnified on average (Geiler-Samerotte et al. 2016). The overall implication is that natural selection differentially promotes alleles whose effects are buffered by chaperones, not the other way around.

This being said, although it is unlikely that chaperones have been advanced to enhance long-term evolvability, they may nonetheless play indirect roles in short-term evolutionary processes. An example of how chaperones might mediate the evolution of a novel protein function is provided by an experiment in which an expendable protein in *Pseudomonas aeruginosa,* phosphotriesterase, was selected for a novel arylesterase function (Wyganowski et al. 2013). In the experimental system, by controlling the expression of GroEL, it was possible to select for protein function under conditions of either high or low chaperone activity. High chaperone levels allowed the advancement of protein variants with elevated catalytic activity but low folding stability, whereas subsequent return to a low level of GroEL imposed strong selection for compensatory mutations against destabilizing mutations. Several rounds of such selection eventually led to a 10^4-fold increase in arylesterase activity and a near absence of GroEL dependency. Additional experiments of this nature have led to the improvement of the catalytic performance of other enzymes at the expense of self-folding capacity (Tokuriki and Tawfik 2009).

This kind of experimental result, reliant on a contrived experimental setup – alternating periods of high and low GroEL expression, and selection on a nonessential protein, needs to be tempered vis-á-vis the patterns actually seen with natural GroEL clients. Contrary to expectations under the hypothesis that chaperones lead to a relaxation of selection on protein evolution and/or facilitate movement into new adaptive domains, the client proteins of GroEL tend to be slowly evolving (Williams and Fares 2010). Although the subset of clients that are obligately dependent on GroEL and Hsp70 do evolve somewhat more rapidly at the protein-sequence level (Bogumil and Dagan 2010; Williams and Fares 2010; Aguilar-Rodríguez et al. 2016; Kadibalban et al. 2016; Alvarez-Ponce et al. 2019), such a pattern could also exist for reasons unassociated with folding.

In summary, all of the preceding observations strongly support the view that the function of chaperones is to suppress the negative phenotypic consequences of problematic protein folding rather than to store away hopeful monsters. Pushing most organisms beyond their physiological capacities invariably leads to aberrant, pathological phenotypes, so there is nothing particularly unique about the phe-

notypic consequences of overtaxed chaperones. More generally, the broader idea that various biological features have emerged specifically to enhance the long-term evolvability of species is without support and largely incompatible with evolutionary theory, conveniently ignoring the fact that selection operates on individuals in the present and has no capacity to see into the future. There is no known evolutionary mechanism to advance a cellular feature for the specific purpose of allowing the long-term accumulation of suppressed variation with conditionally beneficial effects in some future environment. Indeed, population-genetic theory demonstrates that a release of hidden genetic variation on a mutant background (or in a stressful environment) is a generic property of complex genetic systems, regardless of the prior state of buffering, and not an indicator of the prior evolution of a mechanism for ensuring robustness (Hermisson and Wagner 2004).

It is true that any mechanism that can sufficiently increase the robustness of an organism to perturbations can be selectively favored (de Visser et al. 2003), provided that the strength of selection exceeds that of random genetic drift and that the energetic cost is not too great. Chaperones do indeed expand the capacity of organisms to survive through stressful conditions. However, it does not follow that the assimilation of such a mechanism into a species owes its existence to selection for the long-term evolutionary flexibility of the lineage, nor even that there are any long-term benefits. More likely, there are disadvantages. Although selection for a robustness-enhancing feature may hide background defects in the short-term, in the long run, a new load of defects is expected to bring the population back to the previous fitness state, but with the added expense of maintaining a new layer of surveillance machinery (Frank 2007; Gros and Tenaillon 2009; Lynch 2012). In this sense, the idea that natural selection produces fundamentally superior organisms by adding layers and layers of buffering mechanisms to stabilize high fitness is an illusion (Chapter 20).

Protein Disposal

All organisms are confronted with the challenge of eliminating proteins that are structurally aberrant (owing to improper folding), functionally unnecessary or inappropriate (owing to the completion of prior tasks), or damaged by a wide variety of intracellular effects (such as thermal denaturation and oxidation). To accomplish such tasks, most prokaryotes and possibly all eukaryotes harbor a special molecular machine, the proteasome, which carries out processive protein degradation in an ATP-consuming process. The proteasome consists of a barrel-like structure, reminiscent of that found for the CCT noted above, which provides a safe compartment for restricting protease activity to target proteins and protecting other desirable proteins from proteolysis.

The proteasome exhibits a phylogenetic gradient in complexity similar to that seen for CCT chaperones. In archaea, eukaryotes, and a few bacteria, the barrel consists of four layers of heptameric rings, with the outer rings forming pores through which cargoes are delivered. In most archaea, the two inner (β) rings are homomeric, comprised of catalytic subunits encoded by a single locus, whereas the outer (α)

scaffold rings are homomers of another gene product. In contrast, in eukaryotes each of the fourteen subunits (seven for the α and β rings, respectively) are encoded separately (Pühler et al. 1993). Deviating substantially from the situation in archaea and eukaryotes, the bacterial proteasome is generally comprised of two homomeric rings with six subunits, although archaeal-like structures with seven subunits are found sporadically throughout the bacterial domain (Valas and Bourne 2008; Fuchs et al. 2017, 2018).

Thus, although the proteasome dates back to LUCA, we are again confronted with both an increase in complexity and an expansion in the number of subunits of the eukaryotic version, which must have involved an evolutionary alteration of binding interfaces (Foundations 14.1). Based on their phylogenetic distribution, the origins of all fourteen distinct eukaryotic subunits predate LECA (Bouzat et al. 2000). Along with this shift in proteasome complexity, the regulator proteins that control the entry of cargo proteins into the proteasome consist of at least six different subunit types throughout eukaryotes but only one in archaea (Fort et al. 2015).

In parallel with the proteasome, numerous other proteases operate in both prokaryotic and eukaryotic cells (e.g., Clausen et al. 2011). Many of these complete the degradational process, as proteasome degradation only reduces substrates to short oligopeptides, not single amino acids. Additional machinery, the exosome and its regulatory proteins,is deployed in the selective degradation of specific RNAs (Makino et al. 2013). As in the case of the proteasome, the nine subunit barrel of the exosome has experienced an increase in complexity from archaea to eukaryotes (three vs. nine distinct contributing proteins).

The selective targeting of proteins for disposal is generally orchestrated by pathways dedicated to marking molecules with specific degradation signals. In eukaryotes, the most prominent mechanism by far is the ubiquitylation pathway (Mogk et al. 2007; Sriram et al. 2011; Varshavsky 2011, 2019). In a series of three enzymatically guided steps, ubiquitin is delivered and ligated to specific lysine residues on target molecules in an ATP-dependent manner (Figure 14.3). From this starting point, chains of polyubiquitin are then grown, providing a signal for proteasome delivery. Deubiquitylation occurs prior to entry into the proteasome, sparing the ubiquitin molecules from degradation. The presence of all components of this pathway in some lineages of archaea implies a pre-LECA origin, apparently with independent expansions and specialization of component parts occurring in animals and land plants (Grau-Bové et al. 2015). Pathways with essentially the same features but quite different molecular participants exist in bacteria (Mogk et al. 2007; Mukherjee and Orth 2008), so an even earlier origin cannot be ruled out. In a related eukaryotic pathway, acetylation of specific residues provides another signal for degradation (Hwang et al. 2010; Shemorry et al. 2013).

In addition to its central role in protein degradation, the eukaryotic ubiquitylation / deubiquitylation pathway provides a means for dynamically switching proteins between alternative activity states in a wide variety of cellular functions. These include the cell cycle, DNA repair, vesicle trafficking, and signal transduction (e.g., Hirsch et al. 2009; Raiborg and Stenmark 2009; Ulrich and Walden 2010). Remarkably, a number of pathogenic bacteria have independently evolved molecular mimics of ubiquitin ligases, enabling them to commandeer various aspects of the

machinery of host cells (Hicks and Gálan 2010).

The ubiquitin-proteasome degradation system provides still another example of the importance of intracellular molecular languages in guiding key cellular events. In this case: 1) specific amino-acid residues at the termini of proteins (usually the N ends) define their susceptibility to ubiquitylation; 2) specific internal sites of the target molecules (usually lysine residues) are post-translationally modified by the covalent conjugation of ubiquitin (usually as polyubiquitin chains); and 3) the resultant linked ubiquitin moieties serve as indicators for delivery of the modified protein to the proteasome. Ubiquitylation is mediated by ubiquitin ligases, which rely on specific amino-acid sequence motifs for precise ubiquitin conjugation. Hundreds of such ligases with unique recognition sequences are often encoded within individual genomes, providing both specificity and an immense functional reach of the overall system.

The recognition determinants for protein degradation generally consist of specific amino-acid residues at the N- or C-termini of proteins, referred to as N-degrons or C-degrons (Figure 14.4). The exact nature of degrons (i.e., the recognition language) can differ among major groups of organisms (e.g., bacteria, land plants, and animals; Mogk et al. 2007). Further complicating things is the presence of enzymes for removing the initial methionine residues on polypeptide chains, other endopeptidases for severing small N-terminal peptide chains (thereby exposing new degradation determinants), and still others for converting some N-terminal amino-acid residues to others (e.g., Asn to Asp and Gln to Glu in eukaryotes). The latter residues can be viewed as secondary/tertiary destabilizing N-terminal residues, as they are only effective after modification, and even then, often only after covalent attachment of yet another amino acid that serves as the primary determinant (Arg in the case of eukaryotes). In the case of *E. coli*, N-terminal Arg and Lys serve as secondary destabilizing factors residues, which become active after terminal attachment of Leu. It has been suggested that the Arg-transferase utilized in eukaryotes is related to the one of the Leu-transferases in bacteria (Graciet et al. 2006).

Together then, two signals, one for denoting stabilization/destabilization status and the other for ligase-mediated attachment of ubiquitin to specific sites, largely determine the half lives of individual proteins. Notably, the internal lysine sites involved in ubiquitylation are only slightly more conserved over evolutionary time than other adjacent lysine residues, suggesting a high degree of redundancy with respect to location combined with stabilizing selection on the total numbers of such sites per protein (Hagai et al. 2012; Lu et al. 2017), a point that will be highlighted below for phosphorylation sites.

Finally, although bacteria have a pathway for protein disposal that is similar to that of eukaryotes, including the use of N-end rules, the bacterial pathway is substantially simpler than that in its eukaryotic analogs. Notably, however, a number of the destabilizing N-terminal amino acids in eukaryotes are the same as those in bacteria, suggesting a common ancestry of this N-degron system that emerged prior to LUCA. Varshavsky (2011) has suggested that, despite the increased complexity of the system that mediates processive proteolysis in eukaryotes, the eukaryotic system is no more efficient than that in bacteria, with "overdesign" in the former having arisen by effectively neutral processes operating during phases of reduced effective population sizes.

Post-translational Modification

An additional stage in the life histories of many proteins involves post-translational covalent linkage of small molecules to certain amino-acid residues. The diversity of molecular moieties that can be conjugated to proteins is substantial, ranging from small phosphoryl, adenyl, acetyl, and amide groups to larger molecules such as sugars and fatty-acid chains, and even to entire proteins such as ubiquitin (just discussed). Although nearly all amino-acid residues can participate in such modifications, the exact residue marked in any situation depends on the organism and cellular context. The precise functions of such markings are known in just a fraction of cases, but there is little question that post-translational modifications can lead to changes in structure, stability, and/or localization of the affected proteins, thereby modulating their functions. Two cases stand out in particular – the eukaryotic cell cycle (Chapter 10) and signal-transduction systems used in environmental sensing (Chapter 22).

Thus, although the classical view of gene regulation focuses on gene-expression modification at the level of transcription (Chapter 21), post-translational modifications yield additional dimensions to the overall complexity of regulation in both prokaryotes and eukaryotes. Although the mechanics differ dramatically, there are many similarities between the evolutionary features of transcriptional and post-translational regulation: both involve *trans*-regulating proteins with interactions targeting simple binding sites (DNA in the first case, and proteins in the second); both are subject to divergence with nonfunctional consequences; and both are typically under some form of purifying selection.

Although post-translational modification is largely uncharted territory for the field of evolutionary biology, one major target of study involves phosphorylation (Bradley 2022). Phosphoryl-group (PO_4^{3-}) additions are generally restricted to serine, threonine, and tyrosine residues in animals, and to arginine, aspartate, cysteine, and histidine in bacteria (Chapter 22). Covalent attachment of phosphoryl groups is generally carried out by specialized enzymes called kinases, most of which have simple recognition sites comprised of a substrate amino-acid residue plus just two to four flanking residues (Ubersax and Ferrell 2007; Miller and Turk 2018; Ochoa et al. 2018). Such simplicity raises the potential for substantial promiscuity, often rendering inferences on functional significance of specific sites quite uncertain. Gratuitous phosphorylation may be difficult to select against, as the cost of just a few extra ATP hydrolyses is relatively small compared to the total cost of building a protein (the average cost of an amino acid being ~ 30 ATPs; Chapter 17). However, certain forms of inappropriate phosphorylation may have negative functional consequences (Brunk et al. 2018; Cantor et al. 2018; Viéitez et al. 2022).

The immediate effect of phosphorylation is the addition of a negative charge to the acceptor residue. Such a change can often have downstream effects such as protein activation or inhibition. In addition, protein phosphorylation can be rapidly reversed by use of specific phosphatases (Chapter 22). Substantial numbers of eukaryotic genes are dedicated to post-translational modifications of this sort. For example, $\sim 2\%$ of the yeast genome encodes for protein kinases, with $\sim 40,000$ phosphosites distributed throughout the proteome (Zhu et al. 2000; Lanz et al. 2021). More than 500 kinases and 200 phosphatases are encoded in the human

genome (Manning et al. 2002; Alonso et al. 2004).

Proteome-wide data provide insights into the long-term evolutionary stability of phosphorylation sites. For example, comparative studies in yeasts and mammals indicate that many phosphorylated serines and threonines are under purifying selection to retain their phosphosite status (Gray and Kumar 2011; Levy et al. 2012), and sites known to have functionally relevant phosphorylation are more conserved than those with no known function. There is also evidence that sets of phosphorylation sites undergo subfunctionalization following gene duplication (Amoutzias et al. 2010; Freschi et al. 2011; Kaganovich and Snyder 2012), with each member of a paralogous pair partitioning up the ancestral sites, although the functional significance of this remains unclear. While a large fraction of such sites appears free to vary among species in terms of status and location, not all phosphosites are simply evolving neutrally (Moses et al. 2007; Holt et al. 2009; Landry et al. 2009; Nguyen Ba and Moses 2010; Freschi et al. 2014; Studer et al. 2016). For example, only $\sim 5\%$ of all phosphosites appear to have been conserved across the entire yeast lineage (dating back ~ 700 million years), and even when the same phosphorylatable residue is present in two moderately related species, their phosphorylation status may differ.

Phosphosites tend to be clustered on the surface of a protein or in disordered regions, and the critical feature may simply be the production of functionally appropriate local charge. Notably, the negatively charged aspartate and glutamate residues often serve as replacements (and/or sources) for their phosphorylatable counterparts (although the amino-acid interconversions require two nucleotide substitutions per codon), i.e., phosphosites often evolve from phosphomimetic Asp and Glu sites and vice versa (Kurmangaliyev et al. 2011; Pearlman et al. 2011; Diss et al. 2012).

Taken together, these observations suggest a scenario whereby the degree of a protein's phosphorylation is under stabilizing selection for an appropriate total negative charge, with the specific locations of many of the affected residues relatively free to wander in a quasi-neutral fashion (Lienhard 2008; Landry et al. 2014). That is, the level of phosphorylation of individual proteins appears to operate as a sort of quantitative trait, with the total number of phosphorylated residues being conserved, but also with enough degrees of freedom that there can be considerable turnover of specific phosphosites on evolutionary timescales (Foundations 14.2).

Summary

- All organisms harbor subsets of proteins whose proper folding requires assistance from chaperones. These molecular guardians appear to have been present in LUCA, and likely paved the way for the establishment of long proteins incapable of self-folding. Despite their critical functions, the families of chaperones have diversified substantially with respect to multimeric structures, with expansions in complexity being most extreme in eukaryotes.

- The process of chaperone-assisted folding is relatively cheap, on the order of the

biosynthetic cost of one to a few amino acids per protein molecule, although the structures themselves are often complex and can constitute a substantial fraction of the total protein within a cell.

- As the number of chaperone systems per cell is dwarfed by the number of their client proteins, coevolutionary conflicts arise with respect to the recognition of specific clients, with the fine-tuning to any one particular client reducing the affinity to others.

- Chaperones are commonly exploited by viruses as assembly chambers for viral capsids, imposing still additional constraints on the evolution of chaperone recognition capacities. The very high evolutionary rates of some eukaryotic chaperones may be a consequence of host-pathogen coevolutionary arms races.

- The extent to which client proteins become addicted to the safe havens of chaperones and embark on a path of no return to self-folding is unclear, but the relatively low cost of such dependence may mean that many proteins are not far from drifting down a path of chaperone dependence by effectively neutral processes.

- It has been argued that chaperones serve as capacitors of adaptive evolution, by masking the deleterious effects of mutations in benign environments but releasing novel phenotypes when stressful environments overwhelm surveillance systems. There is, however, no direct support for the idea that chaperones are maintained to enhance the evolvability of species, and multiple lines of evidence are inconsistent with it.

- Essentially all organisms have systems for selectively destroying damaged or superfluous proteins, largely via a barrel-like machine called the proteasome. Selective protein degradation relies on a detailed set of communication rules involving sequence motifs on target molecules and a system of enzymes for marking specific sites as indicators for disposal. The baseline system for regulated protein degradation dates back to LUCA, although the complexity of the processes has expanded in eukaryotes.

- Across the Tree of Life, the structures and functions of numerous proteins are influenced by post-translational modifications involving the covalent conjugation of various side groups to specific amino-acid residues. Best studied is the case of phosphorylation, wherein the specific locations of many target phosphosites appear relatively free to wander in an effectively neutral fashion over evolutionary timescales provided their local density does not change significantly.

Foundations 14.1. The CCT complex. The CCT (chaperonin containing tailless complex) presents a striking example of a transition of a multimeric protein from a homomeric to a complex heteromeric state. Restricted to archaea and eukaryotes, CCT chaperonins are generally double-barreled hexadecamers (occasionally octodecamers), i.e., with 8 or 9 monomeric subunits per barrel (Archibald et al. 1999, 2001).

In archaea, the overall structure is homomeric or heteromeric with two or three alternating subunits (in 8- or 9-component barrels, respectively). The evidence suggests that conditionally deleterious mutations have accumulated in the contact regions between paralogous subunits in heteromeric archaeal CCTs, with compensatory mutations then serving to create a sort of evolutionary entrapment (Ruano-Rubio and Fares 2007). Under this hypothesis, the ancestral CCT was a homo-oligomer that then diversified in architecture following gene duplication, via an effectively neutral evolutionary pathway and with no significant change at the functional level (Archibald et al. 1999). Nonetheless, the evolution of complexity is not unidirectional in CCT, as there are examples of the reversion of heteromeric complexes to homomers.

In contrast, all of CCT subunits in eukaryotes are encoded by separate genes. With eight different subunits per ring in the eukaryotic version of CCT, there are thousands of possible arrangements under random assembly, and yet it is thought that just one assembly is consistently achieved in the cell (Kalisman et al. 2012), i.e., there are precisely calibrated binding affinities between the eight subunits. The underlying duplication and divergence of subunits occurred early in eukaryotic history, apparently pre-LECA, as the different subunits within a species are more divergent from each other than are orthologous subunits across major phylogenetic lineages (Fares and Wolfe 2003). Moreover, the eight eukaryotic duplicates appear to diverge at the amino-acid sequence level at rates exceeding the neutral expectation, thus suggesting positive selection for diversification in function, potentially with each subunit being relatively specialized to a different set of client proteins (Fares and Wolfe 2003; Joachimiak et al. 2014).

Understanding the evolution of an initially homomeric ring into such a complex heteromeric state imposes several challenges. At each evolutionary step, a mechanism is required to permanently preserve both the new and the old members of the complex, either via the gain of a beneficial function or complementary losses of subfunctions (Chapter 13). Moreover, the addition of each new member of the ring likely requires the fixation of at least two mutations, as ring architectures necessitate that each subunit be involved in two distinct interfaces. Each step of the process also raises the above-noted problem of hetero-oligomerization – the assembly of heterogeneous mixtures of subunits in individual complexes that is likely to persist until a high level of interface specificity has evolved (Figure 14.5).The need for understanding of these kinds of issues is not confined to chaperone evolution, as numerous other cellular features have ring-like structures, e.g., the nuclear pore (Chapter 15), the proteasome (this chapter), and a number of DNA-binding proteins in eukaryotes (Chapter 10).

Foundations 14.2. The evolution of a digital trait. Phosphorylation and other post-translational modifications are examples of digital traits, in the sense that they have a simple molecular basis with the resultant phenotypes taking on integer values (e.g., equal to the number of modified amino-acid residues). Many other cellular features have this property, e.g., the number of residues involved in binding of a protein to its substrate, and the number of saturated and unsaturated bonds in a lipid molecule.

Such restriction of simple molecular traits to discontinuous values may impose

unique evolutionary consequences. For example, the optimum binding energy for a particular trait may be unattainable unless it coincides with an integer multiple of the underlying granularity. If this is not the case, two allelic states straddling the optimum may have nearly the same fitness, resulting in an essentially neutral process of molecular evolution combined with a permanent state of suboptimal fitness. In addition, if certain suboptimal allelic states are more accessible by mutation, this can compete with the ability of natural selection to promote higher-fitness states. As will be discussed below, such conflicts can even be present in the absence of mutation bias. Finally, if a system has excess capacity, such that the typical state (e.g., number of modified residues) is well below the maximum possible value, substantial drift is possible among alternative phenotypes with equivalent effects.

Drawing upon an approach introduced in Chapter 5, here we consider a simple model for exploring these issues, with ℓ equivalent sites (factors), each with two alternative functional states, + and −, contributing positively and negatively to the trait. Under this model, a multiplicity of functionally equivalent classes exists with respect to the number of positive alleles (m). As an example, for the case of $\ell = 4$, there are five genotypic classes (m = 0, 1, 2, 3, and 4), with multiplicities 1, 4, 6, 4, and 1, respectively (Figure 14.6). These multiplicities are equivalent to the coefficients of a binomial expansion, e.g., $(x + y)^m$. With equivalent fitness for all members within a particular class, this variation in multiplicity of states plays an important role in determining the long-term evolutionary distribution of alternative classes – all other things being equal, classes with higher multiplicities are more accessible over evolutionary time.

As discussed in Foundations 5.2, a system like this yields an equilibrium distribution of a population occupying alternative states over a long evolutionary time period, given constancy of the population-genetic environment. That is, over time the mean phenotype is expected to wander within limits dictated by the strength of selection for alternative classes, the degree of mutation bias, and the power of random genetic drift. Justification of this quasi-steady-state view derives from the fact that many cellular traits have functions (and cytoplasmic environments) that may have remained relatively stable for tens to thousands of millions of years (even in the face of a changing external environment).

The probabilities of alternative states depend on the relative magnitudes of the transition coefficients between adjacent classes (Figure 14.6). Each of these coefficients is equal to the product of a multiplicity, a per-site mutation rate, and a probability of fixation of a new mutation. The per-generation mutation rates from the − to + state, and vice versa, are defined to be u_{01} and u_{10}, respectively. The probability of fixation is given by the standard expression outlined in Chapter 4. A haploid, nonrecombining population is assumed here, so that each set of functionally equivalent states comprises a genotypic class.

In the limiting case of neutrality, the equilibrium probability of any site being occupied by a + allele is simply $\eta = u_{01}/(u_{01} + u_{10})$, the fraction of the summed mutation rates in the + direction, and the states of all sites will be independent. (Here, the probability of fixation, $1/N$, factors out because it is identical for all mutations). The neutral probability of a population residing in state m is then simply defined by the binomial distribution,

$$\widetilde{P}_{n,m} = \binom{\ell}{m} \eta^m (1-\eta)^{\ell-m}. \tag{14.2.1}$$

Thus, in this limiting case, the probability distribution for the class types only depends on: 1) the ratio of mutation rates, not on their absolute values; and 2) the binomial coefficient associated with each class, which defines the multiplicity of equivalent states in the class. The long-term mean and variance of the trait under neutrality, defined by the properties of the binomial distribution, are $\mu_n = \ell\eta$ and $\sigma_n^2 = \ell\eta(1-\eta)$, respectively.

Selection alters this baseline distribution by weighting each class by the factor $e^{2N_e s_m}$ (with a 4 being substituted for the 2 under diploidy), where N_e is the effective population size, and s_m is a measure of the class-specific deviation of fitness from some reference point (e.g., from the fitness of the optimal phenotype). The quantity $N_e s_m = s_m/(1/N_e)$ is equivalent to the ratio of the strength of selection relative to that of drift. The basis for this weighting term has already been discussed in Foundations 5.2 – it is the ratio of fixation probabilities from class $m-1$ to m and vice versa.

The overall distribution can then be written as

$$\widetilde{P}_m = \widetilde{P}_{n,m} \cdot e^{2N_e s_m} = C \cdot \binom{\ell}{m} \beta^m e^{2N_e s_m}, \tag{14.2.2}$$

where $\beta = u_{01}/u_{10}$ is the mutation bias (the ratio of mutation rates in both directions), and the normalization constant C is equal to the reciprocal of the sum of the terms to the right of C for $m = 0$ to ℓ, which ensures that the frequencies sum to 1.0. The term $(1-\eta)^\ell$ from Equation 14.2.1 has been absorbed into C, as it is a constant independent of m, and the specific reference from which the class-specific fitness deviations are measured does not matter either, as it cancels out through the normalization constant. The mean phenotype is

$$\mu_m = \sum_{m=0}^{\ell} m \cdot \widetilde{P}_m, \tag{14.2.3}$$

which reduces to $\ell\eta$ in the case of neutrality.

As a specific example of the application of Equation 14.2.2, consider the case of a trait under stabilizing selection, such that the fitness of an individual in genotypic class m is denoted by the Gaussian function,

$$W_m = e^{-(m-\theta)^2/(2\omega^2)}, \tag{14.2.4}$$

where θ is the optimum phenotypic value, and ω is a measure of the width of the fitness function (analogous to the standard deviation of a normal distribution). Selection is purely directional if $m = 0$ or ℓ, and neutrality is approached as $\omega \to \infty$. Although m is confined to integer values, θ need not be. The selection coefficient can be arbitrarily defined as $s_m = W_m - W_0$.

An application of the Gaussian fitness function to Equation 14.2.2, shown in Figure 14.6, illustrates several general points. First, a gradient in the average class value (e.g., the number of phosphorylation sites) is expected with respect to the effective population size, the exact location on the phenotypic scale depending on the strength of selection. When the fitness function is sufficiently flat that $N_e \ll \omega^2$, selection is overwhelmed by the power of drift, and the distribution converges on the neutral expectation, Equation 14.2.1. Only when $N_e \gg \omega^2$ does the force of selection overwhelm the power of drift to the extent that the population will almost always reside near the optimum. The actual optimum will only be achievable if θ is an integer. If this is not the case, the two attainable phenotypes straddling the optimum will be present as alternative states with frequencies depending on their relative fitnesses.

Second, there will frequently be two or more classes with probabilities substantially greater than zero, and sometimes with nearly equivalent values. The fact that populations will frequently have different phenotypic states even in a constant population-genetic environment raises significant reservations about the common practice of assuming that phenotypic differences are a consequence of different forms of selection.

Finally, because of the multiplicity of alternative, functionally equivalent states within each class, populations residing within the same class will commonly have different configurations of $-$ and $+$ states. For example, for the case of $\ell = 10$ and

two populations in state $m = 3$, the probability of no overlapping use of + sites is $[1-(3/10)][1-(3/9)][1-(3/8)] \simeq 0.29$. At equilibrium in state $m > 0$, the probability of any specific + site in one population being − in another is $(\ell - m)/\ell$. Each of these points is relevant to a diversity of situations in cellular evolution where there are multiple solutions to the same problem, e.g., the specific amino-acids residues serving as phosphosites on a post-translationally modified protein, or serving as binding residues on the interfaces in protein complexes.

Literature Cited

Abascal, F., A. Corpet, Z. A. Gurard-Levin, D. Juan, F. Ochsenbein, D. Rico, A. Valencia, and G. Almouzni. 2013. Subfunctionalization via adaptive evolution influenced by genomic context: the case of histone chaperones ASF1a and ASF1b. Mol. Biol. Evol. 30: 1853-1866.

Aguilar-Rodríguez, J., B. Sabater-Muñoz, R. Montagud-Martínez, V. Berlanga, D. Alvarez-Ponce, A. Wagner, and M. A. Fares. 2016. The molecular chaperone DnaK is a source of mutational robustness. Genome Biol. Evol. 8: 2979-2991.

Alonso, A., J. Sasin, N. Bottini, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, and T. Mustelin. 2004. Protein tyrosine phosphatases in the human genome. Cell 117: 699-711.

Alvarez-Ponce, D, J. Aguilar-Rodríguez, and M. A. Fares. 2019. Molecular chaperones accelerate the evolution of their protein clients in yeast. Genome Biol. Evol. 11: 2360-2375.

Amoutzias, G. D., Y. He, J. Gordon, D. Mossialos, S. G. Oliver, and Y. Van de Peer. 2010. Posttranslational regulation impacts the fate of duplicated genes. Proc. Natl. Acad. Sci. USA 107: 2967-2971.

Archibald, J. M., T. Cavalier-Smith, U. Maier, and S. Douglas. 2001. Molecular chaperones encoded by a reduced nucleus: the cryptomonad nucleomorph. J. Mol. Evol. 52: 490-501.

Archibald, J. M., J. M. Logsdon, and W. F. Doolittle. 1999. Recurrent paralogy in the evolution of archaeal chaperonins. Curr. Biol. 9: 1053-1056.

Bittner, A. N., A. Foltz, and V. Oke. 2007. Only one of five groEL genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. J. Bacteriol. 189: 1884-1889.

Bogumil, D., D. Alvarez-Ponce, G. Landan, J. O. McInerney, and T. Dagan. 2014. Integration of two ancestral chaperone systems into one: the evolution of eukaryotic molecular chaperones in light of eukaryogenesis. Mol. Biol. Evol. 31: 410-418.

Bogumil, D., and T. Dagan. 2010. Chaperonin-dependent accelerated substitution rates in prokaryotes. Genome Biol. Evol. 2: 602-608.

Bouzat, J. L., L. K. McNeil, H. M. Robertson, L. F. Solter, J. E. Nixon, J. E. Beever, H. R. Gaskins, G. Olsen, S. Subramaniam, M. L. Sogin, et al. 2000. Phylogenomic analysis of the alpha proteasome gene family from early-diverging eukaryotes. J. Mol. Evol. 51: 532-543.

Bradley, D. 2022. The evolution of post-translational modifications. Curr. Opin. Genet. Dev. 76: 101956.

Brunk, E., R. L. Chang, J. Xia, H. Hefzi, J. T. Yurkovich, D. Kim, E. Buckmiller, H. H. Wang, B. K. Cho, C. Yang, et al. 2018. Characterizing posttranslational modifications in prokaryotic metabolism using a multiscale workflow. Proc. Natl. Acad. Sci. USA 115: 11096-11101.

Cantor, A. J., N. H. Shah, and J. Kuriyan. 2018. Deep mutational analysis reveals functional trade-offs in the sequences of EGFR autophosphorylation sites. Proc. Natl. Acad. Sci. USA 115: E7303-E7312.

Carretero-Paulet, L., V. A. Albert, and M. A. Fares. 2013. Molecular evolutionary mechanisms driving functional diversification of the HSP90A family of heat shock proteins in eukaryotes. Mol. Biol. Evol. 30: 2035-2043.

Charlesworth, D., N. H. Barton, and B. Charlesworth. 2017. The sources of adaptive variation. Proc. R. Soc. B 284: 20162864.

Chen, G., W. D. Bradford, C. W. Seidel, and R. Li. 2012. Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. Nature 482: 246-250.

Clausen, T., M. Kaiser, R. Huber, and M. Ehrmann. 2011. HTRA proteases: regulated proteolysis in protein quality control. Nat. Rev. Mol. Cell Biol. 12: 152-162.

Craig, E. A., and J. Marszalek. 2011. Hsp70 chaperones. Encycl. Life Sci., John Wiley & Sons, Ltd., Chichester, UK.

de Visser, J. A., J. Hermisson, G. P. Wagner, L. A. Meyers, H. Bagheri-Chaichian, J. L. Blanchard, L. Chao, J. M. Cheverud, S. F. Elena, W. Fontana, et al. 2003. Evolution and detection of genetic robustness. Evolution 57: 1959-1972.

Diss, G., L. Freschi, and C. R. Landry. 2012. Where do phosphosites come from and where do they go after gene duplication? Int. J. Evol. Biol. 2012: 843167.

Fares, M. A., E. Barrio, B. Sabater-Muñoz, and A. Moya. 2002. The evolution of the heat-shock protein GroEL from *Buchnera,* the primary endosymbiont of aphids, is governed by positive selection. Mol. Biol. Evol. 19: 1162-1170.

Fares, M. A., A. Moya, and E. Barrio. 2004. GroEL and the maintenance of bacterial endosymbiosis. Trends Genet. 20: 413-416.

Fares, M. A., and K. H. Wolfe. 2003. Positive selection and subfunctionalization of duplicated CCT chaperonin subunits. Mol. Biol. Evol. 20: 1588-1597.

Fort, P., A. V. Kajava, F. Delsuc, and O. Coux. 2015. Evolution of proteasome regulators in eukaryotes. Genome Biol. Evol. 7: 1363-1379.

Frank, S. A. 2007. Maladaptation and the paradox of robustness in evolution. PLoS One 2: e1021.

Freschi, L., M. Osseni, and C. R. Landry. 2014. Functional divergence and evolutionary turnover in mammalian phosphoproteomes. PLoS Genet. 10: e1004062.

Freschi, L., M. Courcelles, P. Thibault, S. W. Michnick, and C. R. Landry. 2011. Network rewiring by gene duplication. Mol. Syst. Biol. 7: 504.

Fuchs, A. C. D., V. Alva, L. Maldoner, R. Albrecht, M. D. Hartmann, and J. Martin. 2017. The architecture of the Anbu complex reflects an evolutionary intermediate at the origin of the proteasome system. Structure 25: 834-845.

Fuchs, A. C. D., L. Maldoner, K. Hipp, M. D. Hartmann, and J. Martin. 2018. Structural characterization of the bacterial proteasome homolog BPH reveals a tetradecameric double-ring complex with unique inner cavity properties. J. Biol. Chem. 293: 920-930.

Fujiwara, K., Y. Ishihama, K. Nakahigashi, T. Soga, and H. Taguchi. 2010. A systematic survey of *in vivo* obligate chaperonin-dependent substrates. EMBO J. 29: 1552-1564.

Geiler-Samerotte, K. A., Y. O. Zhu, B. E. Goulet, D. W. Hall, and M. L. Siegal. 2016. Selection transforms the landscape of genetic variation interacting with Hsp90. PLoS Biol. 14: e2000465.

Geller, R., S. Taguwa, and J. Frydman. 2012. Broad action of Hsp90 as a host chaperone required for viral replication. Biochim. Biophys. Acta 1823: 698-706.

Georgopoulos, C. 2006. Toothpicks, serendipity and the emergence of the *Escherichia coli* DnaK

(Hsp70) and GroEL (Hsp60) chaperone machines. Genetics 174: 1699-1707.

Graciet, E., R. G. Hu, K. Piatkov, J. H. Rhee, E. M. Schwarz, and A. Varshavsky. 2006. Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. Proc. Natl. Acad. Sci. USA 103: 3078-3083.

Graciet, E., and F. Wellmer. 2010. The plant N-end rule pathway: structure and functions. Trends Plant Sci. 15: 447-453.

Grau-Bové, X., A. Sebé-Pedrós, and I. Ruiz-Trillo. 2015. The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. Mol. Biol. Evol. 32: 726-739.

Gray, V. E., and S. Kumar. 2011. Rampant purifying selection conserves positions with posttranslational modifications in human proteins. Mol. Biol. Evol. 28: 1565-1568.

Gros, P. A., and O. Tenaillon. 2009. Selection for chaperone-like mediated genetic robustness at low mutation rate: impact of drift, epistasis and complexity. Genetics 182: 555-564.

Hagai T, Á. Tóth-Petróczy, A. Azia, and Y. Levy. 2012. The origins and evolution of ubiquitination sites. Mol. Biosyst. 8: 1865-1877.

Henderson, B., M. A. Fares, and P. A. Lund. 2013. Chaperonin 60: a paradoxical, evolutionarily conserved protein family with multiple moonlighting functions. Biol. Rev. Camb. Phil. Soc. 88: 955-987.

Herbeck, J. T., D. J. Funk, P. H. Degnan, and J. J. Wernegreen. 2003. A conservative test of genetic drift in the endosymbiotic bacterium *Buchnera:* slightly deleterious mutations in the chaperonin groEL. Genetics 165: 1651-1660.

Hermisson, J., and G. P. Wagner. 2004. The population genetic theory of hidden variation and genetic robustness. Genetics 168: 2271-2284.

Hicks, S. W., and J. E. Galán. 2010. Hijacking the host ubiquitin pathway: structural strategies of bacterial E3 ubiquitin ligases. Curr. Opin. Microbiol. 13: 41-46.

Hirakawa, Y., S. Suzuki, J. M. Archibald, P. J. Keeling, and K. Ishida. 2014. Overexpression of molecular chaperone genes in nucleomorph genomes. Mol. Biol. Evol. 31: 1437-1443.

Hirsch, C., R. Gauss, S. C. Horn, O. Neuber, and T. Sommer. 2009. The ubiquitylation machinery of the endoplasmic reticulum. Nature 458: 453-460.

Hirtreiter, A. M., G. Calloni, F. Forner, B. Scheibe, M. Puype, J. Vandekerckhove, M. Mann, F. U. Hartl, and M. Hayer-Hartl. 2009. Differential substrate specificity of group I and group II chaperonins in the archaeon *Methanosarcina mazei.* Mol. Microbiol. 74: 1152-1168.

Holt, L. J., B. B. Tuch, J. Villén, A. D. Johnson, S. P. Gygi, and D. O. Morgan. 2009. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science 325: 1682-1686.

Horwich, A. L., A. C. Apetri, and W. A. Fenton. 2009. The GroEL/GroES cis cavity as a passive anti-aggregation device. FEBS Lett. 583: 2654-2662.

Hwang, C. S., A. Shemorry, and A. Varshavsky. 2010. N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327: 973-977.

Ishimoto, T., K. Fujiwara, T. Niwa, and H. Taguchi. 2014. Conversion of a chaperonin GroEL-independent protein into an obligate substrate. J. Biol. Chem. 289: 32073-32080.

Joachimiak, L. A., T. Walzthoeni, C. W. Liu, R. Aebersold, and J. Frydman. 2014. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. Cell 159: 1042-1055.

Johnson, J. L. 2012. Evolution and function of diverse Hsp90 homologs and cochaperone proteins. Biochim. Biophys. Acta 1823: 607-613.

Johnson, J. L., and C. Brown. 2009. Plasticity of the Hsp90 chaperone machine in divergent eukaryotic organisms. Cell Stress Chaperones 14: 83-94.

Kadibalban, A. S., D. Bogumil, G. Landan, and T. Dagan. 2016. DnaK-dependent accelerated evolutionary rate in prokaryotes. Genome Biol. Evol. 8: 1590-1599.

Kaganovich, M., and M. Snyder. 2012. Phosphorylation of yeast transcription factors correlates with the evolution of novel sequence and function. J. Proteome Res. 11: 261-268.

Kalisman, N., C. M. Adams, and M. Levitt. 2012. Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. Proc. Natl. Acad. Sci. USA 109: 2884-2889.

Karttunen, J., S. Mäntynen, T. O. Ihalainen, J. K. Bamford, and H. M. Oksanen. 2015. Non-structural proteins P17 and P33 are involved in the assembly of the internal membrane-containing virus PRD1. Virology 482: 225-233.

Keppel, F., M. Rychner, and C. Georgopoulos. 2002. Bacteriophage-encoded cochaperonins can substitute for *Escherichia coli*'s essential GroES protein. EMBO Rep. 3: 893-898.

Kerner, M. J., D. J. Naylor, Y. Ishihama, T. Maier, H.-C. Chang, A. P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann, et al. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli.* Cell 122: 209-220.

Keskin, O., I. Bahar, D. Flatow, D. G. Covell, and R. L. Jernigan. 2002. Molecular mechanisms of chaperonin GroEL-GroES function. Biochemistry 41: 491-501.

Kurmangaliyev, Y. Z., A. Goland, and M. S. Gelfand. 2011. Evolutionary patterns of phosphorylated serines. Biol. Direct 6: 8.

Laksanalamai, P., T. A. Whitehead, and F. T. Robb. 2004. Minimal protein-folding systems in hyperthermophilic archaea. Nat. Rev. Microbiol. 2: 315-324.

Landry, C. R., L. Freschi, T. Zarin, and A. M. Moses. 2014. Turnover of protein phosphorylation evolving under stabilizing selection. Front. Genet. 5: 245.

Landry, C. R., E. D. Levy, and S. W. Michnick. 2009. Weak functional constraints on phosphoproteomes. Trends Genet. 25: 193-197.

Lanz, M. C., K. Yugandhar, S. Gupta, E. J. Sanford, V. M. Faça, S. Vega, A. M. N. Joiner, J. C. Fromme, H. Yu, and M. B. Smolka. 2021. In-depth and 3-dimensional exploration of the budding yeast phosphoproteome. EMBO Rep. 22: e51121.

Levy, E. D., S. W. Michnick, and C. R. Landry. 2012. Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367: 2594-2606.

Levy, S. F., and M. L. Siegal. 2008. Network hubs buffer environmental variation in *Saccharomyces cerevisiae.* PLoS Biol. 6: e264.

Lienhard, G. E. 2008. Non-functional phosphorylations? Trends Biochem. Sci. 33: 351-352.

Llorca, O., J. Martín-Benito, J. Grantham, M. Ritco-Vonsovici, K. R. Willison, J. L. Carrascosa, and J. M. Valpuesta. 2001. The 'sequential allosteric ring' mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. EMBO J. 20: 4065-4075.

Lu, L., Y. Li, Z. Liu, F. Liang, F. Guo, S. Yang, D. Wang, Y. He, J. Xiong, D. Li, et al. 2017. Functional constraints on adaptive evolution of protein ubiquitination sites. Sci. Rep. 7: 39949.

Lund, P. A. 2009. Multiple chaperonins in bacteria – why so many? FEMS Microbiol. Rev. 33: 785-800.

Lynch, M. 2012. Evolutionary layering and the limits to cellular perfection. Proc. Natl. Acad. Sci. USA 109: 18851-18856.

Lynch, M., and K. Hagner. 2015. Evolutionary meandering of intermolecular interactions along the drift barrier. Proc. Natl. Acad. Sci. USA 112: E30-E38.

Macario, A. J., L. Brocchieri, A. R. Shenoy, and E. Conway de Macario. 2006. Evolution of a protein-folding machine: genomic and evolutionary analyses reveal three lineages of the archaeal hsp70(dnaK) gene. J. Mol. Evol. 63: 74-86.

Maisnier-Patin, S., J. R. Roth, A. Fredriksson, T. Nyström, O. G. Berg, and D. I. Andersson. 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat. Genet. 37: 1376-1379.

Makino, D. L., F. Halbach, and E. Conti. 2013. The RNA exosome and proteasome: common principles of degradation control. Nat. Rev. Mol. Cell. Biol. 14: 654-660.

Manning, G., D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. 2002. The protein kinase complement of the human genome. Science 298: 1912-1934.

Miller, C. J., and B. E. Turk. Homing in: mechanisms of substrate targeting by protein kinases. Trends Biochem. Sci. 43: 380-394.

Mogk, A., R. Schmidt, and B. Bukau. 2007. The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol. 17: 165-172.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA 93: 2873-2878.

Moses, A. M., M. E. Liku, J. J. Li, and R. Durbin. 2007. Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. Proc. Natl. Acad. Sci. USA 104: 17713-17718.

Mukherjee, S., and K. Orth. 2008. A protein pupylation paradigm. Science 322: 1062-1063.

Nakonechny, W. S., and C. M. Teschke. 1998. GroEL and GroES control of substrate flux in the *in vivo* folding pathway of phage P22 coat protein. J. Biol. Chem. 273: 27236-27244.

Nguyen Ba, A. N., and A. M. Moses. 2010. Evolution of characterized phosphorylation sites in budding yeast. Mol. Biol. Evol. 27: 2027-2037.

Niwa, T., T. Kanamori, T. Ueda, and H. Taguchi. 2012. Global analysis of chaperone effects using a reconstituted cell-free translation system. Proc. Natl. Acad. Sci. USA 109: 8937-8942.

Ochoa, D., D. Bradley, and P. Beltrao. 2018. Evolution, dynamics and dysregulation of kinase signalling. Curr. Opin. Struct. Biol. 48: 133-140.

Pearlman, S. M., Z. Serber, and J. E. Ferrell, Jr. 2011. A mechanism for the evolution of phosphorylation sites. Cell 147: 934-946.

Powers, E. T., and W. E. Balch. 2013. Diversity in the origins of proteostasis networks – a driver for protein function in evolution. Nat. Rev. Mol. Cell Biol. 14: 237-248.

Pühler, G., F. Pitzer, P. Zwickl, W. Baumeister. 1993. Proteasomes: multisubunit proteinases common to *Thermoplasma* and eukaryotes. Syst. Appl. Microbiol. 16: 734-741.

Raiborg, C., and H. Stenmark. 2009. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. Nature 458: 445-452.

Rebeaud, M. E., S. Mallik, P. Goloubinoff, and D. S. Tawfik. 2021. On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life. Proc. Natl. Acad. Sci. USA 118: e2020885118.

Rosenzweig, R., N. B. Nillegoda, M. P. Mayer, and B. Bukau. 2019. The Hsp70 chaperone network. Nat. Rev. Mol. Cell Biol. 20: 665-680.

Rousseau, F., L. Serrano, and J. W. Schymkowitz. 2006. How evolutionary pressure against protein aggregation shaped chaperone specificity. J. Mol. Biol. 355: 1037-1047.

Ruano-Rubio, V., and M. A. Fares. 2007. Testing the neutral fixation of hetero-oligomerism in the archaeal chaperonin CCT. Mol. Biol. Evol. 24: 1384-1396.

Rüßmann, F., M. J. Stemp, L. Mönkemeyer, S. A. Etchells, A. Bracher, and F. U. Hartl. 2012. Folding of large multidomain proteins by partial encapsulation in the chaperonin TRiC/CCT. Proc. Natl. Acad. Sci. USA 109: 21208-21215.

Rutherford, S. L., and S. Lindquist. 1998. Hsp90 as a capacitor for morphological evolution. Nature 396: 336-342.

Sabater-Muñoz, B., M. Prats-Escriche, R. Montagud-Martínez, A. López-Cerdán, C. Toft, J. Aguilar-Rodríguez, A. Wagner, and M. A. Fares. 2015. Fitness trade-offs determine the role of the molecular chaperonin GroEL in buffering mutations. Mol. Biol. Evol. 32: 2681-2693.

Santra, M., D. W. Farrell, and K. A. Dill. 2017. Bacterial proteostasis balances energy and chaperone utilization efficiently. Proc. Natl. Acad. Sci. USA 114: E2654-E2661.

Schilke, B., B. Williams, H. Knieszner, S. Pukszta, P. D'Silva, E. A. Craig, and J. Marszalek. 2006. Evolution of mitochondrial chaperones utilized in Fe-S cluster biogenesis. Curr. Biol. 16: 1660-1665.

Sharma, S. K., P. De los Rios, P. Christen, A. Lustig, and P. Goloubinoff. 2010. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. Nat. Chem. Biol. 6: 914-920.

Shemorry, A., C. S. Hwang, and A. Varshavsky. 2013. Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol. Cell 50: 540-551.

Siegal, M. L., and J. Masel. 2012. Hsp90 depletion goes wild. BMC Biol. 10: 14.

Specchia, V., L. Piacentini, P. Tritto, L. Fanti, R. D'Alessandro, G. Palumbo, S. Pimpinelli, and M. P. Bozzetti. 2010. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. Nature 463: 662-665.

Spreter, T., M. Pech, and B. Beatrix. 2005. The crystal structure of archaeal nascent polypeptide-

associated complex (NAC) reveals a unique fold and the presence of a ubiquitin-associated domain. J. Biol. Chem. 280: 15849-15854.

Sriram, S. M., B. Y. Kim, and Y. T. Kwon. 2011. The N-end rule pathway: emerging functions and molecular principles of substrate recognition. Nat. Rev. Mol. Cell. Biol. 12: 735-747.

Stirling, P. C., S. F. Bakhoum, A. B. Feigl, and M. R. Leroux. 2006. Convergent evolution of clamp-like binding sites in diverse chaperones. Nat. Struct. Mol. Biol. 13: 865-870.

Studer, R. A., R. A. Rodriguez-Mias, K. M. Haas, J. I. Hsu, C. Viéitez, C. Solé, D. L. Swaney, L. B. Stanford, I. Liachko, R. Böttcher, et al. 2016. Evolution of protein phosphorylation across 18 fungal species. Science 354: 229-232.

Taipale, M., D. F. Jarosz, and S. Lindquist. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat. Rev. Mol. Cell Biol. 11: 515-528.

Taipale, M., I. Krykbaeva, M. Koeva, C. Kayatekin, K. D. Westover, G. I. Karras, and S. Lindquist. 2012. Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. Cell 150: 987-1001.

Tian, G., I. E. Vainberg, W. D. Tap, S. A. Lewis, and N. J. Cowan. 1995. Specificity in chaperonin-mediated protein folding. Nature 375: 250-253.

Tokuriki, N., and D. S. Tawfik. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459: 668-673.

Tomala, K., and R. Korona. 2008. Molecular chaperones and selection against mutations. Biol. Direct 3: 5.

Ubersax, J. A., and J. E. Ferrell, Jr. 2007. Mechanisms of specificity in protein phosphorylation. Nat. Rev. Mol. Cell. Biol. 8: 530-541.

Ueno, T., H. Taguchi, H. Tadakuma, M. Yoshida, and T. Funatsu. 2004. GroEL mediates protein folding with a two successive timer mechanism. Mol. Cell 14: 423-434.

Ulrich, H. D., and H. Walden. 2010. Ubiquitin signalling in DNA replication and repair. Nat. Rev. Mol. Cell. Biol. 11: 479-489.

Valas, R. E., and P. E. Bourne. 2008. Rethinking proteasome evolution: two novel bacterial proteasomes. J. Mol. Evol. 66: 494-504.

Varshavsky, A. 2011. The N-end rule pathway and regulation by proteolysis. Protein Sci. 20: 1298-1345.

Varshavsky, A. 2019. N-degron and C-degron pathways of protein degradation. Proc. Natl. Acad. Sci. USA 116: 358-366.

Viéitez, C., B. P. Busby, D. Ochoa, A. Mateus, D. Memon, M. Galardini, U. Yildiz, M. Trovato, A. Jawed, A. G. Geiger, et al. 2022. High-throughput functional characterization of protein phosphorylation sites in yeast. Nat. Biotechnol. 40: 382-390.

Wang, J. D., C. Herman, K. A. Tipton, C. A. Gross, and J. S. Weissman. 2002. Directed evolution of substrate-optimized GroEL/S chaperonins. Cell 111: 1027-1039.

Wang, Y., W. Y. Zhang, Z. Zhang, J. Li, Z. F. Li, Z. G. Tan, T. T. Zhang, Z. H. Wu, H. Liu, and Y. Z. Li. 2013. Mechanisms involved in the functional divergence of duplicated GroEL chaperonins in *Myxococcus xanthus* DK1622. PLoS Genet. 9: e1003306.

Warnecke, T. 2012. Loss of the DnaK-DnaJ-GrpE chaperone system among the Aquificales. Mol. Biol. Evol. 29: 3485-3495.

Warnecke, T., and E. P. Rocha. 2011. Function-specific accelerations in rates of sequence evolution suggest predictable epistatic responses to reduced effective population size. Mol. Biol. Evol. 28: 2339-2349.

Weissenbach, J., J. Ilhan, D. Bogumil, N. Hülter, K. Stucken, and T. Dagan. 2017. Evolution of chaperonin gene duplication in stigonematalean cyanobacteria (subsection V). Genome Biol. Evol. 9: 241-252.

Williams, T. A., and M. A. Fares. 2010. The effect of chaperonin buffering on protein evolution. Genome Biol. Evol. 2: 609-619.

Wong, P., and W. A. Houry. 2004. Chaperone networks in bacteria: analysis of protein homeostasis in minimal cells. J. Struct. Biol. 146: 79-89.

Wyganowski, K. T., M. Kaltenbach, and N. Tokuriki. 2013. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. J. Mol. Biol. 425: 3403-3414.

Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, et al. 2011. Global analysis of protein activities using proteome chips. Science 293: 2101-2105.

Figure 14.1. An idealized cross-sectional view of GroEL-assisted protein folding. The GroEL chaperone has a double barrel (back-to-back) shape, with each barrel consisting of a ring of seven subunits. Loading of the top barrel with a client protein combined with seven ATPs leads to binding of the lid (GroES) and release of the processed protein from the bottom barrel.

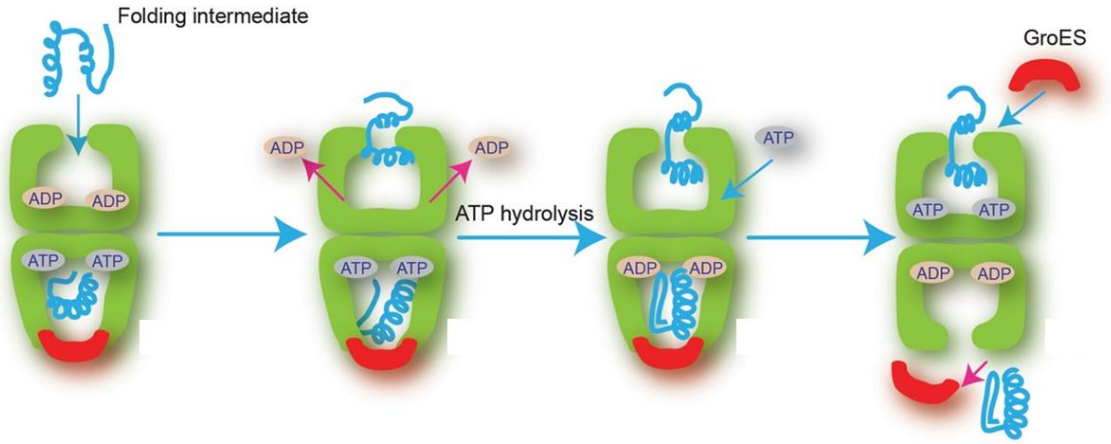

Figure 14.2. An idealized view of the recognition "sequence space" for client proteins of a chaperone. The central red dot represents the position of the chaperone relative to the recognition profiles of its various client proteins. The closer the chaperone is to a hypothetical client protein within this space, the better the recognition, with the black circle denoting the minimum distance necessary for recognition. In the upper panel, the chaperone has just one client protein, so the pair is free to wander through sequence space, so long as the matching specificity is kept within the minimum limit (denoted by the dashed arrows). In the lower panel, the chaperone has four client proteins, and this prevents the chaperone sequence from wandering, as any improvement with respect to one client protein is likely to reduce the affinity for others, e.g., movement of the red dot (chaperone) towards a particular client protein (other colors).

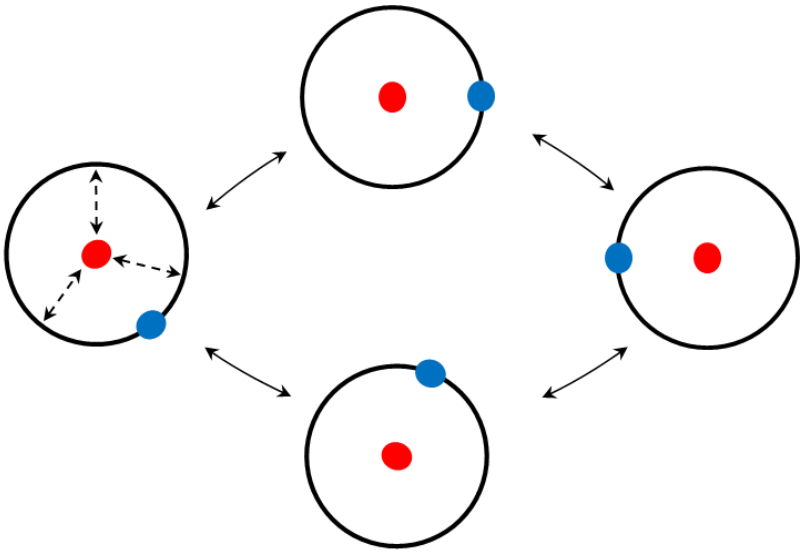

Figure 14.3. The ubiquitin-proteasome system for protein degradation. E1, E2, and E3 are enzymes involved in sequestering and covalently conjugating ubiquitin moieties to specific sites on a target protein, with the build-up of polyubiquitin chains serving as a signal for the recognition and processive degradation of the protein by the proteasome.

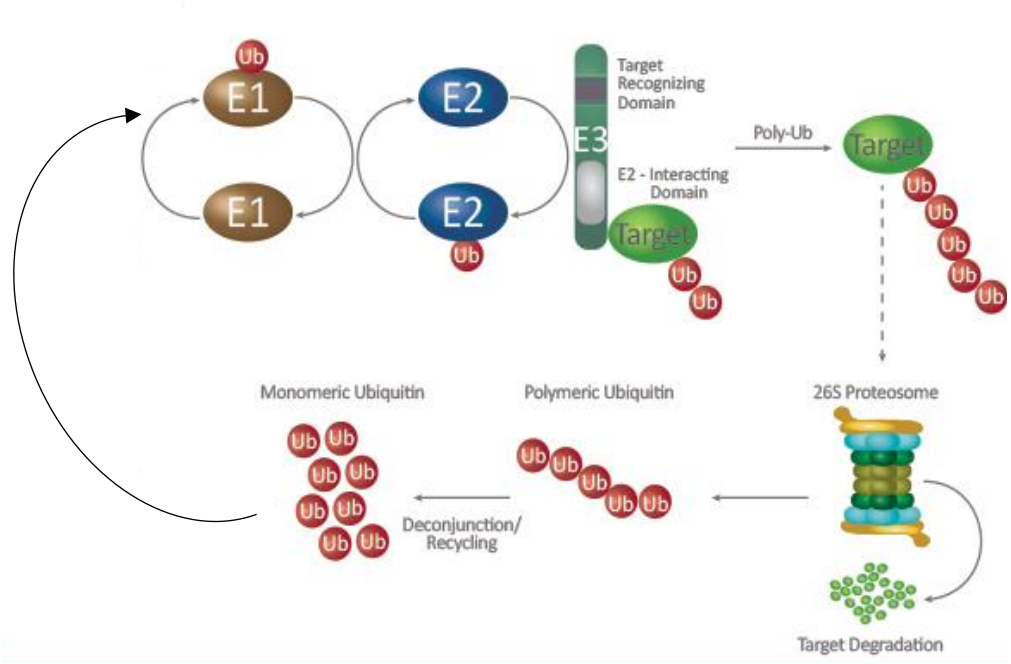

Figure 14.4. Some of the N-end rules for the acquisition of polyubiquitin signals that mediate processive protein-degradation by the proteasome (known from yeast, land plant, and mammalian cells). In the lower panel, the tertiary signal asparagine (N) is converted to the secondary signal aspartic acid (D) by deamidation, and likewise for glutamine (Q) to glutamic acid (E), and then a transferase adds arginine (R, the primary destabilizing residue), thereby enabling the recognition and marking of the resultant protein by ubiquitin ligases. From Graciet and Wellmer (2010) and Varshavsky (2011, 2019).

Primary destabilizing and stabilizing residues:

Modification and addition of destabilizing residue:

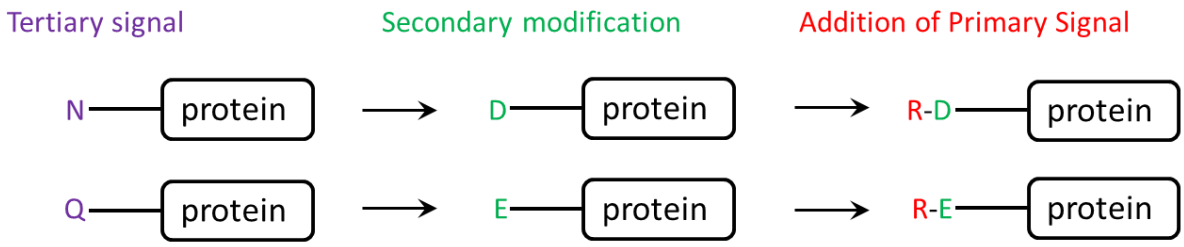

Figure 14.5. A simplified view of some of the challenges to the evolution of a heteromeric ring molecule. When the first (red) variant of the monomeric subunit appears (either as an allelic variant, or as a duplicate gene), prior to sufficient sequence divergence, the two types are likely to form a diversity of hetero-oligomeric structures within a cell. If a pair of sufficiently distinct interfaces can be established, an organized architecture involving alternating subunits might be acquired, e.g., alternating white and red subunits in the case of an even-mer. A ring with an odd number of subunits imposes additional challenges; for example, in the first step (with two subunit types), the positions cannot be evenly divided between two monomeric subtypes. A final structure involving eight or nine distinct members of the ring requires several additional gene duplications followed by the evolution of two distinct binding interfaces by each monomeric subunit, with each step introducing stoichiometric challenges.

Transition steps from a homomeric to a heteromeric even-mer (8 subunits):

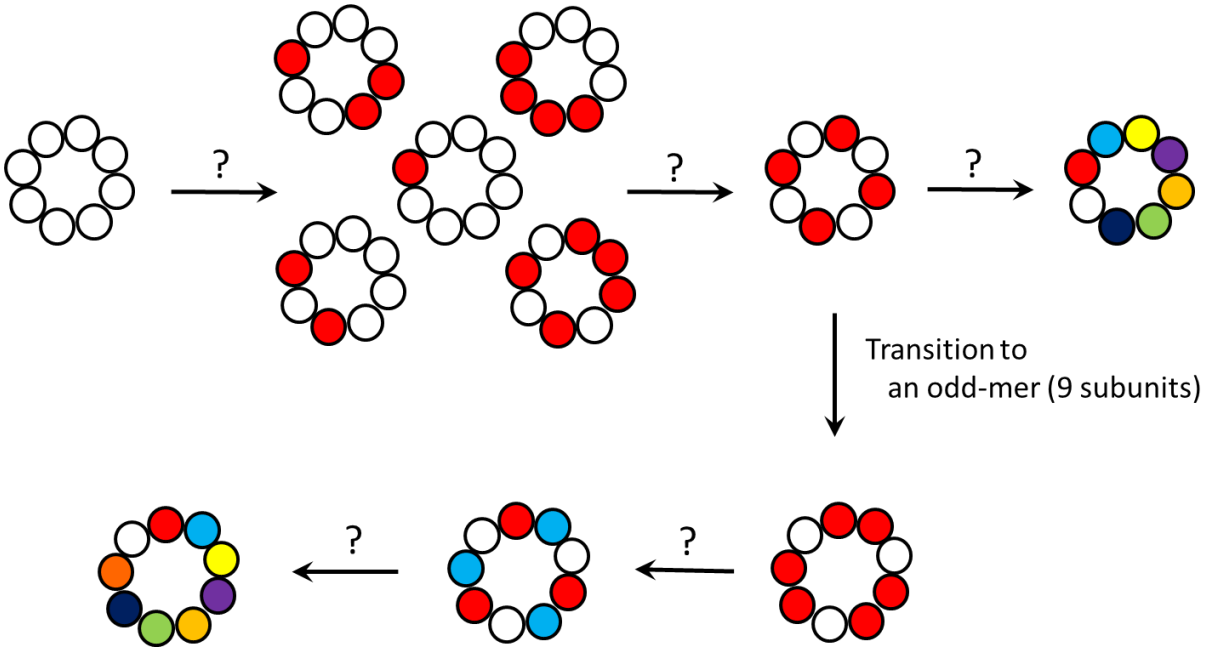

Figure 14.6. Top) Schematic for the transition rates (terms on arrows) between adjacent classes under the sequential-fixation model for the case of $\ell = 4$ sites. Under this model, transitions are rare enough (owing to small enough population sizes and/or mutation rates) that populations essentially always reside in pure states relative to the much less common polymorphic transition periods. Mutation rates towards + and− alleles are denoted by u_{01} and u_{10}, respectively, and ϕ_{xy} denotes the probability of fixation of a new mutation of state y arising in a population of state x. **Bottom)** Equilibrium haplotype (genotype) distributions for four effective population sizes (N_e), given for the situation in which the capacity of the system is $\ell = 20$ sites, and selection is of a stabilizing nature with optimum genotypic value (for the number of + alleles) $\theta = 7.0$ and width of the fitness function $\omega = 5000$. The mutation rate in the direction of − alleles is assumed to be $10\times$ that in the opposite direction. Results are derived by use of Equations 14.2.2 and 14.2.4. As $N_e \to \infty$, the mean genotypic value converges on the optimum (in this case, $\theta = 7$), although even at very large N_e, substantial variation among population states can remain. Note that only integer values of the average value of m (dots) are possible under the sequential-fixation model, and these are connected by lines merely for visualization.