æ

13. PROTEIN MANAGEMENT

28 October 2018

Some of the more fundamental features of proteins were reviewed in the two preceding chapters, with a primary focus on their structure in the native state. We now move on to additional issues central to the life histories of proteins, most notably matters associated with folding assistance, post-translational modification, and protein disposal. Whereas much attention has been given to transcriptional control of gene expression, the latter two phenomena provide additional dimensions to such management.

As noted in Chapter 9, many small proteins are capable of folding on their own without any external physical assistance or energy source. Such proteins must be endowed with amino-acid sequences carrying all of the "information" essential to acquiring proper three-dimensional structures. However, in virtually all organisms, numerous proteins require some form of folding assistance supplied by a specific set of 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 that might lead to cellular malfunctions. Properly folded proteins must often be disposed of after completing their missions, and such selective disposal invovles particular markings directing delivery to the cellular disposal machinery. Still other post-translational markings on proteins are specifically utilized as modifiers for particular subcellular functions.

While outlining general aspects of protein management, this chapter also provides numerous examples of the evolutionary properties of cellular features that reinforce the principles outlined in preceding chapters. For example, much of the machinery associated with protein folding assistance and disposal consists of higher-order multimers that have frequently changed with respect to subunit number and type (Chapter 12). The coevolution of chaperones and their client genes raise issues of how a cellular feature with multiple substrates might become constrained by a "jack of all trades, master of none" syndrome; and once reliant on chaperone-assisted folding a client protein may experience relaxed selection on self-folding capacity, and hence become locked into a state of dependency (Chapter 10). The sites of post-translational markings appear to be free to wander evolutionarily over protein surfaces, providing a means for the effectively neutral re-wiring of regulatory mechanisms (Chapters 6 and 20).

Chaperone Assistance

Chaperones operate by confining individual client proteins within a protective environment and restricting the ways in which they can move. This lowers the energetic barrier necessary to achieve a stable folding configuration, while also minimizing the potential for harmful interactions of unfolded proteins with others in the cell. The widespread use of molecular chaperones across the Tree of Life poses numerous evolutionary questions, not all of which have been answered.

First, what are the mechanisms by which chaperones recognize their appropriate client proteins? Second, do certain classes of chaperones coevolve with individual client proteins in ways that make them less effective with others? Third, once a protein becomes reliant on chaperone assistance for proper folding, does this act as a evolutionary trap by further relaxing selection on features essential to self-folding? 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 to 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, possibly paving the way to the establishment of proteins longer than capable of self-folding. However, the substantial diversity of chaperone types within lineages also leads to the conclusion that these helper molecules have evolved more than once, often converging on similar molecular structures and mechanisms (Schilke et al. 2006; Stirling et al. 2006). Functional diversification following gene duplication and sub/neofunctionalization has occurred 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), likely driven by adaptive conflicts imposed by alternative client protein pools. Moreover, as will be outlined below, striking examples of evolutionary transitions between homomeric and hetermomeric chaperone structures exist.

Unfortunately, chaperone-family nomenclature is difficult to navigate, as the notation for orthologous genes is often inconsistent among lineages. To avoid this morass, an attempt is made below to simplify via a slight abuse of taxon-specific notation. Many chaperones are referred to as heat-shock proteins, owing to their typical elevation in expression at extreme temperatures (and other extreme conditions), and such labels are often followed by a number referring to the approximate size in kiloDaltons (a measure of mass, with one KD equivalent to ~ 7.5 amino acids), a notation that will be adhered to in a number of cases below. However, not all heat shock proteins are exclusively involved in protein folding, with some being more closely associated with protein degradation and/or disaggregation. Only the most well-studied chaperone families in the three major domains (bacteria, archaea, and eukaryotes) will be introduced here.

In the bacterial domain, 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. The molecular architectures of each of these classes are diverse, and they 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 func-

tion. Hsp40 (a tweezer-like dimer) acts as a cochaperone for Hsp70 (a monomer), binding exposed hydrophobic patches on unfolded proteins and targeting them to the latter, which stabilizes the protein in an ATP-dependent manner. Hsp70 can also operate as an "unfoldase," consuming ~ 5 ATPs per protein in the process (Sharma et al. 2010). Although this three-member system is extremely widespread among bacteria, at least one lineage appears to have lost it (Warnecke 2012).

The best studied of the bacterial chaperones is GroEL (more generally known as chaperonin 60, with the more specific name being that for the *E. coli* protein). GroEL has a large barrel-like structure, consisting of two heptameric rings (a homomer with all 14 subunits derived from the same genetic locus), stacked back to back (Figure 13.1). Each ring comprises a separate chamber within which the folding of individual client proteins proceeds after closure by a cochaperone lid (another heptamer, called GroES). GroEL/ES function involves a form of allostery, with cycles of enclosure and release, as the binding of ATPs to one ring result in the release of the GroES cap from the other. Substrate molecules are captured via interactions with their hydrophobic residues and then stretched and remodeled within the hydrophilic folding cage. Each round of turnover of a protein requires ~ 11 seconds and consumes seven ATPs, regardless of the substrate, one ATP for each of the subunits of the ring (Keskin et al. 2002; Ueno et al. 2004; Horwich et al. 2009). In *E. coli,* the half-time of assisted folding is ~ 45 sec (Kerner et al. 2005), suggesting that an average client protein engages in ~ 4 folding attempts before success is achieved.

GroEL is present throughout the entire bacterial phylogeny, with some species harboring multiple variants that likely have become subfunctionalized with respect to client proteins (Lund 2009; Henderson et al. 2013). However, a few bacterial species (e.g., some species of *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 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 possibly being 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 have secondary functions in a number of species, e.g., adhesion to host cells, secretion, DNA binding, cell-cell communication, and even toxicity (Henderson et al. 2013).

Many archaea, like bacteria, deploy chaperones in the Hsp40/Hsp70/Nucleotide Exchange Factor 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, and those containing them 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 dealing with 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) with substantial structural similarity (Details 13.1). As with bacterial GroEL, the archaeal CCT, also known as the thermosome, 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. The apical lids of CCT close in a manner similar to a camera iris, leaving a small opening, which may enable proteins too large to enter the chamber in their entirety to experience progressive folding by threading through the apical pore (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, prefoldin is a heterohexamer, consisting of two monomeric subunit types, that serves to transfer proteins to CCT.

Eukaryotes deploy several protein complexes for assistance 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. Unlike the situation in archaea, the eukaryotic prefoldin hexamer consists of six subunit types, and all of the monomeric subunits of eukaryotic CCT are also encoded by different proteins (Foundations 13.1). Eukaryotic organelles (mitochondria and chloroplasts) utilize eubacterial-derived orthologs of GroEL and GroES, 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. These have commonly diversified into a dozen or more copies in various eukaryotes, with even more Hsp40 proteins present (Craig and Marszalek 2011; Bogumil et al. 2014). Moreover, the ancestry is mixed, with some showing greater phylogenetic affinities to bacteria and others to archaea. As in bacteria, operation of the eukaryotic system generally involves a Hsp40 protein recruiting a client protein and then stimulating Hsp70 ATPase activity to capture and 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). Likewise, whereas a specific chaperone system operates as a chaperone in the mitochondrion, it is also involved in other functions, most notably 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 patchily distributed among various branches of the eukaryotic tree, and also have numerous secondary functions (Johnson and Brown 2009; Taipale et al. 2010; Johnson 2012).

If there are any generalities to emerge from this morass of complexity, it is that all of cellular life is dependent on protein complexes specifically assigned to protein folding. Yet, despite the essentiality of these conserved functions, the strutural constituents of chaperones have been modified on numerous occasions. This must have been accomplished in such a way that basic folding capacity remained uncompromised during such 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.

Foundations 13.1. The CCT complex. The CCT (chaperonin containing tailless complex) presents a striking example of a transition from a complex homomeric to heteromeric structure. Restricted to archaea and eukaryotes, CCT chaperonins are generally hexadecamers (occasionally octodecamers), i.e., with 8 or 9 monomeric subunits per barrel (Archibald et al. 1999, 2001). In archaea, the overall structure is generally either homomeric or heteromeric with two or three alternating subunits (in 8- or 9-component barrels, respectively). The evidence suggests that 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, but with effectively neutral evolution and no necessary change at the functional level (Archibald et al. 1999). However, the evolution of complexity is not unidirectional in CCT, as there are examples of both the reversion of heteromeric complexes to homomers.

In contrast, all of the subunits in eukaryotes are encoded by separate genetic loci. This 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 protein level at rates exceeding the neutral expectation, 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).

With eight different subunits per ring in the eukaryotic version of CCT, there are $8! = 40,320$ possible arrangements, and yet it is thought that just one assembly is consistently achieved in the cell (Kalisman et al. 2012). Understanding the evolution of an initially homomeric ring into such a complex heteromeric state imposes several challenges, as 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. Each step of the process also raises the problem of hetero-oligomerization – the production of heterogeneous mixtures of subunits in individual complexes that is likely to exist until a high level of interface specificity has evolved (Figure 13.2). If rings of only particular architectures are fully functional, such a condition would presumably impose a further barrier against the emergence of heteromeric complexes. The addition of each new member of the ring likely requires the fixation of at least two mutations, as such architectures require that each subunit be involved in two distinct interfaces. Understanding these kinds of issues is not peculiar to chaperone evolution, as numerous other cellular features have ring-like structures, e.g., the nuclear pore, the proteasome, and a number of DNA-binding proteins in eukaryotes.

Client-chaperone coevolution. Unlike enzymes, chaperones typically have a wide variety of client substrates. In *E. coli,* for example, at least 250 proteins appear to rely on GroEL for proper folding, while another 400 or so rely on Hsp40/70, about 170 are serviceable by both, and fewer than 20 are substantially influenced by Trigger Factor (Kerner et al. 2005; Fujiwara et al. 2010; Niwa et al. 2012). In yeast, $\sim 20\%$

of proteins are targeted by Hsp90 alone (Taipale et al. 2010).

Such a vast repertoire of substrates raises questions about the degree to which the features of chaperone systems are compromised by the numbers of client genes, the issue being that any evolutionary movement toward a better fit of one client may diminish the effectiveness with others (Lynch and Hagner 2015). Wang et al. (2002) acquired some insight into this matter by engineering *E. coli* to carry a foreign GFP (green fluorescent protein, carried on a plasmid) and then imposing a selective challenge on cultures to improve GFP folding into functional molecules. This resulted in the advancement 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 (Figure 13.3).

Eukaryotic chaperone dependence of the production of monomeric subunits of both actin and tubulin represents another potential example of such a compromise. Together, these molecules form the cytoskeleton, serve as highways for the transport of various cargos, and have roles in numerous other central cellular functions (Chapter 15). 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 both actin and tubulin, it is incapable of reducing them to their native conformations (Tian et al. 1995). Moreover, unlike the situation with GroEL, where proteins are processed in a chamber with hydrophilic walls, folding assistance by CCT involves binding of the substrate to the apical domains of the internal chamber. Given their centrality as primary clients, it is plausible that fine-tuning for processing actin and tubulin imposes strong constraints on the capacity to assist alternative substrates.

A third example of chaperones coevolving with client features involves populations experiencing 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 protein-sequence evolution and possible deleterious-mutation accumulation 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 driving force is adaptive remodeling of key chaperone motifs in response to major mutations in specific client genes, and/or whether chaperone over-expression is an evolutionary compensation for its own reduced catalytic capacity.

Given their large size, high baseline 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, and experimental data do suggest that bacterial cells respond physiologically to the accumulation of deleterious mu-

tations by up-regulating GroEL expression, which presumably results in improved protein folding and stability. Maisnier-Patin et al. (2005) found such an increase in GroEL expression in lines of *Salmonella* after accumulating random mutations, with additional artificial enhancement of GroEL expression resulting in still further increase in fitness. Similarly, Fares et al. (2002) found that lines of *E. coli* allowed to accumulate enough mutations to reduce fitness by $\sim 50\%$ were restored to $\sim 90\%$ fitness following the overexpression of GroEL; and this type of observation extends to Hsp70 (Aguilar-Rodriguez 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).

Selection for improved client-protein folding may not be the only constraint on the architectural features of chaperones. Most notably, owing to their roles as safe havens for protein assembly, chaperones are vulnerable to exploitation by foreign agents. 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). Host-encoded chaperones are essential to the morphogenesis of many other bacteriophage (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, selection to avoid processing proteins associated with cellular parasites may directly conflict with selection for efficient handling of a cell's endogenous proteins.

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 scenario outlined in Chapter 10, with a reliable mechanism of assisted folding, mutations that would otherwise prevent self-folding of a protein would be expected to accumulate. However, 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).

The precise mechanisms by which chaperones identify their client proteins remains poorly understood, but will be central to understanding these kinds of processes 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 unfolded, but that such regions tend to be flanked with positively charged (arginine, lysine, and proline) that are targets of chaperones. Less clear is whether the latter sequences arose in response to the accumulation of aggregative sequences, or appeared first and simply paved the way for the safe accumulation of otherwise adhesive residues. If the former, the challenging question is why selection should not minimize the accumulation of aggregative 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 such a way as to modify the course of evolution. One extreme view is that chaperones might facilitate adaptive evolution by buffering the normally deleterious effects of mutant alleles, thereby encouraging effectively neutral build-up of a cryptic load of variation. Such variation might then be exposed if the chaperone system becomes overwhelmed in a stressful environment, leading to the suggestion that chaperones can act as "capacitors" for evolutionary change by fostering the expression of novel variants in extreme environments (Rutherford and Lindquist 1998). Further imagining that such environments are precisely the ones within which aberrant phenotypes are most likely to have utility, the speculation is that chaperones enhance the ability of populations to adapt to extreme selective challenges. If sustained, this might eventually lead to the constitutive expression of the previously suppressed variant, moving a population into an entirely new phenotypic domain.

Although this kind of narrative fuels the idea that various biological features have emerged specifically to enhance the evolvability of species, it does not follow that a cellular trait that influences the evolutionary flexibility of a species owes its existence to selection for such purposes. There is no evidence that any chaperone has been promoted for anything other than the immediate cellular benefits of protein management. Pusing most genes beyond their capacities will lead to aberrant, pathological phenotypes, so there is nothing particularly unique about the consequences of overtaxed chaperones.

The idea that chaperone expression is fine tuned so as to enhance long-term evolutionary potential is confronted with several other issues (Levy and Siegal 2008; Tomala and Korona 2008; Siegal and Masel 2012). 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, one must inquire how a variant that is not expressed for considerable periods of time can avoid the neutral accumulationi of even 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 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. Finally, for the entire scenario to work, chaperone stress must last long enough to keep the extreme phenotype exposed to selection to mutationally acquire a mechanism for constitutive expression, but short enough not to avoid population extinction.

Implicit in the argument that compromised chaperone capacity leads to a release of latent variation is the assumption that chaperones normally serve to enhance phenotypic robustness by buffering the effects of new mutations. In fact, the empirical evidence suggests otherwise. In yeast, whereas the effects of standing variation are muted by chaperone activity, the phenotypic effects of *de novo* mutations are actually magnified on average (Geiler-Samerotte et al. 2016). Thus, natural selection differentially promotes alleles whose effects are buffered by chaperones, creating the illusion that outbreaks of variation caused by compromised chaperones are promoted

by selection.

Although chaperones have not been advanced to enhance long-term evolvability, they may nonetheless play 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, it was possible to control the expression of GroEL, and hence 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, with subsequent return to a low level of GroEL imposing 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 by Tokuriki and Tawfik (2009) involving repeated rounds of mutagenesis and selection in the presence of overexpressed GroEL led to the improvement of the catalytic performance of other enzymes at the expense of self-folding capacity.

This kind of experimental result, reliant on a highly contrived situation – alternating periods of high and low GroEL expression, and selection on a nonessential protein, needs to be tempered with the kinds of 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 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-Rodriguez et al. 2016; Kadibalban et al. 2016), such a pattern could also exist for reasons unassociated with folding. Thus, for natural client proteins exposed to native chaperone expression levels, it remains unclear whether chaperone-dependence alters the evolutionary trajectories of protein function.

In summary, all of the preceding observations strongly support the view that the function of chaperones promoted directly by natural selection is to suppress the negative phenotypic consequences of deleterious mutations rather than to store away hopeful monsters. There is no known 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. Assuming the energetic cost is not too great, any mechanism that can sufficiently suppress the negative fitness consequences of deleterious mutations and/or increase the robustness of an organism to other perturbations can be selectively favored (de Visser et al. 2003). However, it does not follow that the assimilation of such a mechanism into a species has any long-term benefits, and there may even be disadvantages (Frank 2007; Gros and Tenaillon 2009; Lynch 2012). Although selection for a robustness-enhancing feature may serve to partially 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. In this sense, the idea that evolution has produced fundamentally superior organisms with layers and layers of buffering mechanisms to stabilize high

fitness is an illusion.

Disposal by Proteasomes

All organisms are confronted with the challenge of disposing of proteins that are structurally aberrant (owing to improper folding), functionally inappropriate (owing to the prior completion of their tasks), or damaged by a wide variety of intracellular effects (such as thermal denaturation and oxidation). Most prokaryotes and possibly all eukaryotes harbor a special molecular machine, the proteasome, to carry out such processes 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 confining protease activity on target proteins to the interior of the structure, thereby protecting desirable proteins from proteolysis.

The proteasome also 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 a pore through which cargoes are delivered. In most archaea, the two inner β active rings are monomeric, comprised of catalytic subunits encoded by a single locus, and 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, but only six of the seven subunits on the internal ring carry active sites (Puhler et al. 1994). Consistent with the hypothesis that substantial gene duplication (perhaps even whole-genome duplication) occurred on the branch between FECA and LECA, based on their dispersion across the entire eukaryotic tree, the origin of all fourteen distinct subunits predates LECA (Bouzat et al. 2000).

The bacterial proteasome is generally comprised of two homomeric rings with six subunits, although archaeal-like structures with seven identical subunits are found sporadically throughout the bacterial domain (Valas and Bourne 2008; Fuchs et al. 2017, 2018). Thus, although varying in structure, the proteasome dates back to LUCA, and we are again confronted with both an increase in the complexity and an expansion in the number of subunits of the eukaryotic proteasome, which requires an evolutionary alteration of binding interfaces, combined with the loss of catalytic ability of one of the subunits. In parallel with this shift in complexity, the proteasome regulator proteins in control of the entry of cargo proteins consist of at least six different subunit types throughout eukaryotes but only one in archaea (Fort et al. 2015).

Numerous other proteases (peptidases) besides the proteasome operate in both prokaryotic and eukaryotic cells (e.g., Clausen et al. 2011), many of which complete the degradational process, as the proteasome only degrades substrates to short oligopeptides. Moreover, a parallel set of machinery, the exosome and its regulatory proteins, exists for the disposal of 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 polypeptides).

The targeting of proteins for removal is generally orchestrated by pathways dedicated to marking molecules with specific degradation signals. In eukaryotes, the most prominent mechanism by far is the ubiquitinylation pathway (Mogk et al.

2007; Sriram et al. 2011; Varshavsky 2011). In a series of three enzymatically guided steps, ubiquitin is delivered and ligated to specific lysine residues on target molecules in an ATP-dependent process. Chains of polyubiquitin are further grown by internal binding of these molecules to each other, providing a signal for proteasome delivery. Deubiquitinylation occurs prior to entry into the proteasome, sparing the ubiquitin molecules from degradation. All of the components of this pathway are found in some lineages of archaea, implying a pre-LECA origin, apparently with independent expansions and specialization of component parts in animals and land plants (Grau-Bové et al. 2015). Pathways with essentially the same features but quite different (and in some cases possibly unrelated) molecular participants exist in bacteria (Mogk et al. 2007; Mukherjee and Orth 2008), so an even earlier origin cannot be ruled out.

In addition to its central role in protein degradation, ubiquitinylation combined with deubiquitinylation provides a means of dynamically switching proteins between alternative functional 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).

This ubiquitin-proteasome degradation pathway provides still another example of the importance of intracellular molecular languages in guiding key cellular events: 1) specific amino acids at the N-end of molecules define their susceptibility to ubiquitinylation; 2) specific internal sites of the target molecules are post-translationally modified by ubiquitin addition; and 3) the marked regions are read as indicators for proteasome delivery.

The recognition determinants for protein degradation generally consist of single amino acids terminating the N-end of the molecule, each serving as a stabilizing or destabilizing signal (Figure 13.5). The nature of these determinants (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 initially terminal methionine residues, other endopeptidases for severing small N-terminal peptide chains (thereby exposing new degradation determinants), and still others for converting some amino acids to others (e.g., Asn and Gln to Asp and Glu, respectively, in eukaryotes). The latter residues can be viewed as tertiary/secondary degradation determinants, as they are only effective after modification, and even then so often only after attachment of another amino acid that serves as the primary determinant (Arg in the case of eukaryotes). In a related eukaryotic pathway, acetylation of specific N-end residues provides a signal for degradation (Hwang et al. 2010; Shemmory et al. 2013). In the case of *E. coli*, Arg and Lys serve as secondary destabilizing factors, which become active after an attachment of Leu. It has been suggested that the Arg transferase utilized in eukaryotes is related to the Leu transferase in bacteria (Graciet et al. 2006).

Ubiquitinylation itself is guided by ubiquitin ligases, which recognize specific amino-acid sequence motifs, usually with a Lys serving as the ligation site. Dozens to hundreds of such proteins with unique recognition sequences are often encoded within individual genomes, providing specificity to the overall system. Together,

these two signals, one for stabilization/destabilization signals and the other for ligase recognition, determine the half lives of individual proteins. As in the case of phosphorylation sites, the lysine residues specifically designated for ubiquitylation are only slightly more conserved over evolutionary time, again suggesting a high degree of redundancy with respect to location but stabilizing selection on the total numbers of such sites per protein (Hagai et al. 2012; Lu et al. 2017).

Although bacteria have a pathway for protein disposal quite like that of eukaryotes, including the use of N-end rules (albeit involving some different key amino acids), the overall system in bacteria is substantially simpler than in eukaryotes. Varshavsky (2011) has suggested that the latter may be no more efficient than the former, with "overdesign" in the eukaryotic system resulting from effectively neutral processes that can drive the increase in complexity during phases of reduced effective population sizes. A number of the destabilizing N-terminal amino acids in eukaryotes and the same as those in bacteria, suggesting a common ancestry of this degradation system in these two phylogenetic lineages.

Post-translational Modification

A key stage in the life histories of many proteins is the post-translational covalent linkage of small molecules to a small fraction of amino-acid residues. The number of known modifications is substantial, and includes additions of small molecules such as phosphate, adenyl, acetyl, and amide groups, to larger molecules such as sugars and fatty-acid chains, to entire proteins such as ubiquitin (as just discussed). It appears that all amino acids can participate in such interactions, the exact residue depending on the particular binding group and cellular pathway. Although the functions of such markings are known in just a few cases, post-translational modification can lead to changes in structure, stability, localization, and/or function of the affected proteins. Thus, although the classical view of gene regulation focuses on expression modification at the level of transcription (Chapter 20), post-translational modification adds a number of additional dimensions in both prokaryotes and eukaryotes. There are many similarities between transcription-level and phosphorylation-level regulation: both involve *trans*-regulating proteins with interactions specific to simple *cis*-binding sites; both are subject to divergence with nonfunctional consequences; and both are typically under some form of purifying selection.

Although post-translational modification is largely unchartered territory for the field of evolutionary biology, one major target of inquiry involves phosphorylation. Additions of phosphoryl groups are generally restricted to serine and threonine residues, but extend to tyrosine in animals, and arginine, aspartic acid, cysteine, and histidine in bacteria. Covalent attachment of PO_4^{3-} subgroups is generally carried out by specialized enzymes called kinases, most of which have simple recognition sites comprised of the substrate amino acid plus just two to four flanking residues (Ubersax and Ferrell 2007; Ochoa et al. 2018). Such simplicity raises the opportunity for substantial promiscuity, often rendering inferences of functional significance from the simple observation of a phosphorylation site 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

(Chapter 16).

The immediate effect of phosphorylation is the addition of a negative charge to the amino-acid substrate, which will often have downstream effects including protein activation or inhibition. Protein phosphorylation has the additional property of being rapidly reversible by use of specific phosphatases (Chapter 21). Substantial genomic resources are invested in post-translational modification. For example, $\sim 2\%$ of the yeast genome is devoted to protein kinases (Zhu et al. 2000), with $\sim 10,000$ phosphosites distributed over $\sim 2,300$ proteins. Over 500 kinases and 200 phosphatases are encoded in the human genome (Manning et al. 2002; Alonso et al. 2004).

Although the functions of most modified sites are unknown, proteome-wide comparative data provide insights into the long-term evolutionary stability of modified residues. 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, suggesting that not all such sites are simply evolving neutrally (Landry et al. 2009; Ba and Moses 2010). There is also evidence that phosphorylation sites undergo subfunctionalization following gene duplication (Amoutzias et al. 2010; Freschi et al. 2011; Kaganovich and Snyder 2011), with each member of a paralogous pair partitioning up the ancestral sites, although the functional significance of this remains unclear. Nonetheless, a large fraction of phosphosites appear free to vary among species in terms of status and location (Moses et al. 2007; Holt et al. 2009; Freschi et al. 2014; Studer et al. 2016). For example, only $\sim 5\%$ of all yeast phosphorylation sites appear to have been conserved across the entire lineage (dating back ~ 700 million years). Even when the same phosphorylatable residue is present in two moderately related species, their phosphorylation status may differ.

Phosphorylation sites are typically clustered on the surface of a protein or in disordered regions, and the critical feature may simply be the acquisition of an adequate local charge. Notably, aspartic-acid and glutamic-acid residues, which are naturally charged positively, serve as potential replacements (and/or sources) for their phosphorylatable counterparts (although interconverions require two nucleotide substitutions), 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 essentially a quantitative trait under stabilizing selection for an appropriate positive charge with the specific locations of many of the affected residues free to wander in an effectively 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 under stabilizing selection, but enough degrees of freedom that there can be considerable turnover of the specific phosphosites (Foundations 13.2).

Foundations 13.2. The evolution of a digital trait. Phosphorylation and other post-translational modifications are examples of digital traits, in the sense that they have a very simple molecular basis with the possible phenotypes taking on integer values (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 selection to promote higher-fitness states. As will be discussed below, this is virtually always the case, even 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 11, 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, 4, and 5), with multiplicities 1, 4, 6, 4, and 1, respectively (Figure 13.6). These multiplicities are equivalent to the coefficients of a binaomial 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 11.2, 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 internal cellular traits have functions (and cytoplasmic environments) that have remained relatively stable for tens to thousands of millions of years (even in the face of a changing external environment).

In general, the probabilities of alternative states depend on the relative magnitudes of the transition coefficients between adjacent classes (Figure 13.6). 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 the + state, and vice versa, are defined to be u_{01} and u_{10}, respectively. A haploid, nonrecombining population is assumed here, so each set of functionally equivalent states comprises a genotypic class. The probability of fixation is given by the standard expression outlined in Chapter 8.

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,

$$\widetilde{P}_{n,m} = \binom{\ell}{m} \eta^m (1-\eta)^{\ell-m}. \tag{13.2.1}$$

Thus, in this limiting case, the probability distribution for the class types only depends on the ratio of mutation rates, not on their absolute values, and on the binomial coefficient associated with each class, which defines the multiplicity of 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 11.2 – it is the ratio of fixation probabilities from class $m-1$ to m and vice versa.

The overall distribution can then be written as

$$\widetilde{P}_m = \widetilde{P}_{n,m} \cdot e^{2N_e s_m} = C \cdot \binom{\ell}{m} \beta^m e^{2N_e s_m}, \tag{13.2.2}$$

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 (13.2.1) 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

$$\mu_m = \sum_{m=0}^{\ell} m \cdot \widetilde{P}_m, \tag{13.2.3}$$

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

As a specific example of the application of Equation (13.2.2), 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,

$$W_m = e^{-(m-\theta)^2/(2\omega^2)}, \tag{13.2.4}$$

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 $s_m = W_m - W_0$.

Application of the Gaussian fitness function to Equation (13.2.2), one example of which is shown in Figure 13.6, 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 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 (13.2.1). 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

in state θ. This selection limit assumes that θ is an integer, and 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 commonly 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, the probability of any specific activated site in one population not being activated in another in state m is $(\ell - m)/\ell$.

Literature Cited

Abascal, F., A. Corpet, Z. A. Gurard-Levin, D. Juan, F. Ochsenbein, D. Rico, A. Valencia, and G. Almouzni. 20913. 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.

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., et al. 2000. Phylogenomic analysis of the alpha proteasome gene family from early-diverging eukaryotes. J. Mol. Evol. 51: 532-543.

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.

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. 2011 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., 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.

Doyle, S. M., O. Genest, and S. Wickner. 2013. Protein rescue from aggregates by powerful molecular chaperone machines. Nat. Rev. Mol. Cell. Biol. 14: 617-629.

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(1): 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. Aminoacyltransferases 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, Azia A, Levy Y. 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. Philos. 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.

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.

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., 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. GelfandS. 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.

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., 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. 2014. 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. 2913. 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.

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.

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.

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.

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, Fares MA. 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.

Sabate, R., N. S. de Groot, and S. Ventura. 2010. Protein folding and aggregation in bacteria. Cell. Mol. Life Sci. 67: 2695-2715.

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.

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., et al. 2016. Evolution of protein phosphorylation across 18 fungal species. Science 354: 229-232.

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.

Toyama, B. H., and M. W. Hetzer. 2013. Protein homeostasis: live long, won't prosper. Nat. Rev. Mol. Cell. Biol. 14: 55-61.

Tyedmers, J., A. Mogk, and B. Bukau. 2010. Cellular strategies for controlling protein aggregation. Nat. Rev. Mol. Cell. Biol. 11: 777-788.

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.

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., et al. 2011. Global analysis of protein activities using proteome chips. Science 293: 2101-2105.

Figure 13.1. An idealized cross-sectional view of protein processing by the GroEL chaperone. This molecular machine has a double (back-to-back) barrel shape, with each barrel consisting of a ring of seven monomeric units. Loading of the top barrel with a client protein combined with seven ATPs results in binding of the lid (GroES) and release of the processed protein from the bottom barrel.

Figure 13.2. 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. 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 (above). 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 stoichiometry issues.

Figure 13.3. 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 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. 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.

Figure 13.4. The ubiquitin-proteasome pathway for protein degradation. E1, E2, and E3 are enzymes involved in sequestering and transferring ubiquitin molecules to specific sites on a target protein, with the build-upinto poly-ubiquitin chains serving as a signal for harvest and partial digestion by the proteasome.

Figure 13.5. Some of the N-end rules for the acquisition of protein-degradation signals 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), endowing the protein with a destabilizing residue. From Graciet et al. (2010) and Varshavsky (2011).

Figure 13.6. Top) Schematic for the transition rates between adjacent classes under the sequential-fixation model for the case of $\ell = 4$ sites. **Bottom)** Equilibrium genotype distributions for four effective population sizes, given for the situation in which the capacity of the system is $\ell = 20$, 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 (13.2.2) and (13.2.4).

Figure 13.1

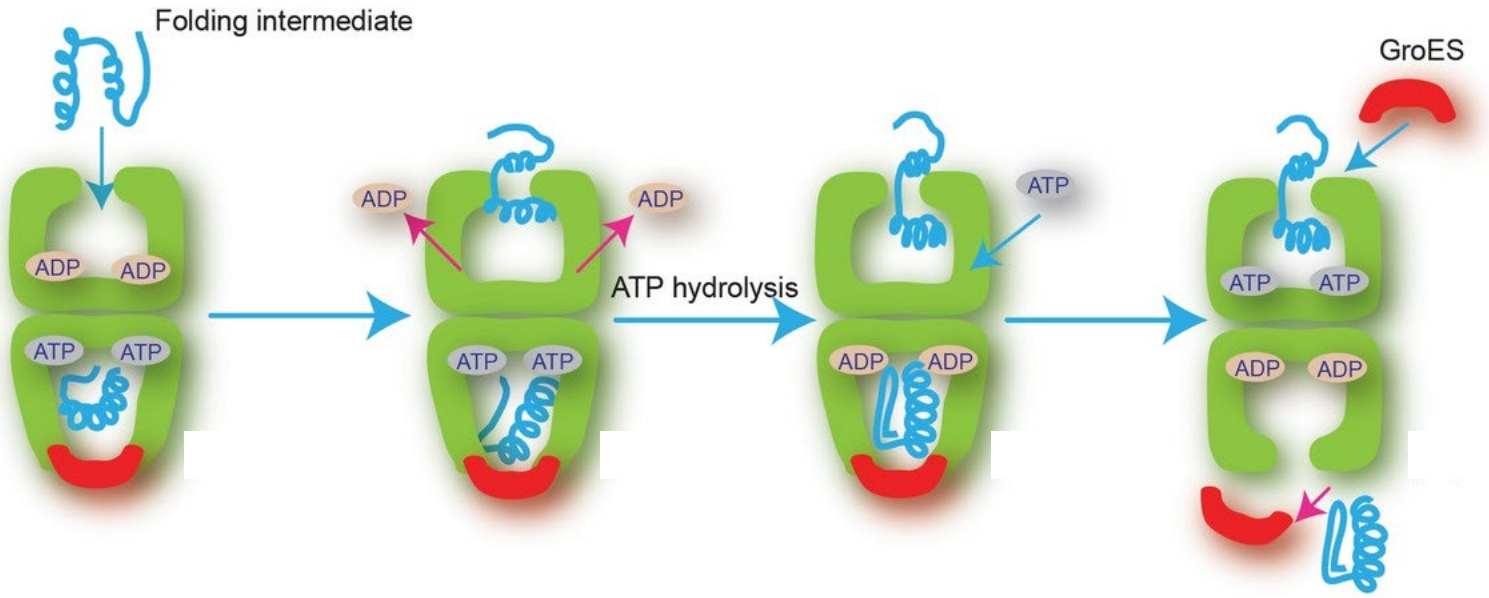

Folding intermediate
ADP
ADP
ATP
ATP
ADP
ADP
ATP
ATP
ATP hydrolysis
ATP
ADP
ADP
GroES
ATP
ATP
ADP
ADP

Figure 13.2

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

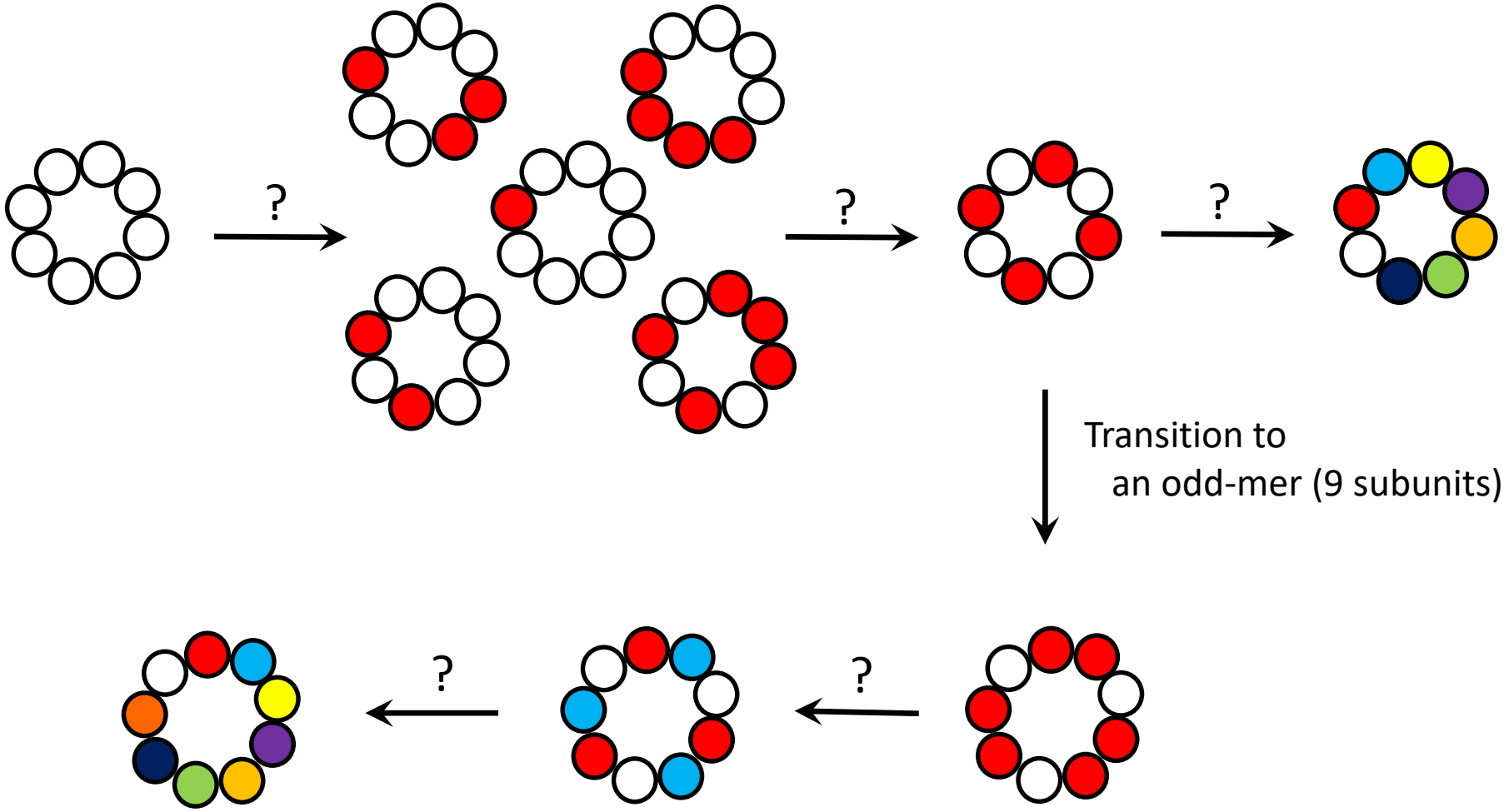


Figure 13.3

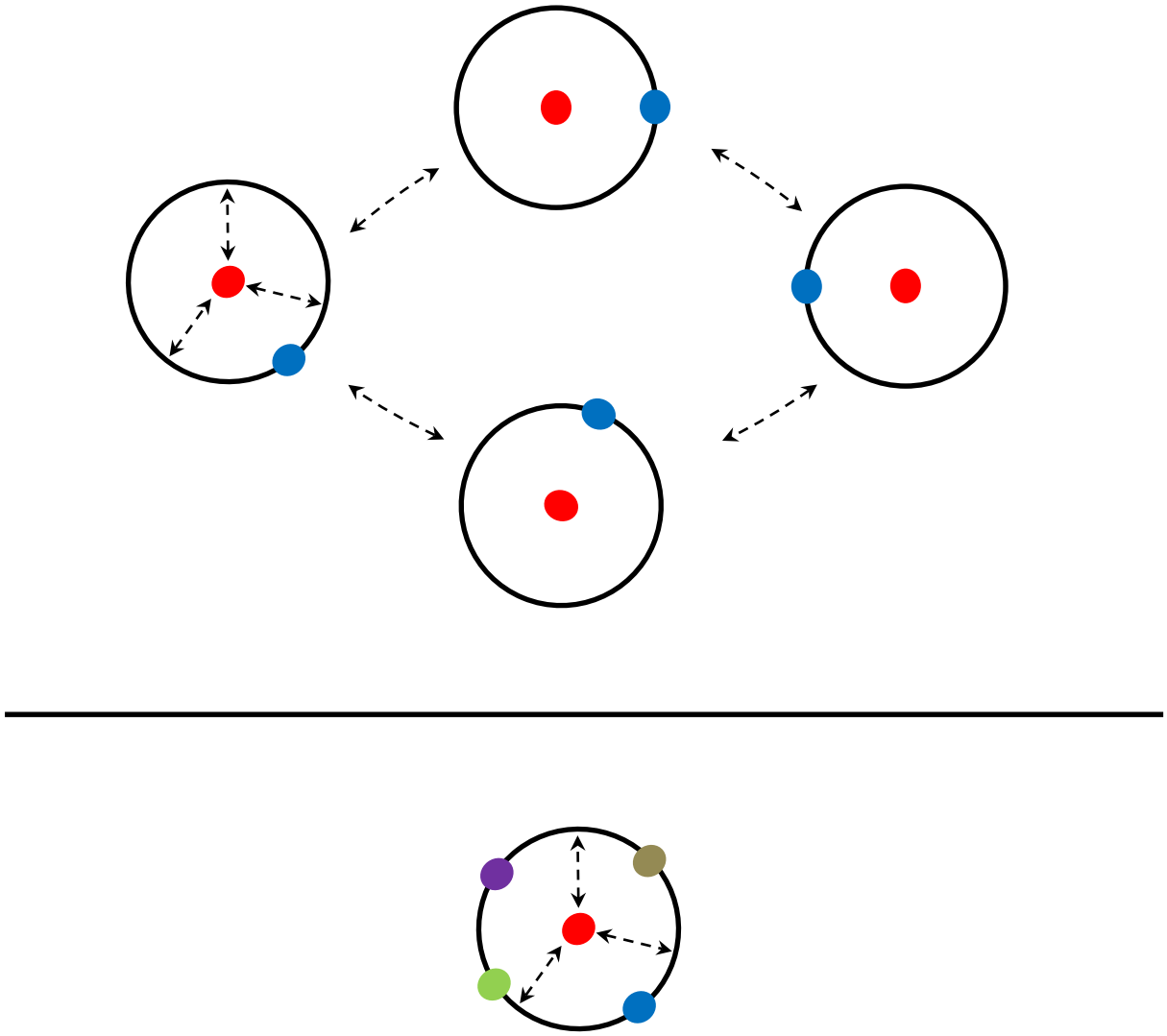


Figure 13.4

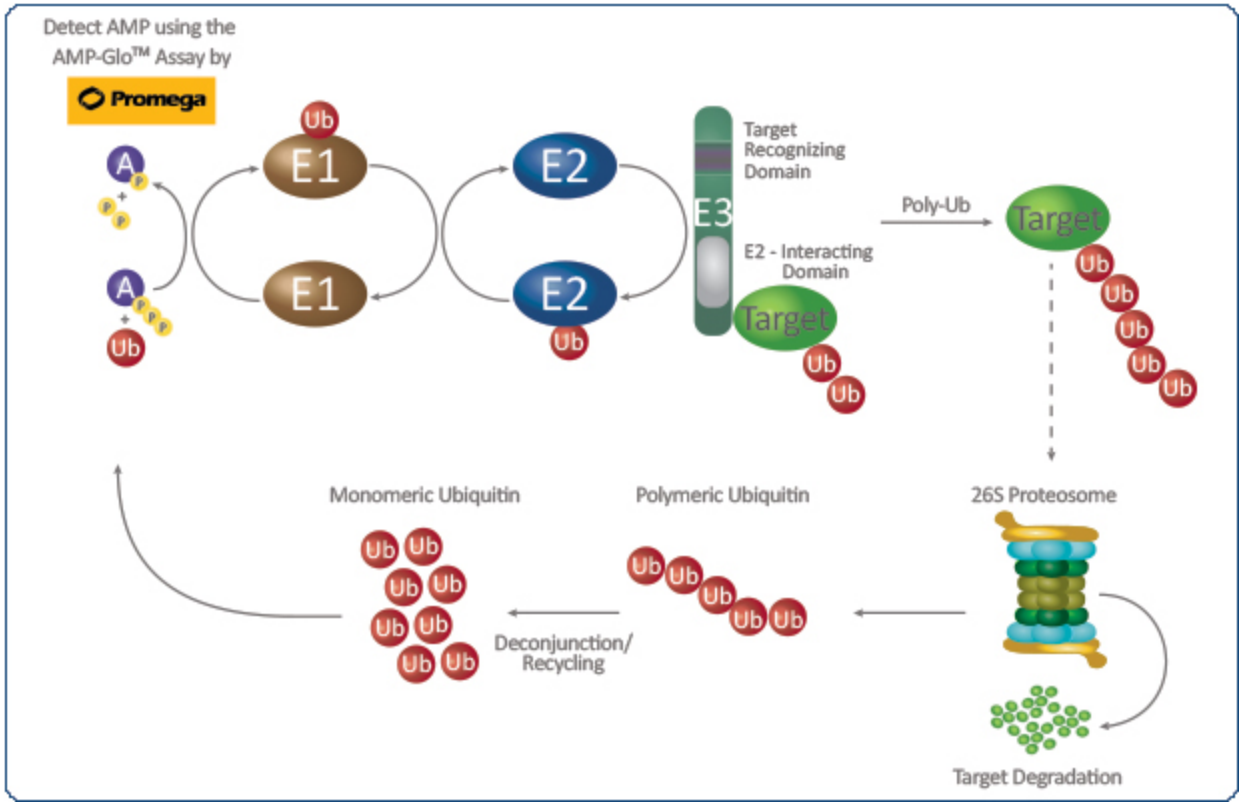

Detect AMP using the
AMP-Glo™ Assay by
Promega
Ub
E1
E1
E2
E2
Ub
E3
Target
Recognizing
Domain
E2 - Interacting
Domain
Target
Ub
Ub
Poly-Ub
Target
Ub
Ub
Ub
Ub
Ub
26S Proteosome
Polymeric Ubiquitin
Monomeric Ubiquitin
Deconjunction/
Recycling
Target Degradation

Figure 13.5

Primary destabilizing and stabilizing residues:

Modification and addition of destabilizing residue:

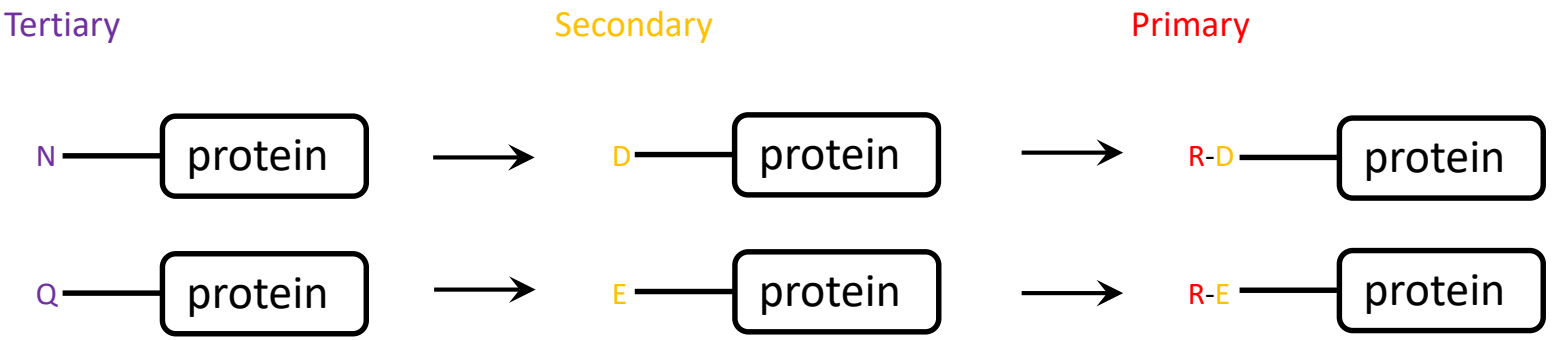
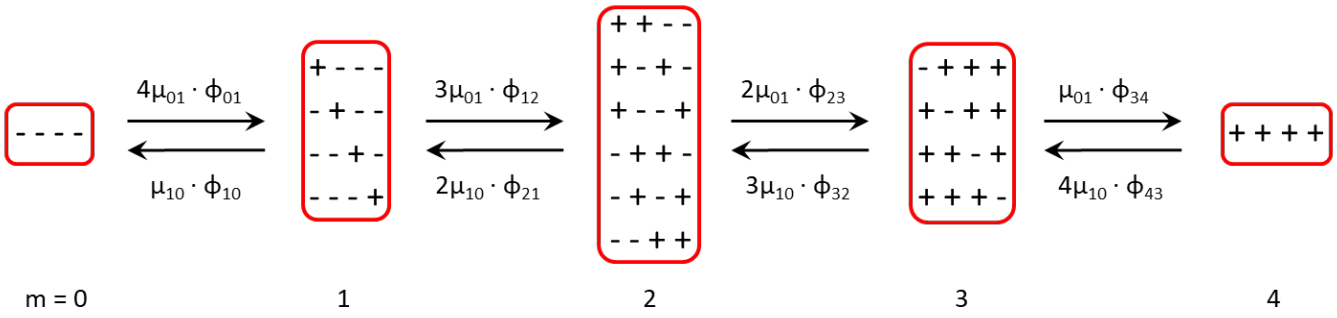
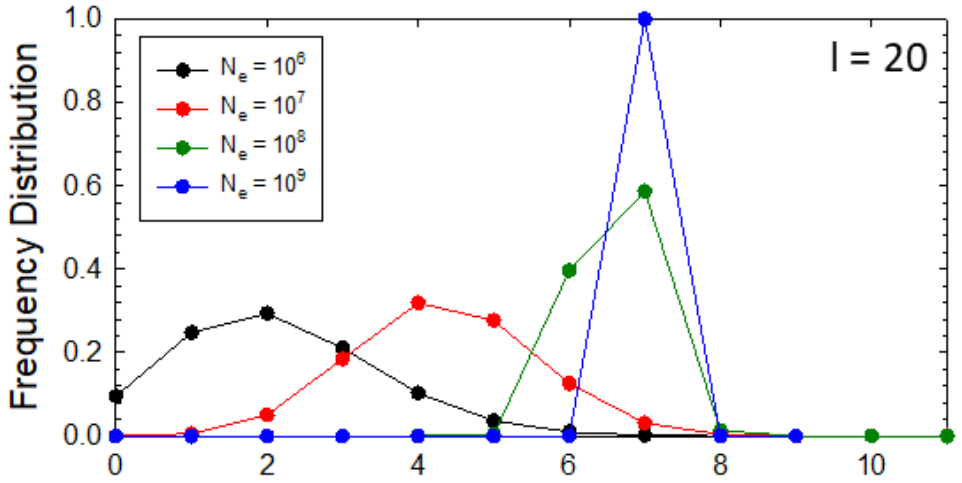


Figure 13.6


- - - -
$4\mu_{01} \cdot \phi_{01}$
$\mu_{10} \cdot \phi_{10}$
+ - - -
- + - -
- - + -
- - - +
$3\mu_{01} \cdot \phi_{12}$
$2\mu_{10} \cdot \phi_{21}$
+ + - -
+ - + -
+ - - +
- + + -
- + - +
- - + +
$2\mu_{01} \cdot \phi_{23}$
$3\mu_{10} \cdot \phi_{32}$
- + + +
+ - + +
+ + - +
+ + + -
$\mu_{01} \cdot \phi_{34}$
$4\mu_{10} \cdot \phi_{43}$
+ + + +
m = 0
1
2
3
4


Frequency Distribution
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
$N_e = 10^6$
$N_e = 10^7$
$N_e = 10^8$
$N_e = 10^9$
l = 20