Published in final edited form as:

Nat Chem Biol. 2014 November ; 10(11): 884–891. doi:10.1038/nchembio.1670.

Energy landscapes of functional proteins are inherently risky

Anne Gershenson¹, Lila M Gierasch^{1,2,*}, Annalisa Pastore³, and Sheena E Radford⁴

¹Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, Massachusetts, USA.

²Department of Chemistry, University of Massachusetts Amherst, Amherst, Massachusetts, USA.

³Department of Clinical Neurosciences, King's College London, Denmark Hill Campus, London, UK.

⁴Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds, UK.

Abstract

Evolutionary pressure for protein function leads to unavoidable sampling of conformational states that are at risk of misfolding and aggregation. The resulting tension between functional requirements and the risk of misfolding and/or aggregation in the evolution of proteins is becoming more and more apparent. One outcome of this tension is sensitivity to mutation, in which only subtle changes in sequence that may be functionally advantageous can tip the delicate balance toward protein aggregation. Similarly, increasing the concentration of aggregation-prone species by reducing the ability to control protein levels or compromising protein folding capacity engenders increased risk of aggregation and disease. In this Perspective, we describe examples that epitomize the tension between protein functional energy landscapes and aggregation risk. Each case illustrates how the energy landscapes for the at-risk proteins are sculpted to enable them to perform their functions and how the risks of aggregation are minimized under cellular conditions using a variety of compensatory mechanisms.

Evolution selects protein sequences for the ability to function and not simply for stability or ability to fold, as long as folding is efficient enough for cell survival. This principle has many consequences. The sequence of a functional protein encodes: the ability of the protein to sample alternate conformations; the probability with which this sampling occurs, i.e., the conformational dynamics of the protein; the ease with which the protein folds and finds its native, functional state; and the tendency of the protein to interact with other proteins or itself. All of this is described by a protein's energy landscape: the multidimensional surface that describes the choreography of protein folding. The energy landscapes for most proteins in the functional proteome will be evolutionarily selected to be rough, which means that a protein will have multiple energy minima available to it, depending on its environment and

© 2014 Nature America, Inc. All rights reserved.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

*Correspondence and requests for materials should be addressed to L.M.G. gierasch@biochem.umass.edu.

Competing financial interests The authors declare no competing financial interests.

interactions. Rough energy landscapes correlate with frustration in folding¹. They also explain the observation that kinetically stable states (deep minima with high barriers around them) may frequently be populated during folding^{2,3} and may form part of the ensemble of states populated under a given set of cellular conditions (pH, temperature or concentrations of partner ligands, for example). For folded proteins, cellular conditions and availability of binding partners may shift the energy landscape such that the bound conformation represents a deeper energy well and partially folded states are relatively disfavored. Only recently has an unavoidable dark side of functional protein energy landscapes—kinetic traps, frustration, metastability and sampling of alternative folds—been fully recognized.

Despite our ability to lock proteins into crystalline forms, it has long been known that they are not rocks. Quite the contrary: proteins are molecular machines composed of movable parts that work together to accomplish a wide range of physiological functions. From an energy landscape point of view, protein dynamics translate into excursions from one low-energy conformation to another. Hence, the dynamics of protein structures are crucial for their function by enabling the population of alternative conformations. The alternative conformations may enable functional interactions by exposing interactive surfaces, providing opportunities for new, favorable interactions (in functional terms). However, there is also a chance that the exposed interaction surfaces are aggregation prone, thus creating a risk of dysfunctional interactions, which are causative in an increasing family of pathologies^{4,5}.

Even simple two-dimensional representations of energy landscapes can help to clarify these concepts. The depth of energy minima describes a protein's thermodynamic stability; the heights of the barriers separating energy minima dictate the kinetic stability of a protein, that is, how readily it can leave one conformation and sample another; and the width of minima correlates with the breadth of the conformational ensemble within the energy well. The shape of a protein's energy landscape is dictated by myriad weak and competing interactions that define the search for its native fold. These same interactions also enable proteins to undergo conformational changes and to use conformational malleability and dynamics for their roles, such as binding-induced folding⁶ and allosteric regulation^{7–9}. As a consequence, the attributes of energy landscapes are sculpted by the requirements for proteins to perform their functions. In addition, the ability of proteins to evolve new functions places constraints on their energy landscapes that compete with the optimization of stability and function¹⁰. As illustrated by the examples discussed in this Perspective, the functionally required features of protein energy landscapes may put proteins at risk of misfolding, aggregation and/or polymerization.

Schematic examples of possible energy landscapes for functional proteins are illustrated in Figure 1. Many proteins must change their conformations away from their native, or lowest, energy state for ligands to gain access to their binding site (or sites). An example of this behavior is the family of intracellular lipid-binding proteins (iLBPs) discussed in the first section below. iLBPs are characterized by an energy landscape that incorporates a near-native apo state (N^* ; Fig. 1a). In the absence of ligand, N^* is dynamically visited from the native state N . Binding of ligand stabilizes and rigidifies the protein (Fig. 1a), leading to the native holoprotein (N). N and N^* are generally separated by a relatively low energy barrier

such that they readily interconvert. If the functionally required N* state exposes hydrophobic surface, it will also be susceptible to aggregation.

The energy landscapes of intrinsically disordered proteins (IDPs) are devoid of a single energy minimum and instead are characterized by many possible nearly isoenergetic conformational states separated by very low energy barriers, as schematically shown in Figure 1b. Interactions with partner proteins select from this heterogeneous collection of conformational ensembles and shift the landscape so that one conformation becomes preferentially populated. For these proteins as well as for natively folded or partially folded proteins, stability is enhanced by formation of a complex with ligands or protein partners. Thus, for IDPs, major changes in the energy landscape can result from binding, which can result in the formation of a clear energy minimum in the energy landscape (Fig. 1b), akin to that of their natively folded apo-protein counterparts.

Minima on protein conformational energy landscapes may be separated by high barriers, leading to the possibility of kinetically stable states. The example in Figure 1c shows a free energy landscape that includes a kinetically trapped metastable state (M) separated by a large energy barrier from the native state (N), such as that found in the serpin family or for proteins with a *cis*-proline, such as β_2 -microglobulin (β_2m). The conformational population will shift to the more stable N state either through a slow, spontaneous process that may be catalyzed by environmental perturbations (serpin) or via proline isomerization (β_2m). Here, as in the case of the N* state of the iLBP, the M state is vulnerable to intermolecular association to form pathological aggregates, amyloid fibrils or polymers.

In all of these situations, non-native or unfolded protein conformations will inevitably be populated under some cellular conditions, for example, during protein biosynthesis or when a protein's partner is absent or degraded. Because the evolutionary pressures for proteins to perform their functions drive the features of their energy landscapes, it is not surprising that we are increasingly recognizing new aggregation-based diseases and associating them with features encoded in protein sequences that correlate with their functions. The development of proteins as drugs is accompanied by similar risks, as aggregation poses one of the major bottlenecks in the exploitation of biologics for intervention in human health. It is also not surprising that a wide array of compensatory cellular strategies has evolved to minimize the accumulation of high-risk protein conformational states. Cellular factors that can offset the risk of aggregation include adjustments in expression levels¹¹ and protein turnover^{12,13}, both of which modulate cellular concentration; protection by molecular chaperones¹⁴; or the presence of intracellular ligands¹⁵. In addition, amino acid sequences themselves have evolved ways to minimize aggregation, including the appropriate positioning of proline or glycine residues; use of gatekeeper residues; minimization of sequence segments with high hydrophobicity, high β -sheet propensity and low net charge; insertion of protective elements at edge β -strands; and, indeed, folding into stable globular structures¹³. In a recent intriguing example of a protective mechanism to enable function and avoid aggregation, a chaperone in the type III bacterial secretion system exists as a dynamic, molten globule homodimer, poised to bind substrate at the dimer interface¹⁶. The aggregation-prone substrate-binding surface is protected by dimerization, and the dimer dynamics enable substrate binding. A similar mechanism has been reported by the HdeA family of bacterial chaperones^{17,18}.

Another recent study postulates that the mitochondrial protein frataxin experiences frustration owing to competition between folding and function; in this case, aggregation is avoided because aggregation-prone sequences are sequestered by misfolding events along the folding pathway¹⁹. Even though mechanisms have arisen to minimize the risks, vulnerability to aggregation is an inevitable cost of selective pressures for protein function, as seen for the examples discussed in the sections below: iLBPs, serpins, the immunoglobulin domain β_2m and ataxin-3 (Atx3). These four very different proteins serve as exemplars of the delicate balance between the requirements for protein function (which frequently involve dynamics), rough energy landscapes and the threat of aggregation. Together, the examples portray how nature has used an array of strategies to enable function to evolve while avoiding the potential risks of aggregation.

iLBPs dynamically open to bind hydrophobic substrates

iLBPs comprise a family of β -barrel proteins that bind hydrophobic ligands in an interior cavity²⁰. Their roles are varied, but all of them rely on their ability to solubilize and transport otherwise water-insoluble ligands. No aggregation or misfolding diseases have been reported for iLBPs, although *in vitro* studies have shown that the archetypal iLBP, cellular retinoic acid-binding protein 1 (CRABP1), is prone to aggregation as a purified protein or when expressed in *Escherichia coli*²¹. What is the origin of this aggregation propensity? How is aggregation avoided under physiological conditions? Formal possibilities include regulation of CRABP1 expression and clearance, so as to reduce its cellular concentration, or protection from aggregation by binding to ligand. Both may be protective in cells. Additionally, recent work has suggested that the folding mechanism of CRABP1 protects aggregation-prone sequences early in the formation of native structure, providing a neat mechanism of minimizing the risk of aggregation²².

From early on, it was observed that iLBPs are dynamic in the absence of their ligand²³. This dynamic character is intimately linked to the functional requirements and gives rise to rugged energy landscapes of iLBPs. The helical region of these proteins, which comprises a helix-loop-helix motif, has been dubbed ‘the helical portal’, as its movements allow the ligand access to the interior cavity within the ten-stranded β -barrel, where the ligand-binding surface is largely presented (Fig. 2). NMR studies of many iLBPs, including ileal lipid-binding protein²⁴, intestinal fatty acid-binding protein (FABP)²⁵, bile acid-binding protein²⁶, liver FABP²⁷ and CRABP1 (ref. 28), all show substantially higher fluctuations in the helical portal than in the β -barrel. The open form of iLBPs that allows ligand binding exemplifies an N* state, (Fig. 1a). In all cases, binding of ligand rigidifies the iLBP and, in particular, dampens movement of the helical portal, shifting the population to the N state.

Several lines of evidence point to the importance of the dynamics of apo-iLBPs for ligand entry and egress from the binding site²⁹. Some members of the subclass of iLBPs known as FABPs are postulated to bind their ligand directly from a lipid bilayer and then release the ligand into the bilayer. A portal-open form of the FABP has been postulated to dock onto the lipid bilayer, enabling the membrane-resident ligand to gain access to the binding cavity³⁰. In other cases, such as the retinoic acid-binding iLBPs, direct interactions are postulated to occur with a protein partner, the nuclear retinoid receptor, and the ligand is proposed to be

passed onto the partner via a mechanism that also requires opening of the helical portal³¹. The dynamics of the helical portal of the iLBP structure are therefore crucial to many of its functions. Major questions ask how much, or how often, the apo-protein is present inside cells and how cells protect themselves from the threat of apo-state aggregation.

Extensive studies of the folding and aggregation of CRABP1 have been performed^{21,22,32–36}. This has led to intriguing linkages: CRABP1 folds via multiple kinetic steps in which early hydrophobic collapse occurs in a few milliseconds along with formation of its helix-loop-helix motif. In a subsequent 100-ms step, the barrel topology organizes. Stable hydrogen bonding and van der Waals packing of side chains throughout the barrel occur only much later (time constant ~1 s), in a cooperative manner in the rate-determining step. Analysis of which topological interactions are organized earliest in CRABP1 folding showed that interactions between strands 10 and 1, and between helix 1, turn IV and strand 9 (Fig. 2), all favoring barrel closure, are organized first. These structural features sequester β -strand regions that are predicted to have the highest aggregation propensity. Indeed, assessment of the involvement of two different regions of the CRABP1 sequence in aggregates formed in *E. coli* and *in vitro* showed that they comprise β -strand 3–turn II– β -strand 4 and β -strands 9 and 10 (Fig. 2). Formation of structural contacts concomitant with early barrel closure sequester these very regions from intermolecular interactions²². It is thus striking that the functionally required exposure of hydrophobic surface in this archetypal iLBP gives rise to a high risk of aggregation. The residues evolutionarily conserved to mediate ligand binding are precisely those that create sequences of high predicted aggregation propensity, and a folding mechanism has evolved that enables protection of these regions during folding, not unlike the example of frataxin described above¹⁹. The hydrophobic collapse step in CRABP1 folding can be viewed as a kind of specific intramolecular aggregation. By the evolution of a folding mechanism that sequesters aggregation-prone sequences early through this collapse, intermolecular aggregation is effectively outcompeted. This iLBP exemplar epitomizes the ying of functionally required dynamics that leads to aggregation susceptibility and the yang of simultaneous offsetting coping mechanisms—in this case, the folding mechanism.

Serpins can get caught in their own ‘mouse trap’

Members of the ‘serine protease inhibitor’ or serpin superfamily are also paradigms of the delicate balance between function and folding-associated vulnerabilities. The inhibitory serpins inactivate their target proteases by dangling a scissile bond on a solvent-exposed reactive center loop (RCL). The RCL links the large central β -sheet A to the smaller β -sheet C in these α - and β -sheet proteins³⁷ (Fig. 3a). Protease cleavage triggers RCL insertion into β -sheet A, whereupon it becomes the fourth strand in the now six-stranded sheet (Fig. 3a). This springing of the mouse trap translocates the covalently attached protease ~70 Å relative to the serpin, mechanically disrupting the protease active site by pulling on the single acyl-enzyme covalent bond between the protease active site serine (or cysteine) and the RCL^{38,39}. Thus, serpins fold into a functionally active, yet metastable, kinetically trapped conformation (like the M state in Fig. 1c) that is poised to undergo a massive conformational change concomitant with the serpin’s inhibitory action on its target protease^{37,40}. In the absence of cleavage by their target protease, serpins may slowly convert to an N-like state

(Fig. 1c) via insertion of the RCL into β -sheet A; the resulting lower-energy, inactive conformation is termed 'latent'⁴¹. The latency transition can be accelerated or decelerated by environmental perturbations including pH, temperature and the binding of protein partners⁴⁰, all of which can alter the energy barrier between the active and latent states (Fig. 1c). When serpins perform the functional role of inhibiting proteases, proteolytic cleavage of the RCL initiates a similar conformational change, resulting in an even deeper energy minimum on the energy landscape and altered chain connectivity. These conformational changes support the notion that the serpin energy landscape has multiple energy minima, including the 'cocked' inhibitory state, with six strands in β -sheet A, and the more stable, strand-inserted state, with six strands in β -sheet A (Fig. 1c).

A number of serpins, such as the hormone-binding globulins, do not inhibit proteases, but they still use RCL conformational changes to mediate function^{42–44}. In addition, conformational changes in other regions, particularly the N-terminal helical region, often help regulate serpin function^{37,45–48}. Conformational malleability is thus key to function throughout the serpin family.

This functionally essential conformational malleability comes at a cost: serpins form a variety of multimers and polymers (Fig. 3b), and formation of these species leads to loss of function and in some cases toxicity. This problem becomes more pronounced as RCL insertion into β -sheet A becomes more energetically favorable. Formation of serpin multimers is favored *in vitro* when the stability of the folded monomers is perturbed. The RCL with its β -strand character can add to the edge of β -sheets^{49,50}. Serpin multimers can also be formed by domain swaps, either as a dimer (seen for ATIII)⁵¹ or a trimer (observed in α 1-antitrypsin (A1AT)⁵²) (Fig. 3b). In the A1AT trimer, the C-terminal 34 residues are swapped, and *in vitro* folding experiments⁵³ as well as a number of disease- and polymerization-associated mutations at or near the serpin C terminus⁵⁴ underscore the importance of the C terminus for proper folding and function. Hence, the folding energy landscape of serpins must have multiple wells, and destabilization of the metastable functional state can lead to competition with any of a number of inter-molecularly associated states.

The multimerization and polymerization reactions of serpins are a manifestation of the tug of war between the need to fold to a metastable functional state and the vulnerability of serpins to disease-associated polymerization events. Serpin polymers in cells are generally formed in the endoplasmic reticulum (ER), where secretory serpins fold and mature. In most cases, cellular quality control networks keep misfolding and polymerization in check. Nonetheless, the synthesis and accumulation of polymerization-prone serpin mutants in the ER can lead to cell death and diseases termed the serpinopathies⁵⁵. The most prevalent serpinopathy arises from mutations in A1AT and leads to liver disease. Similarly, polymerization-prone neuroserpin mutants lead to epilepsy and dementia⁵⁶. Disease severity correlates with the polymerization propensity of both A1AT and neuroserpin, as determined *in vitro*^{57,58}. In the ER, the protein quality control network helps mutant serpins to fold and targets mutants for degradation through ER-associated degradation and/or autophagy^{59,60}. Misfolded serpins may also be recognized by protein quality control in the Golgi⁶¹. Thus far, the most advanced therapy for A1AT serpinopathies is a small molecule that increases

autophagy⁵⁹. However, it remains unclear which specific members of the quality control network are most important to ameliorate serpinopathologies and how the network can best be tuned to minimize polymerization.

Mutations associated with serpin misfolding and polymerization in the ER often map to structural regions that are important for functionally required conformational changes^{37,40,54}. Most of the neuroserpin mutations are in the shutter region (Fig. 3a), which must open to allow full insertion of the RCL into β -sheet A, and the A1AT Z mutation disrupts a conserved salt bridge that helps control the RCL conformation. This correlation suggests that function and folding rely on an overlapping set of key residues. The ease with which polymers may be formed *in vitro* even from wild-type serpins suggests that metastable serpins are precariously poised on their energy landscape and, given high enough concentrations and enough time in an on-pathway intermediate or misfolded state, serpins will polymerize. Although this may seem an error of evolution, the strong overlap between serpin folding and function suggests that some of the attributes that allow serpins to perform a host of functions and to change functions in a regulated manner are correlated with a susceptibility for polymerization and its consequent deleterious outcomes.

The dangerous plasticity of an immunoglobulin fold

The tug of war between folding and function is also demonstrated in the β -sandwich immunoglobulin (Ig) domains. Widely associated with antibodies, the all- β -sheet family of Ig domains are widespread in nature, with roles as scaffold proteins and binding partners and as tandem repeats within many multidomain proteins. The folding mechanisms of several Ig domains have been mapped by mutation and analysis of the stability and the folding and unfolding kinetics of the resulting sequences⁶². These studies have shown that a central tetrad of β -strands, conserved topologically in all Ig domains, needs to form first, acting as a key stepping stone in the search for the native state. Folding is challenged, however, by several complicating factors: most Ig domains are stabilized by a disulfide bond that links β -strands B and F, several have a *cis* X-Pro peptide bond in one or more of the loops, and the isomerization of this bond slows folding^{63,64}. Additionally, the occurrence of Ig domains in long, tandem arrays increases the probability of protein misfolding by events such as domain swapping⁶⁵. When isolated from their heavy-chain binding partners, aberrant folding of light chains can result in aggregation and disease (reviewed in ref. 66).

A striking example of the tension between function and folding is found in the Ig domain of β_2m . This 99-residue protein has a canonical Ig fold comprising seven β -strands. The functional role of β_2m is to chaperone the folding of its binding partner, the heavy chain of the major histocompatibility complex 1 (MHC-1), which is required for functional antigen presentation at the cell surface⁶⁷. Like many Ig domains, β_2m contains a disulfide bond linking residues Cys25 to Cys80 in β -strands B and F and a *cis* peptide bond linking residues His31 and Pro32. Studies of the folding of β_2m have shown that the protein becomes trapped in a long-lived intermediate, known as I_T , which has to wait for the slow (minute timescale) *trans-cis* isomerization of the His31-Pro32 bond to form its native structure^{63,64,68,69}. The energy landscape of β_2m , thus, is rough, with a near-native, long-lived and trapped intermediate blocking rapid folding to the native state (Fig. 4a). Structural studies of I_T

obtained by creation of a trapped metastable species that mimics I_T by deletion of the N-terminal six amino acids (N6)⁷⁰ have shown that this species is able to fold to an Ig-like structure that differs only marginally from that of the native protein (r.m.s. deviation of backbone atoms of 1.5 Å)⁷⁰. Most importantly, however, and by contrast with native β_2m , I_T and its structural analog, N6, are highly aggregation prone^{64,71,72}, assembling rapidly into amyloid fibrils that resemble those that form in the disease dialysis-related amyloidosis^{73,74}. Moreover, this kinetically trapped non-native conformation of the protein can promote misfolding of the initially innocuous native state into an amyloidogenic conformer, analogous to conformational conversion associated with prion disease⁷⁵. Thus, trapped partially folded proteins formed on rugged folding landscapes not only give rise to difficulties by creating folding bottlenecks and enhancing aggregate potential, but they can also wield a second blow by turning soluble proteins into aggregation precursors.

Why, then, is Pro32 conserved across so many Ig domains in its *cis* isomer? And, why is wild-type β_2m resilient to aggregation, when its close structural homolog N6 aggregates spontaneously and rapidly *in vitro*? The answers, again, include the evolutionary pressure for function: *cis*-Pro32 is solvent exposed in wild-type β_2m , forming a hydrophobic surface involved in binding to the MHC-1 heavy chain (Fig. 4b), thereby enabling it to chaperone MHC-1 complex assembly⁷⁶. Formation of the MHC-1 complex involves interaction with Trp60, which is also highly conserved⁷⁷ (Fig. 4). Despite being solvent exposed, mutation of Trp60 to Gly stabilizes native β_2m and also reduces its propensity to aggregate, providing another example of function winning over folding and stability^{77,78}. In a similar vein, the β_2m variant with mutation D76N, which gives rise to a recently discovered form of hereditary systemic amyloid disease (in the absence of kidney dysfunction)⁷⁹, is also more aggregation prone than its wild-type counterpart as this mutation reduces the stability of the native state relative to I^{80}_T . In this case, however, the MHC-1 heavy chain stabilizes and thereby reduces the aggregation potential of the D76N mutant, protecting against aggregation and disease⁸¹. Other perturbations, for example, binding of Cu^{2+} ions during the dialysis procedure, combined with other truncations and chemical modifications, can enhance aggregation of β_2m by increasing the population of the non-native *trans*-Pro32 forms (reviewed in ref. 82). Given that an intact adaptive immune response is vital for life in all higher eukaryotes, there is immense evolutionary pressure on the sequences of both the MHC-1 heavy chain and β_2m to ensure that they fold and bind efficiently. This explains why the sequence of β_2m is so highly conserved from cartilaginous fish to humans⁸³ and why the smallest changes to its sequence that cause the isomerization of a single peptide bond can lead to aggregation and disease. In addition, the interactions between β_2m and the MHC-1 heavy chain have coevolved in such a way as to minimize aggregation risks, although the susceptibility is ever present when the fragile fold of β_2m is left on its own.

Aggregation-prone protein recognition sites in Atx3

Atx3 provides another example of the competition between function and aggregation and also illustrates a compensatory mechanism that again involves the protective role of functional protein-protein interactions. Atx3 belongs to the protein family associated with neurodegenerative diseases that are caused by the anomalous expansion of polymorphic polyglutamine (polyQ) tracts⁸⁴. When above a threshold of approximately 37 repeats, these

regions promote protein aggregation, misfolding and consequent cell death⁸⁵. The polyQ-associated diseases include the well-known Huntington's chorea as well as several spinocerebellar ataxias, among them spinocerebellar ataxia type 3, which is linked to Atx3. The proteins responsible share no similarities in terms of sequence, cellular localization or function. PolyQ expansion is a necessary requisite for disease development, but it is now established that other regions of the proteins are important contributors to aggregation⁸⁴.

Atx3 is a soluble protein (~350 kDa) that can shuttle in and out of the nucleus (detailed reviews are in refs. 86–88). Although it is mainly cytoplasmic in unaffected brain and in normal neurons, Atx3 partitions to neuronal cell nuclei in diseased brains. Functionally, Atx3 is a deubiquitinating enzyme that preferentially cleaves ubiquitin chains longer than four repeats⁸⁹. It is composed of an evolutionarily conserved N-terminal region, the Josephin domain, and an intrinsically unfolded C-terminal region that contains the polymorphic polyQ tract, a putative coiled-coil^{90,91} and two or three ubiquitin-interacting motifs, depending on the isoform. The Josephin domain accounts for enzymatic activity and forms a globular structure with a papain-like fold. The Atx3 Josephin domain contains a feature that is unusual among deubiquitinating enzymes and more generally among cysteine proteases: a dynamic helical hairpin that protrudes out into solution from the main body of the globular domain.

The full range of possible protein-protein interactions formed by Atx3 remains to be determined, but some of the partners have been identified^{92–94}. Among the better-characterized partners are polyubiquitin chains and other components of the ubiquitin-proteasome pathway, including the transitional ER ATPase valocin-containing protein Vcp and the human homologs of Rad23, the HHR23 proteins. Atx3 can also bind monoubiquitin, and it does so through the multiple binding sites distributed along its length, albeit with low (20–400 μ M) affinity. Two ubiquitin-binding sites have been described on Josephin, and they are located on either side of the helical hairpin⁹⁵ (Fig. 5). Site 1 is essential for enzymatic activity⁹⁶. Site 2, which is more exposed, confers K48 ubiquitin-chain linkage preference to Josephin and overlaps with the surface of interaction for the ubiquitin-like domain of the HHR23B protein⁸⁹. Previous measurements suggest that the dynamics of the helical hairpin, which are distinct from that of the rest of the domain, may have a role in ubiquitin recognition and possibly determine substrate and ubiquitin linkage specificity⁹⁵. The ubiquitin-interacting motifs are thought to be important for polyubiquitin binding and to enhance affinity.

A tight relationship between function and aggregation is clear for Atx3. As in other polyQ-containing proteins, disease is determined by polyQ expansion, but the Josephin domain contributes to the process and is able per se to induce aggregation and misfold^{97–100}. As shown in these references, the aggregates and fibrillar species of Josephin have morphologies and features very similar to those observed for full-length nonexpanded Atx3. Constructs lacking the polyQ-containing C-terminal region were shown to induce neurodegeneration in mice models¹⁰¹. Careful analysis of the regions of Josephin that promote aggregation showed that the two main aggregation hot spots coincide with the two ubiquitin-binding sites 1 and 2 and are formed by two exposed hydrophobic surfaces that have evolved to bind ubiquitin (Fig. 5). Accordingly, aggregation of Atx3 is strongly

mitigated by mutations in Josephin that, at the same time, abolish ubiquitin binding¹⁰². Likewise, the presence of excess monoubiquitin has a similar effect and leads to amorphous aggregates¹⁰².

Consequently, ubiquitin recognition has been suggested to be an important factor that has a role in protecting Atx3 from protein aggregation *in vivo*¹⁰². A similar behavior in which the same regions of a protein are involved both in functional interactions and in aberrant aggregation is observed for ataxin-1, another member of the polyQ disease protein family¹⁰³. This evidence teaches us an important lesson that, while confirming once again an intimate interconnection between a functional energy landscape and aggregation risk, also points to another cellular compensatory mechanism that might be harnessed pharmaceutically: the protective role of protein-protein interactions. Hence, a deeper understanding of protein function may inspire the generation of compounds that mimic cellular interactions and result in protection from aggregation and disease.

Concluding remarks

Life is full of compromises. It is important to remind ourselves that species, organisms and their component cells and molecular machines have been subjected to fitness pressures that require them to function well enough to survive and compete in their evolutionary niche. In the case of the evolutionary selection for molecular function, we are becoming increasingly aware of the compromises that are manifested in the vulnerability of proteins to altered conditions such as pH and temperature, stresses that challenge the cellular context such as proteostatic depletion and mutations that subtly alter the energetic landscape of the protein. The discovery of the role of misfolding in aggregation diseases highlights the fragility of proteins and the risk of the black pit of protein aggregation, which can readily outcompete folding once the scales are tipped toward that intermolecular process.

The four vignettes we have provided in this brief Perspective are intended to illustrate the tenuous balance between functional requirements on proteins and their susceptibility to competing misfolding and aggregation. These are not the only examples of functional constraints exacerbating the risk of protein aggregation—far from it. Nonetheless, these four examples show compellingly how the sculpting of an energy landscape for function can lead to dangerous characteristics, whether they be necessity for dynamics (the iLBPs), kinetically stable partially folded states (Ig folds), conformations that have exposed interactive (often hydrophobic) surfaces (Atx3) or metastability to enable a response required for function (serpins). We hope that the ideas described here will lead readers to look more deeply at the functional energy landscapes of proteins and in so doing gain insights that might help understanding of the cellular roles of proteins, the coping mechanisms employed in biology and possible intervention strategies one might invoke therapeutically.

Acknowledgments

We thank our many collaborators and co-authors whose experiments and ideas have contributed to our work. We also recognize that all relevant examples and references cannot be included in a short perspective such as this, and we apologize to those whose contributions we have omitted. We acknowledge, with thanks, funding from the National Institutes of Health (grants GM027616 and GM101644 to L.M.G., GM094848 to L.M.G. and A.G. and GM060418 to A.G.); the Medical Research Council (grant U117584256 to A.P.) and the European Research

Council under the European Union's Seventh Framework Programme (FP7/2007-2013; 322408) and the Wellcome Trust (WT092896MA to S.E.R.).

References

1. Jenik M, et al. Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.* 2012; 40:W348–W351. [PubMed: 22645321]
2. Sutto L, Latzer J, Hegler JA, Ferreira DU, Wolynes PG. Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc. Natl. Acad. Sci. USA.* 2007; 104:19825–19830. [PubMed: 18077415]
3. Zheng W, Schafer NP, Wolynes PG. Frustration in the energy landscapes of multidomain protein misfolding. *Proc. Natl. Acad. Sci. USA.* 2013; 110:1680–1685. [PubMed: 23319605]
4. Knowles TP, Vendruscolo M, Dobson CM. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 2014; 15:384–396. [PubMed: 24854788]
5. Uversky VN. Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front. Biosci. (Landmark Ed.)*. 2009; 14:5188–5238. [PubMed: 19482612]
6. Wright PE, Dyson HJ. Linking folding and binding. *Curr. Opin. Struct. Biol.* 2009; 19:31–38. [PubMed: 19157855]
7. Ma B, Tsai CJ, Haliloglu T, Nussinov R. Dynamic allostery: linkers are not merely flexible. *Structure.* 2011; 19:907–917. [PubMed: 21742258]
8. Kalodimos CG. Protein function and allostery: a dynamic relationship. *Ann. NY Acad. Sci.* 2012; 1260:81–86. [PubMed: 22256894]
9. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. *Nature.* 2014; 508:331–339. [PubMed: 24740064]
10. Tóth-Petróczy A, Tawfik DS. The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* 2014; 26:131–138. [PubMed: 25038399]
11. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* 2007; 32:204–206. [PubMed: 17419062]
12. De Baets G, et al. An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. *PLOS Comput. Biol.* 2011; 7:e1002090. [PubMed: 21731483]
13. Monsellier E, Chiti F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* 2007; 8:737–742. [PubMed: 17668004]
14. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* 2013; 82:323–355. [PubMed: 23746257]
15. Pastore A, Temussi PA. The two faces of Janus: functional interactions and protein aggregation. *Curr. Opin. Struct. Biol.* 2012; 22:30–37. [PubMed: 22155180]
16. Chen L, et al. Structural instability tuning as a regulatory mechanism in protein-protein interactions. *Mol. Cell.* 2011; 44:734–744. [PubMed: 22152477]
17. Hong W, Wu YE, Fu X, Chang Z. Periplasmic protein HdeA exhibits chaperone-like activity exclusively within stomach pH range by transforming into disordered conformation. *J. Biol. Chem.* 2005; 280:27029–27034. [PubMed: 15911614]
18. Foit L, George JS, Zhang BW, Brooks CL III, Bardwell JC. Chaperone activation by unfolding. *Proc. Natl. Acad. Sci. USA.* 2013; 110:E1254–E1262. [PubMed: 23487787]
19. Gianni S, Camilloni C, Giri R, Toto A, Bonetti D, Morrone A, Sormanni P, Brunori M, Vendruscolo M. Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl. Acad. Sci. USA.* Sep 17, 2014 doi: 10.1073/pnas.1405233111.
20. Bernlohr DA, Simpson MA, Hertz AV, Banaszak LJ. Intracellular lipid-binding proteins and their genes. *Annu. Rev. Nutr.* 1997; 17:277–303. [PubMed: 9240929]
21. Ferrolino MC, Zhuravleva A, Budyak IL, Krishnan B, Gierasch LM. Delicate balance between functionally required flexibility and aggregation risk in a β -rich protein. *Biochemistry.* 2013; 52:8843–8854. [PubMed: 24236614]

22. Budyak IL, et al. Early folding events protect aggregation-prone regions of a β -rich protein. *Structure*. 2013; 21:476–485. [PubMed: 23454187]
23. Banaszak L, et al. Lipid-binding proteins: a family of fatty acid and retinoid transport proteins. *Adv. Protein Chem.* 1994; 45:89–151. [PubMed: 8154375]
24. Lücke C, Zhang F, Ruterjans H, Hamilton JA, Sacchettini JC. Flexibility is a likely determinant of binding specificity in the case of ileal lipid binding protein. *Structure*. 1996; 4:785–800. [PubMed: 8805562]
25. Hodsdon ME, Cistola DP. Ligand binding alters the backbone mobility of intestinal fatty acid-binding protein as monitored by ^{15}N NMR relaxation and ^1H exchange. *Biochemistry*. 1997; 36:2278–2290. [PubMed: 9047330]
26. Eberini I, et al. Conformational and dynamics changes induced by bile acids binding to chicken liver bile acid binding protein. *Proteins*. 2008; 71:1889–1898. [PubMed: 18175325]
27. Cai J, et al. Solution structure and backbone dynamics of human liver fatty acid binding protein: fatty acid binding revisited. *Biophys. J.* 2012; 102:2585–2594. [PubMed: 22713574]
28. Krishnan VV, Sukumar M, Gierasch LM, Cosman M. Dynamics of cellular retinoic acid binding protein I on multiple time scales with implications for ligand binding. *Biochemistry*. 2000; 39:9119–9129. [PubMed: 10924105]
29. Storch J, McDermott L. Structural and functional analysis of fatty acid-binding proteins. *J. Lipid Res.* 2009; 50(Suppl):S126–S131. [PubMed: 19017610]
30. Falomir-Lockhart LJ, Laborde L, Kahn PC, Storch J, Corsico B. Protein-membrane interaction and fatty acid transfer from intestinal fatty acid-binding protein to membranes. Support for a multistep process. *J. Biol. Chem.* 2006; 281:13979–13989. [PubMed: 16551626]
31. Budhu A, Gillilan R, Noy N. Localization of the RAR interaction domain of cellular retinoic acid binding protein-II. *J. Mol. Biol.* 2001; 305:939–949. [PubMed: 11162104]
32. Clark PL, Liu ZP, Rizo J, Gierasch LM. Cavity formation before stable hydrogen bonding in the folding of a β -clam protein. *Nat. Struct. Biol.* 1997; 4:883–886. [PubMed: 9360599]
33. Clark PL, Liu ZP, Zhang J, Gierasch LM. Intrinsic tryptophans of CRABPI as probes of structure and folding. *Protein Sci.* 1996; 5:1108–1117. [PubMed: 8762142]
34. Clark PL, Weston BF, Gierasch LM. Probing the folding pathway of a β -clam protein with single-tryptophan constructs. *Fold. Des.* 1998; 3:401–412. [PubMed: 9806942]
35. Eyles SJ, Gierasch LM. Multiple roles of prolyl residues in structure and folding. *J. Mol. Biol.* 2000; 301:737–747. [PubMed: 10966780]
36. Gunasekaran K, Hagler AT, Gierasch LM. Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins*. 2004; 54:179–194. [PubMed: 14696180]
37. Gettins PGW. Serpin structure, mechanism, and function. *Chem. Rev.* 2002; 102:4751–4804. [PubMed: 12475206]
38. Dementiev A, Dobo J, Gettins PGW. Active site distortion is sufficient for proteinase inhibition by serpins: structure of the covalent complex of α_1 -proteinase inhibitor with porcine pancreatic elastase. *J. Biol. Chem.* 2006; 281:3452–3457. [PubMed: 16321984]
39. Huntington JA, Read RJ, Carrell RW. Structure of a serpin-protease complex shows inhibition by deformation. *Nature*. 2000; 407:923–926. [PubMed: 11057674]
40. Gooptu B, Lomas DA. Conformational pathology of the serpins: themes, variations, and therapeutic strategies. *Annu. Rev. Biochem.* 2009; 78:147–176. [PubMed: 19245336]
41. Mottonen J, et al. Structural basis of latency in plasminogen activator inhibitor-1. *Nature*. 1992; 355:270–273. [PubMed: 1731226]
42. Klieber MA, Underhill C, Hammond GL, Muller YA. Corticosteroid-binding globulin, a structural basis for steroid transport and proteinase-triggered release. *J. Biol. Chem.* 2007; 282:29594–29603. [PubMed: 17644521]
43. Qi X, Chan WL, Read RJ, Zhou A, Carrell RW. Temperature-responsive release of thyroxine and its environmental adaptation in Australians. *Proc. Biol. Sci.* 2014; 281:20132747. [PubMed: 24478298]

44. Zhou A, Wei Z, Read RJ, Carrell RW. Structural mechanism for the carriage and release of thyroxine in the blood. *Proc. Natl. Acad. Sci. USA*. 2006; 103:13321–13326. [PubMed: 16938877]
45. Baek J-H, Yang WS, Lee C, Yu M-H. Functional unfolding of α 1-antitrypsin probed by hydrogen-deuterium exchange coupled with mass spectrometry. *Mol. Cell. Proteomics*. 2009; 8:1072–1081. [PubMed: 19136720]
46. Carrell RW, Stein PE, Fermi G, Wardell MR. Biological implications of a 3 Å structure of dimeric antithrombin. *Structure*. 1994; 2:257–270. [PubMed: 8087553]
47. Trelle MB, Madsen JB, Andreasen PA, Jørgensen TJD. Local transient unfolding of native state PAI-1 associated with serpin metastability. *Angew. Chem. Int. Ed. Engl.* 2014; 53:9751–9754. [PubMed: 25049220]
48. Zhou A, et al. A redox switch in angiotensinogen modulates angiotensin release. *Nature*. 2010; 468:108–111. [PubMed: 20927107]
49. Marszal E, Shrake A. Serpin crystal structure and serpin polymer structure. *Arch. Biochem. Biophys.* 2006; 453:123–129. [PubMed: 16631102]
50. Zhang Q, Law RHP, Bottomley SP, Whisstock JC, Buckle AM. A structural basis for loop C-sheet polymerization in serpins. *J. Mol. Biol.* 2008; 376:1348–1359. [PubMed: 18234218]
51. Yamasaki M, Li W, Johnson DJD, Huntington JA. Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature*. 2008; 455:1255–1258. [PubMed: 18923394]
52. Yamasaki M, Sendall TJ, Pearce MC, Whisstock JC, Huntington JA. Molecular basis of α 1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep*. 2011; 12:1011–1017. [PubMed: 21909074]
53. Dolmer K, Gettins PG. How the serpin α 1-proteinase inhibitor folds. *J. Biol. Chem.* 2012; 287:12425–12432. [PubMed: 22334651]
54. Stein PE, Carrell RW. What do dysfunctional serpins tell us about molecular mobility and disease? *Nat. Struct. Biol.* 1995; 2:96–113. [PubMed: 7749926]
55. Lomas DA, Carrell RW. Serpinopathies and the conformational dementias. *Nat. Rev. Genet.* 2002; 3:759–768. [PubMed: 12360234]
56. Davis RL, et al. Familial dementia caused by polymerization of mutant neuroserpin. *Nature*. 1999; 401:376–379. [PubMed: 10517635]
57. Hagen MC, et al. Encephalopathy with neuroserpin inclusion bodies presenting as progressive myoclonus epilepsy and associated with a novel mutation in the Proteinase Inhibitor 12 gene. *Brain Pathol.* 2011; 21:575–582. [PubMed: 21435071]
58. Miranda E, et al. The intracellular accumulation of polymeric neuroserpin explains the severity of the dementia FENIB. *Hum. Mol. Genet.* 2008; 17:1527–1539. [PubMed: 18267959]
59. Perlmutter DH. α 1-antitrypsin deficiency: importance of proteasomal and autophagic degradative pathways in disposal of liver disease-associated protein aggregates. *Annu. Rev. Med.* 2011; 62:333–345. [PubMed: 20707674]
60. Schipanski A, et al. The lectin OS-9 delivers mutant neuroserpin to endoplasmic reticulum associated degradation in familial encephalopathy with neuroserpin inclusion bodies. *Neurobiol. Aging*. 2014; 35:2394–2403. [PubMed: 24795221]
61. Gelling CL, Dawes IW, Perlmutter DH, Fisher EA, Brodsky JL. The endosomal protein-sorting receptor sortilin has a role in trafficking α 1 antitrypsin. *Genetics*. 2012; 192:889–903. [PubMed: 22923381]
62. Nickson AA, Wensley BG, Clarke J. Take home lessons from studies of related proteins. *Curr. Opin. Struct. Biol.* 2013; 23:66–74. [PubMed: 23265640]
63. Chiti F, et al. Detection of two partially structured species in the folding process of the amyloidogenic protein β 2-microglobulin. *J. Mol. Biol.* 2001; 307:379–391. [PubMed: 11243826]
64. Jahn TR, Parker MJ, Homans SW, Radford SE. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* 2006; 13:195–201. [PubMed: 16491092]
65. Wright CF, Teichmann SA, Clarke J, Dobson CM. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*. 2005; 438:878–881. [PubMed: 16341018]

66. Feige Y, Buchner J. Principles and engineering of antibody folding and assembly. *Biochim. Biophys. Acta*. Jun 13.2014 doi:10.1016/j.bbapap.2014.06.004.
67. Eichner T, Radford SE. Understanding the complex mechanisms of β_2 -microglobulin amyloid assembly. *FEBS J*. 2011; 278:3868–3883. [PubMed: 21595827]
68. Chiti F, et al. A partially structured species of β_2 -microglobulin is significantly populated under physiological conditions and involved in fibrillogenesis. *J. Biol. Chem*. 2001; 276:46714–46721. [PubMed: 11590172]
69. Kameda A, et al. Nuclear magnetic resonance characterization of the refolding intermediate of β_2 -microglobulin trapped by non-native prolyl peptide bond. *J. Mol. Biol*. 2005; 348:383–397. [PubMed: 15811375]
70. Eichner T, Kalverda AP, Thompson GS, Homans SW, Radford SE. Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell*. 2011; 41:161–172. [PubMed: 21255727]
71. Eichner T, Radford SE. A generic mechanism of β_2 -microglobulin amyloid assembly at neutral pH involving a specific proline switch. *J. Mol. Biol*. 2009; 386:1312–1326. [PubMed: 19452600]
72. Esposito G, et al. Removal of the N-terminal hexapeptide from human β_2 -microglobulin facilitates protein aggregation and fibril formation. *Protein Sci*. 2000; 9:831–845. [PubMed: 10850793]
73. White HE, et al. Globular tetramers of β_2 -microglobulin assemble into elaborate amyloid fibrils. *J. Mol. Biol*. 2009; 389:48–57. [PubMed: 19345691]
74. Su Y, et al. Secondary structure in the core of amyloid fibrils formed from human β_2 m and its truncated variant N6. *J. Am. Chem. Soc*. 2014; 136:6313–6325. [PubMed: 24679070]
75. Karamanos TK, Kalverda AP, Thompson GS, Radford SE. Visualization of transient protein-protein interactions that promote or inhibit amyloid assembly. *Mol. Cell*. 2014; 55:214–226. [PubMed: 24981172]
76. Zijlstra M, et al. β_2 -microglobulin deficient mice lack CD4–8⁺ cytolytic T cells. *Nature*. 1990; 344:742–746. [PubMed: 2139497]
77. Ricagno S, Raimondi S, Giorgetti S, Bellotti V, Bolognesi M. Human β_2 microglobulin W60V mutant structure: implications for stability and amyloid aggregation. *Biochem. Biophys. Res. Commun*. 2009; 380:543–547. [PubMed: 19284997]
78. Esposito G, et al. The controlling roles of Trp60 and Trp95 in β_2 -microglobulin function, folding and amyloid aggregation properties. *J. Mol. Biol*. 2008; 378:887–897. [PubMed: 18395224]
79. Valleix S, et al. Hereditary systemic amyloidosis due to Asp76Asn variant β_2 -microglobulin. *N. Engl. J. Med*. 2012; 366:2276–2283. [PubMed: 22693999]
80. Mangione PP, et al. Structure, folding dynamics, and amyloidogenesis of D76N β_2 -microglobulin: roles of shear flow, hydrophobic surfaces, and α -crystallin. *J. Biol. Chem*. 2013; 288:30917–30930. [PubMed: 24014031]
81. Halabelian L, et al. Class I major histocompatibility complex, the trojan horse for secretion of amyloidogenic β_2 -microglobulin. *J. Biol. Chem*. 2014; 289:3318–3327. [PubMed: 24338476]
82. Corlin DB, Heegaard NH. β_2 -microglobulin amyloidosis. *Subcell. Biochem*. 2012; 65:517–540. [PubMed: 23225015]
83. Raimondi S, et al. The two tryptophans of β_2 -microglobulin have distinct roles in function and folding and might represent two independent responses to evolutionary pressure. *BMC Evol. Biol*. 2011; 11:159. [PubMed: 21663612]
84. Zoghbi HY, Orr HT. Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *J. Biol. Chem*. 2009; 284:7425–7429. [PubMed: 18957430]
85. Orr HT. Cell biology of spinocerebellar ataxia. *J. Cell Biol*. 2012; 197:167–177. [PubMed: 22508507]
86. Matos CA, de Macedo-Ribeiro S, Carvalho AL. Polyglutamine diseases: the special case of ataxin-3 and Machado-Joseph disease. *Prog. Neurobiol*. 2011; 95:26–48. [PubMed: 21740957]
87. Paulson H. Machado-Joseph disease/spinocerebellar ataxia type 3. *Handb. Clin. Neurol*. 2012; 103:437–449. [PubMed: 21827905]
88. Almeida B, Fernandes S, Abreu IA, Macedo-Ribeiro S. Trinucleotide repeats: a structural perspective. *Front Neurol*. 2013; 4:76. [PubMed: 23801983]

89. Nicastro G, et al. The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. *Proc. Natl. Acad. Sci. USA*. 2005; 102:10493–10498. [PubMed: 16020535]
90. Fiumara F, Fioriti L, Kandel ER, Hendrickson WA. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell*. 2010; 143:1121–1135. [PubMed: 21183075]
91. Pelassa I, et al. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum. Mol. Genet*. 2014; 23:3402–3420. [PubMed: 24497578]
92. Doss-Pepe EW, Stenroos ES, Johnson WG, Madura K. Ataxin-3 interactions with rad23 and valosin-containing protein and its associations with ubiquitin chains and the proteasome are consistent with a role in ubiquitin-mediated proteolysis. *Mol. Cell. Biol*. 2003; 23:6469–6483. [PubMed: 12944474]
93. Wang G, Sawai N, Kotliarova S, Kanazawa I, Nukina N. Ataxin-3, the MJD1 gene product, interacts with the two human homologs of yeast DNA repair protein RAD23, HHR23A and HHR23B. *Hum. Mol. Genet*. 2000; 9:1795–1803. [PubMed: 10915768]
94. Zhong X, Pittman RN. Ataxin-3 binds VCP/p97 and regulates retrotranslocation of ERAD substrates. *Hum. Mol. Genet*. 2006; 15:2409–2420. [PubMed: 16822850]
95. Nicastro G, et al. Josephin domain of ataxin-3 contains two distinct ubiquitin-binding sites. *Biopolymers*. 2009; 91:1203–1214. [PubMed: 19382171]
96. Nicastro G, et al. Understanding the role of the Josephin domain in the PolyUb binding and cleavage properties of ataxin-3. *PLoS ONE*. 2010; 5:e12430. [PubMed: 20865150]
97. Ellisdon AM, Pearce MC, Bottomley SP. Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J. Mol. Biol*. 2007; 368:595–605. [PubMed: 17362987]
98. Ellisdon AM, Thomas B, Bottomley SP. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J. Biol. Chem*. 2006; 281:16888–16896. [PubMed: 16624810]
99. Gales L, et al. Towards a structural understanding of the fibrillization pathway in Machado-Joseph's disease: trapping early oligomers of non-expanded ataxin-3. *J. Mol. Biol*. 2005; 353:642–654. [PubMed: 16194547]
100. Masino L, et al. Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3. *J. Mol. Biol*. 2004; 344:1021–1035. [PubMed: 15544810]
101. Hübener J, et al. N-terminal ataxin-3 causes neurological symptoms with inclusions, endoplasmic reticulum stress and ribosomal dislocation. *Brain*. 2011; 134:1925–1942. [PubMed: 21653538]
102. Masino L, et al. The Josephin domain determines the morphological and mechanical properties of ataxin-3 fibrils. *Biophys. J*. 2011; 100:2033–2042. [PubMed: 21504740]
103. de Chiara C, Menon RP, Kelly G, Pastore A. Protein-protein interactions as a strategy towards protein-specific drug design: the example of ataxin-1. *PLoS ONE*. 2013; 8:e76456. [PubMed: 24155902]
104. Kleywegt GJ, et al. Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-*trans*-retinoic acid and a synthetic retinoid. *Structure*. 1994; 2:1241–1258. [PubMed: 7704533]
105. Corsico B, Cistola DP, Frieden C, Storch J. The helical domain of intestinal fatty acid binding protein is critical for collisional transfer of fatty acids to phospholipid membranes. *Proc. Natl. Acad. Sci. USA*. 1998; 95:12174–12178. [PubMed: 9770459]
106. Dementiev A, Simonovic M, Volz K, Gettins P. Canonical inhibitor-like interactions explain reactivity of α 1-proteinase inhibitor Pittsburgh and antithrombin with proteinases. *J. Biol. Chem*. 2003; 278:37881–37887. [PubMed: 12860985]
107. McCoy AJ, Pei XY, Skinner R, Abrahams JP, Carrell RW. Structure of β -antithrombin and the effect of glycosylation on antithrombin's heparin affinity and activity. *J. Mol. Biol*. 2003; 326:823–833. [PubMed: 12581643]

108. Tysoe-Calnon VA, Grundy JE, Perkins SJ. Molecular comparisons of the β_2 -microglobulin-binding site in class I major-histocompatibility-complex α -chains and proteins of related sequences. *Biochem. J.* 1991; 277:359–369. [PubMed: 1859364]

Figure 1. Schematic two-dimensional functional energy landscapes

(a) Many proteins sample a near-native state (N^*) to open a binding site and interact productively with a ligand. Formation of the complex with ligand stabilizes the protein in its native state (N). (b) IDPs have sequences that disfavor a unique folded state and instead lead to many conformational possibilities of nearly equivalent energy. Interaction(s) with partners leads to stabilization of the state(s) that have the capacity to bind the partner with the highest affinity, thus shifting the energy landscape in the presence of partner(s) to one or only a few more stable states. (c) The magnitude of barriers on energy landscapes has a profound impact on protein function. A state that is not the thermodynamically most stable state (here M) may be long lived because it is separated from the most thermodynamically stable state (N) by high-energy barriers. The barriers may be reduced in amplitude by various triggers, whether environmental or protein-protein interactions.

Figure 2. CRABP1 exemplifies the structure and dynamics of the iLBP family

(a) Structure of holo-CRABP1 (ref. 104) (Protein Data Bank (PDB) code 1CBR) showing the helical portal domain, proposed to dynamically open to allow ligand entry and exit^{30,105} (light orange ellipse), and several regions implicated in early folding events in CRABP1 (ref. 22) (green). (b) Regions of CRABP1 found to be sequestered in the cores of aggregates formed either *in vitro* or in *E. coli* cells²¹ are shown in orange. Note that the early folding events may provide a potential protective mechanism to offset the risk of aggregation during folding of apo-CRABP1. (c,d) Comparison showing increased dynamics of CRABP1 in the apo form (c) relative to holo-CRABP1 (d). The width of the polypeptide chain depicts the rate of exchange of backbone amides with solvent water, a clear indicator of chain dynamics²⁸. Panel c is modified with permission from a figure in reference 28. Copyright 2000, American Chemical Society.

Figure 3. Functional and nonfunctional serpin conformational gymnastics

(a) On the left is shown the active serpin conformation with the solvent-exposed RCL (red), sheet A (yellow) and the shutter, which must open to allow RCL insertion into sheet A. A target protease is shown as a blue space-filling structure as it interacts with the RCL in an initial encounter complex. On the right is shown the conformational transition required for mechanical inactivation of target proteases. (PDB codes 1OPH¹⁰⁶ and 1EZX³⁹ on the left and right, respectively). (b) Serpins can form multimers and polymers *in vitro* and *in vivo*. Possible interactions include, from left to right: addition of the RCL to the edge of a β sheet as observed in the antithrombin III dimer (PDB code 1E05 (ref. 107)), partial insertion of the RCL into sheet A of an adjacent monomer⁴⁰ and domain swaps⁵².

Figure 4. Folding landscape of β_2m and its assembly with the MHC-1 heavy chain

(a) Schematic free energy landscape of β_2m monomer folding and aggregation. Φ_1 and Φ_2 indicate the increase of native intramolecular interactions during folding and non-native intermolecular interactions during aggregation, respectively. The unfolded protein with *cis*-Pro32 and *trans*-Pro32 are denoted U_C and U_T ; the intermediate ensembles are denoted as I_C , which is populated in vanishingly small amounts, and I_T , which is ~5% populated; and the native state is N . The fibril is indicated as Fib . (b) Structure of β_2m showing *cis*-Pro32 (green), Trp60 (blue) and the disulfide bond (C23–C80) (sticks) (PDB code 2XKS)⁷⁰. The regions involved in binding to the MHC1 heavy chain¹⁰⁸ are shown in orange on the surface. Panel a is reproduced from ref. 64.

Figure 5. Competition between ubiquitin binding and aggregation for the Josephin domain of Atx3

A surface representation of Josephin (PDB code 2JRI) is shown in white with exposed hydrophobic surfaces in green. When in the presence of the natural partner ubiquitin (Ub; the two binding ubiquitins are shown as blue traces), the Josephin binding sites are saturated by the interaction and protected from aggregation. When interaction is impaired, for instance by polyQ expansion, which alters the affinity for the interactor, the two sites promote aggregation, as shown in the right panel. Adapted from ref. 15 with permission from Elsevier.