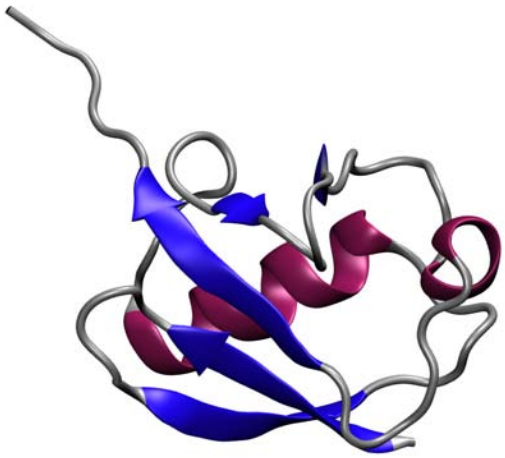
Case Study: Ubiquitin

Eduardo Cruz-Chu and JC Gumbart



1 Introduction

Without a doubt, the most organized and coordinated machine known is the biological cell. Inside its micrometer-scale diameter, a wide variety of macromolecules (DNA, proteins, sugars, lipids, etc.) work together in a co-operative way, balancing energy and matter to keep the cell alive. Within the cell, proteins are the overachievers. They allow the movement of water and ions through the cell membrane, help ATP to store energy, assist DNA during replication, recognize foreign infections, and more. However, all of these functions don't work independently of each other. To maintain harmony and efficiency between various functions, most processes have to be turned on or off according to different cellular stages and changes within the environment.

To this end, together with the mechanisms to assemble functional proteins and to turn on their functions, there should be counterparts to suppress and disassemble proteins when they are no longer needed. The cellular machine depends on assembly and disassembly to regulate the effective concentration of proteins and their corresponding activities [1]. Furthermore, defective

proteins also need to be removed. Consequently, protein degradation, the process of disposing of proteins, is a fundamental task. Inside the cell, such a critical function is a cooperative effort that depends on many different proteins, a "pathway" that involves different enzymes and reactions.

In this case study, we are going to focus on ubiquitin, a key player in eukaryotic intracellular protein degradation. Its relevance has been widely recognized in the scientific community, and the pioneering researchers, Aaron Ciechanover, Avram Hershko and Irwin Rose were awarded the Noble Prize in Chemistry in 2004 "for the discovery of ubiquitin-mediated protein degradation".
Here, we will use the following files:

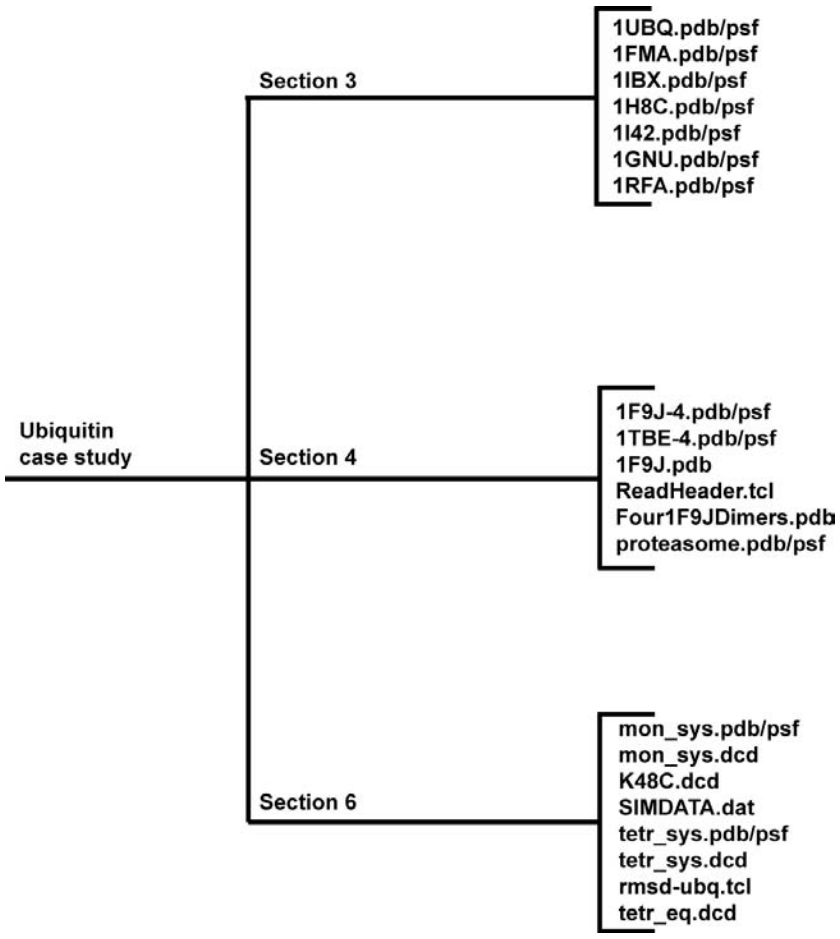


Figure 1: Sequence alignment of ubiquitins from different organisms, colored according to amino acid conservation. In blue are identical residues, in green are conserved substitutions, in light brown are semi-conserved substitutions, and in red there is no conservation among residues. In the far-right column, protein accession numbers to the NCBI database are given (http://www.ncbi.nlm.nih.gov/entrez/).

2 Ubiquitin and Evolution

As its name suggests, ubiquitin is found in many life forms. From humans to yeast, ubiquitin is consistently present throughout all eukaryotes. Remarkably, its genetic sequence is preserved without almost any modification. In Figures 1 and 2, we show a comparative analysis of different ubiquitins' amino acid sequences that illustrate the high degree of conservation.

Even though proteins are linear polymers, they do not assume a linear shape. In order to be functional, a protein must fold into a particular, usually compact geometry. The native conformation found in living cells is mainly determined by the amino acid sequence. Hence, proteins with similar sequences are expected to have similar folded structures. The more similar sequences are, the more likely it is that they share a common structure.

Ubiquitin's sequence conservation becomes obvious when we compare animals and plants. Humans, mice, pigs, guinea pigs, rabbits, chickens, and fruit flies have exactly the same ubiquitin sequence. A similar situation is seen with the soybean, garden pea, oat, wild oat, barley, wheat, maize, common sunflower, tomato, potato, garden asparagus, rice, carrot and turnip - all

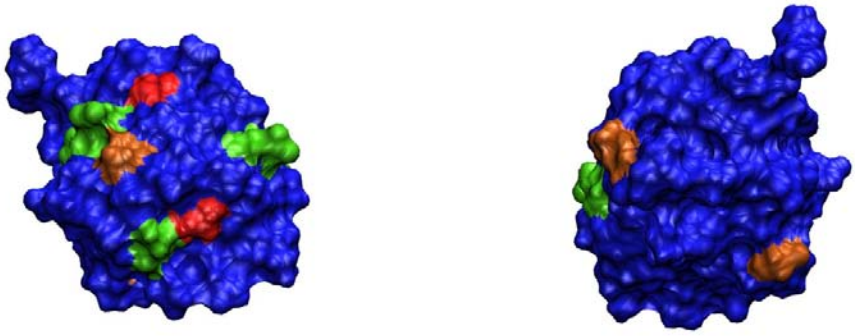


Figure 2: Surface representation of ubiquitin using VMD. The views are related by a 180^o rotation. Colors were assigned according to the amino acid conservation in Figure 1.

ubiquitins in this group share the same sequence. Between both groups, the difference in sequence is just two amino acids. That high degree of conservation has been considered indicative of the importance of each amino acid for the functionality of ubiquitin.

This conservation over millions of years of evolution leads to one conclusion about ubiquitin: its function is so crucial to the survival of any eukaryotic cell that it was practically perfected before multi-cellular organisms arose. In fact, it is a key regulatory label for many different cellular processes in addition to the degradation of either unassembled or misfolded proteins. These functions rely on ubiquitin's ability to be covalently linked to other proteins as well as on its particular structural features.

3 Ubiquitin's Profile

Besides its biological relevance, ubiquitin's physical and structural features make it an attractive candidate for experimental and theoretical studies of proteins. First, it is small, composed of just 76 amino acids and with a molecular weight of 8433 Da. It is also a high-temperature thermostable globular protein; it is very soluble and, at neutral pH, its folded structure is quite stable. To unfold ubiquitin through heating in solution, one needs to reach temperatures around 100 ^{o}C [3], i.e. the temperature of boiling water!

We can use structures resulting from x-ray crystallography to examine ubiquitin in more detail. Its compact structure becomes evident as seen in

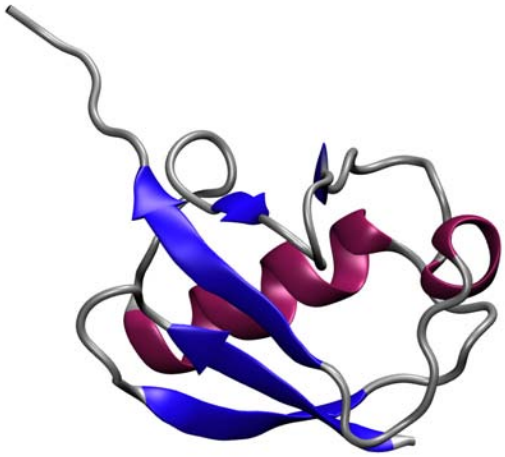


Figure 3: A structural view of ubiquitin with helices in purple, the β-sheet in blue, and turns and coil in grey [2]. (The figure can be reproduced using the included file 1UBQ.pdb.)

Figure 3. At this resolution, 1.8 Å, individual atoms can be seen, with the exception of hydrogens [2]. Ubiquitin's secondary structure also has three and one-half turns of α-helix, a short piece of 3_{10} helix (a helix with three residues per turn instead of 3.6 for α-helices), a mixed β-sheet that contains five strands, and seven reverse turns [2]. Its core is organized in a $\beta(2)$-α-$\beta(2)$ fashion known as the β-grasp fold. Many other proteins share this kind of fold, and due to the popularity of ubiquitin, it has also been called the ubiquitin-like fold (Figure 4). Look at ubiquitin in VMD and explain why is it called a β-grasp fold?

In Figure 5 we can see an alignment of different ubiquitin-like fold proteins. Multiple *structural* alignments do not include sequence information but rather just the three dimensional organization of the protein. These comparative analyses of proteins, at the level of sequence and structure, have became powerful tools. They are used to trace evolutionary relationships as well as to predict folded structures. However, the accuracy of alignment predictions relies on sequence and structure databases as well as alignment algorithms.

As previously mentioned, ubiquitin functions by covalently attaching itself to other proteins, known as substrate proteins. The process of conju-

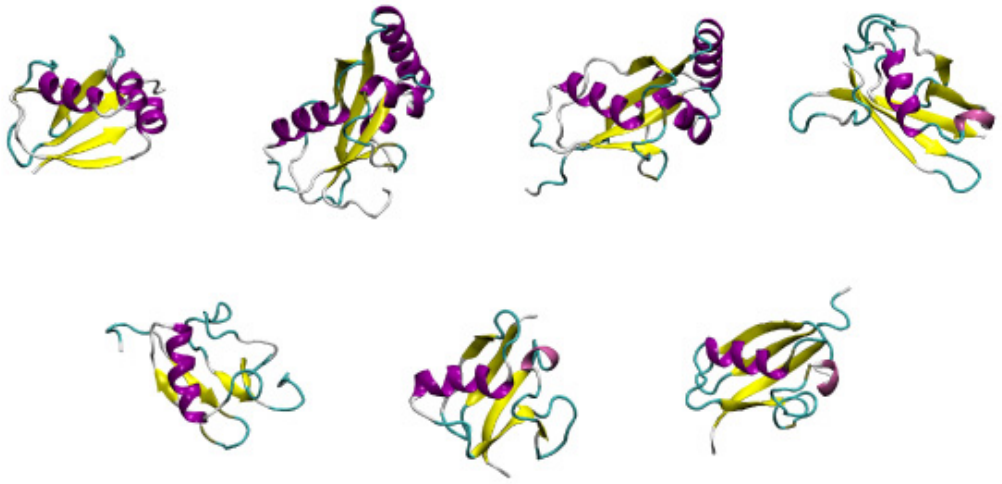


Figure 4: The simplicity of the β-grasp fold is shared by many different proteins. In the figures, from left to right; on the top : 1FMA.pdb, 1IBX.pdb, 1H8C.pdb, 1I42.pdb; on the bottom : 1GNU.pdb, 1RFA.pdb and 1UBQ.pdb. Try aligning these using the Multiseq extension of VMD, then look to Figure 5 to compare (all pdb files are provided).

Figure 5: Multiple structural alignment using VMD for the structures given in Figure 4. On the left, alignment colored according to *structure* similarity. Blue means high similarity, red low. Comparison with ubiquitin (in the center) shows that the $\beta(2)$-α-$\beta(2)$ motif is preserved. On the right, the same alignment is colored according to sequence similarity. *Sequence* similiarity is low among the different structures.

gation between ubiquitin and the substrate protein is called ubiquitination. The linkage is made using ubiquitin's GLY76 carboxyl group and the ε-amino group of a Lys residue in the substrate through an isopeptide bond. Once the bond is made, the attached ubiquitin can use a lysine residue to link another ubiquitin, producing a poly-ubiquitin chain. The number of ubiquitins and how they are linked determines the location in the cell where a tagged protein is going to go. When the substrate protein reaches its target, ubiquitin is detached and released in order to be used again. The conjugation process is energy driven by ATP. Ubiquitin is also assisted in the conjugation process by three other proteins that will be discussed in the next section.

By 1980, it was still not completely clear why proteins would need energy for degradation inside the cell. Outside the cell, catalyzed protein breakdown is energy independent. For example, proteins taken with food are degraded in the intestines before being absorbed. However, this energy requirement to degrade proteins in the cell is not without purpose.

4 Broad Functionality of Ubiquitin is the Result of Team Work

Ubiquitin definitely deserves its name; not only for being omnipresent, but also because it is involved in a diversity of cell functions. It participates in the G1 phase of the cell cycle, DNA repair, embryogenesis, immune defense, transcription, apoptosis and even preventing self-pollination in plants. How can a small protein with almost no variation among different organisms play so many different and vital functions in the cell?

This multiplicity of functions exists because ubiquitin works in a team. For ubiquitination and the subsequent degradation, the cell needs need more than ubiquitin, substrate proteins, and ATP. Ubiquitin is assisted by three groups of proteins, E1, E2, and E3, and a protein complex called the proteasome (discussed further in sections below); all of them work together composing the ubiquitin-pathway. E1, E2, and E3 handle the substrate protein and tag it with ubiquitin, providing the selectivity needed between both proteins. After ubiquitination is complete, the tagged protein is delivered to the proteasome which is in charge of the degradation itself. The energy provided by ATP is needed for controlling the specificity in the first and last step of the pathway.

Ubiquitin is only a label indicating that the tagged protein has to be degraded. This is why ubiquitin has been called "the kiss of death" protein. Beyond the relevance of ubiquitin's physical features, we have to keep in mind the whole ubiquitin pathway to understand how ubiquitin regulates cell functions. Malfunctioning of the ubiquitin pathway disrupts this regulation in the cell and can lead to neurodegenerative diseases, immune and inflammatory disorders, cystic fibrosis and cancer [4].

Once ubiquitin is attached, the protein's fate, usually degradation in the proteasome, is decided. It was first thought that the central function of ubiquitination was to recognize and destroy abnormal proteins, but later evidence has shown that the ubiquitin pathway can also destroy a "correctly" folded enzyme when it is no longer needed. In this instance, ubiquitin modulates protein turnover and, consequently, determines the half-life of an enzyme.

We can summarize the ubiquitin functions into three groups:

1. *Determination of short half-life for proteins at regulatory points.* Enzymes with regulatory function have to be turned on and off according to the cell's demands. Ubiquitination is a way to label enzymes for destruction, supressing their effects on the cell.

2. *Degradation of misfolded or damaged proteins.* The formation of abnormal proteins leads to an increase in the ubiquitination process. Even a growing protein in the ribosome that has not completed its sequence can be mistaken as a misfolded protein and ubiquitinated for degradation. It is estimated that 30 % of the newly-synthesized polypeptides are ubiquitinated and degraded [5].

3. *Non-proteolytic functions.* Mono-ubiquitinated proteins don't end up in the proteasome, but are directed towards endocytosis and secretion.

4.1 Insight into Function: The Ubiquitin Structure

By looking at the structure of ubiquitin alone, we can hope to derive some basic ideas about how it functions. For example, we can compare the hydrophobic/hydrophilic nature of the protein as seen in Figure 6. Looking at the external solvent-accessible surface, a majority, 62%, is covered with hydrophilic residues both charged and polar. While we naïvely might expect nearly 100%, the hydrophobic patches serve an important function; they are crucial for recognition at the proteosome [6]. By what mechanism is unclear.

Two possibilities are either for direct recognition of the patch itself or instead for ubiquitin-ubiquitin interactions to stabilize a specific structure and allow for the recognition of other residues [7].

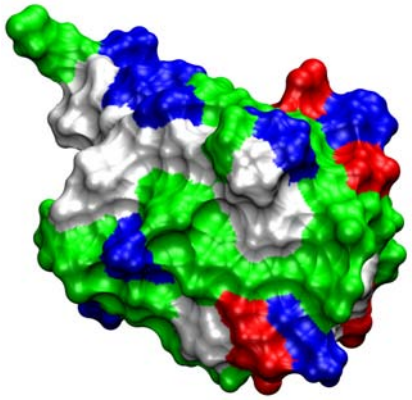


Figure 6: A view of the solvent excluded surface (MSMS in VMD) colored by residue type (red - negatively charged, blue - positively charged, green - hydrophilic, and white - hydrophobic. (Generate a similar view using the provided file 1UBQ.pdb.)

The secondary structure also reveals some clues as to the function of ubiquitin. It is strongly hydrogen-bonded with around 87% of the amino acids being part α-helical, β-sheet, or reverse turns [2] (this is shown in Figure 3). A quick look at ubiquitin's Ramachandran plot in Figure 7 confirms this. Proposed by G.N. Ramachandran in 1963, these plots were originally developed to explore secondary structure since they picture the sterically allowed regions based on the ψ and ϕ angles of the amino acid backbone [8]. This tightly bound secondary structure along with the compactness may be necessary to avoid unfolding ubiquitin itself while the substrate protein is unfolded.

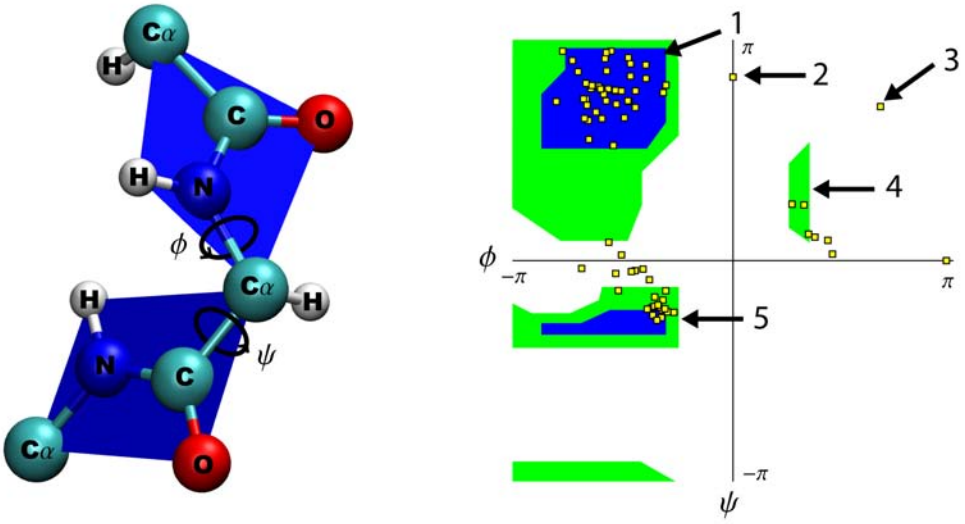


Figure 7: (*Left*) ψ, ϕ angles. Two peptide (N-C) bonds are shown along with three C_α atoms (side groups not shown). The N and C bonds are planar, as indicated. Therefore, between two residues, different *dihedral* angles (based on four consecutive atoms, the dihedral is the angle between the plane of the first three and the plane of the last three) can be defined. Two such common angles, ψ and ϕ are shown. (*Right*) Ramachandran plot for the cyrstal structure of ubiquitin. Points 1, 4, and 5 indicate the areas denoting typical secondary structures. Points 2 and 3 are specific residues worth identifying on your own.

Exercise 1: Ramachandran Plot. Load the file 1UBQ.pdb and make a Ramachandran plot for ubiquitin using VMD (in the menu Extensions/Analysis/Ramachandran Plot). Answer the following questions:
(a) In Figure 7, identify what secondary structure corresponds to each of the regions 1, 4 and 5.
(b) Points 2 and 3 are not located in defined secondary structure regions. Which residues correspond to these points and where are they located in the protein? (You can select points in the plot to view their properties.)
(c) Identify the location of GLY76 in the Ramachandran plot. Is it located in a defined secondary structure region? Do glycine residues have a broader or narrower range of conformations? (Consider the size of the side chains!) Why is the flexibility of GLY76 relevant for ubiquitin's function?
(d) The L-conformation of the amino acids produces the asymmetry of the Ramachandran plot. Why? (Hint : the D-conformation of the amino acids would give an inverted Ramachandran plot).

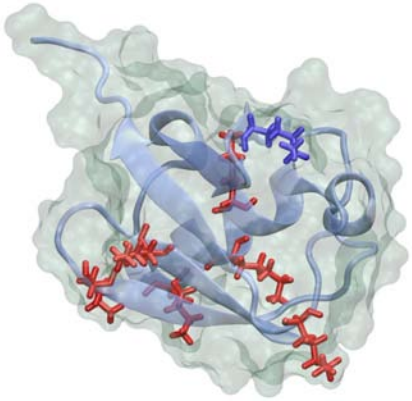


Figure 8: Ubiquitin is seen in both a cartoon and transparent MSMS representation. The seven conserved lysines are explicitly shown in the licorice representation and colored red with the exception of Lys48 which is shown in blue. This helps one see which residues are on the surface. (Prepare this view or a similar one using the file 1UBQ.pdb provided.)

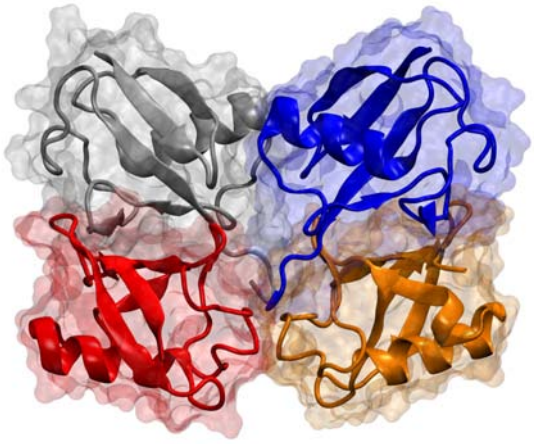


Figure 9: Four ubiquitins joined together through the C-K48 linkage. They are colored individually in order to see orientation as well as matching surfaces [10]. (Prepare the image yourself using the file 1F9J-4.pdb provided.)

We can also use the genetic sequence combined with structural data to gain new insight. Even though we know that almost all residues are well conserved, functionally the most important of these are the seven lysines, necessary to join ubiquitin to substrate proteins. By looking at the structure again (with the lysines highlighted) in Figure 8, we can see that at least four of them provide potential binding sites for forming ubiquitin chains due to their exposure on the surface [9]. The hydrophobic residues L8, I44, and V70 are known to be important to ubiquitin's function as well [6]; they form a hydrophobic patch together on the surface.

The crystal structure for a ubiquitin chain of four monomers reveals much about the process of ubiquitination [10]. This chain is formed by linking

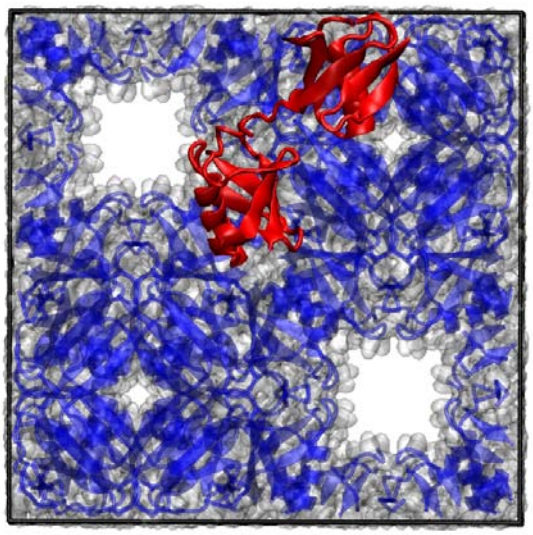


Figure 10: Polyubiquitin unit cell. A single dimer is colored in red.

from Lys48 (also called K48) of one ubiquitin to the C terminus of the next. Experimental studies show that at least four ubiquitins in this chain are required for effective recognition at the proteasome. The structure for this, a tetrameric form of ubiquitin, is shown in Figure 9. All four ubiquitins interact with the proteasome and, in fact, more tetrameric units actually increases the affinity between the chain and the proteasome [11]. From this, we can conclude that the tetramer is the actual unit for signaling in the case of proteolysis. This tetramer demonstrates how ubiquitin chains in various forms can represent multiple directives for proteins in the cell [11].

Exercise 2: Tetramer Structure. One of the most typical methods for resolving protein structures is a technique known as x-ray cyrstallography. With this method, protein crystals are grown and then placed in the path of an x-ray beam in order to produce a diffraction pattern. From this, electron densities and ultimately atom locations are determined.
Together with the atom coordinates (ATOM lines), the PDB file usually contains extra information like the expression system, structure resolution, authors, etc.

The script provided (ReadHeader.tcl) reads the transformation matrix (REMARK 290 lines in the pdb file) and produces the crystallographic unit cell (Figure 10). To use it, type:
- source ReadHeader.tcl
- DrawCUC 1F9J.pdb NoWrap

To see how the protein is packed within the cell, type:
- DrawCUC 1F9J.pdb Wrap

You can draw the neighboring unit cells of any of these two representations using VMD (Graphics/Representations/Periodic). Answer the following questions:
(a) As you can see, there is some "empty space" in the cell. That space is filled with water and ions that can not be solved by x-ray crystallography (unless they have fixed positions). In fact, as a general rule, a protein crystal is around 50% water and 50% protein. How then does the water-filled space affect the crystal's hardness? Would it be similar to a salt crystal?
(b) In the UB tetramer structure 1F9J.pdb, the linkage between the two UB dimers was not resolved because three C terminal residues on the second moiety were omitted. Your job will be to reconstruct the tetramer. Load Four1F9JDimers.pdb. There you have four dimers taken from the unit cell. Which pairs of dimers produce an allowed (K48-C) tetramer conformation? (There are more than two!)
(c) One major problem of x-ray crystallography is growing the protein crystal. Briefly describe another technique for solving protein structures, NMR, that does not require a crystal.

4.2 E1, E2, E3

The process of breaking down proteins is not unguided but rather a precise interplay among different enzymes. Degradation within the cell takes place through a methodical process that is carefully regulated, selecting only certain proteins for degradation, depending on the cell's requirements.

The variability in the enzymes E1, E2, and E3 allows ubiquitin to be

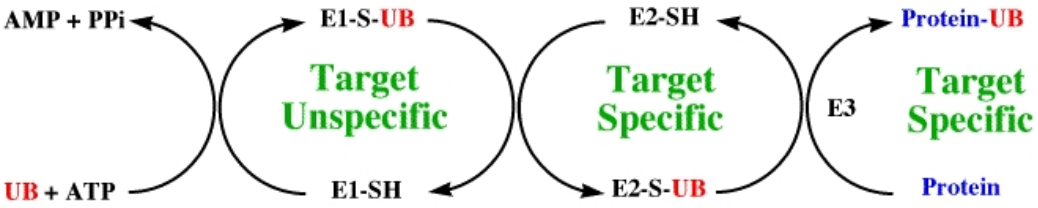


Figure 11: Schematic diagram of the reactions involving the E1, E2, and E3 enzymes and ubiquitin

tagged to many different substrate proteins. E1, E2, and E3 are families of enzymes and their structures have been tuned according to the substrate protein. While E1 doesn't have great variability, E2 and E3 are two large families of proteins and they can even be confined in specific cellular regions. This makes sense because E1 only binds ubiquitin while E2 and E3 are in close contact with the substrate protein; hence, their variability is a consequence of the diversity of the target.

How does the ubiquitin pathway decide when a subtrate protein has to be degraded? The answer is not completely clear yet. Ubiquitin enzymes have to recognize different degradation signals. Apparently, misfolded or damaged proteins feature flexible surfaces that would be recognized by the E3 enzymes.

Certainly, the identity of the N-terminal residue in the substrate protein is critical for determining the half-life of the substrate protein and it can also be considered as a degradation signal. Somehow, some amino acids in the N-terminal position provide a longer half-life than others [12]. Even though it is still uncertain why the N terminus would affect ubiquitination, it also seems to be a signal well preserved during evolution.

Other than the identification of the degradation signal, the ubiquitination process has been clearly understood.

The enzyme reactions involving E1, E2, and E3 are shown in Figure 11 and can be summed up in the following three steps.

1. E1, known as a ubiquitin-activating enzyme, activates ubiquitin using ATP. Ubiquitin is attached by its C-terminal Gly76 to a Cys in E1, forming a high-energy thioester bond. From here ubiquitin moves to E2.

2. E2, known as a ubiquitin-conjugating enzyme, receives ubiquitin from

E1 by making another thioester bond between a Cys in E2 and ubiquitin's Gly 76. It catalyzes the transfer from E2 to the substrate protein.

3. E3 is known as a ubiquitin-ligase. E3 doesn't make any covalent bonds with ubiquitin, but works together with E2 to ubiquitinate the substrate. E3 should be in charge of identifying the protein substrate.

4.3 Proteasome and the End of the Way

Once the protein is poly-ubiquitinated, it goes to its final destination: a cylinder-shaped, ATP-dependent protein complex called the proteasome. The proteasome is ultimately in charge of the degradation. Many copies of the proteasome are present in the cytoplasm and nucleus; proportionally, it comprises about 1% of the cellular proteins [5]. It exists in some form in both eukaryotes and prokaryotes indicating that it predates even ubiquitin.

The proteasome is composed of three sub-units, a central 20S complex in charge of the substrate proteolysis and two 19S complexes that cap the 20S from both ends. The 19S complex acts as a filter for the protein substrate. Polyubiquitin chains are recognized by the 19S complex; it has been hypothesized that the ubiquitin chain works as an anchor while the substrate protein is unfolded [13]. After all, one would not expect the proteasome to be capable of handling all proteins and the wide variety of sizes and shapes they have. In fact, the pore size in the proteasome for insertion of the protein is only around 10-20 Å meaning that most folded proteins will not fit [14]. Association between the 19S and the polyubiquitin chains would give more time to unfold the substrate, facilitating the translocation through the inner face and the consequent degradation. Furthermore, the 19S complex removes the ubiquitin chain before translocation so that ubiquitin can be used again. Once inside the 20S complex, the substrate is degraded into small peptides of seven to nine amino acids that can be later degraded into single amino acids to be reused again. Thus, we can say that the substrate protein is recycled.

There are a few structures available for portions of the proteasome. One we can briefly examine is the 20S subunit of a mammalian proteasome, located in proteasome.psf and proteasome.pdb [15]. Shown in Figure 12, it is a very large complex comprised of almost 100,000 atoms. The structure is composed of four heptameric rings with 28 subunits overall. Using VMD, we can also explore the electrostatics of the proteasome. By running the VMD plugin PME Electrostatics, we can generate what is called an isosur-

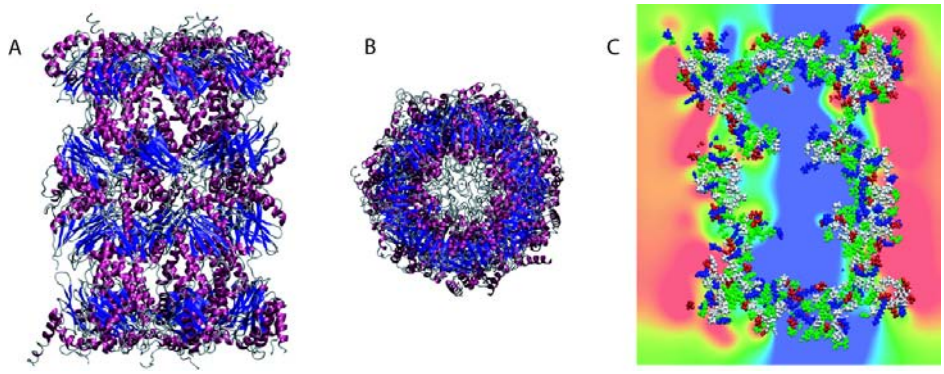


Figure 12: A picture of the mammalian proteasome 20S subunit. **(a)** A side view of the subunit shown in the same representation as Figure 3. **(b)** A top view of the same protein complex. **(c)** A PME slice showing the channel inside. A portion of the protein is shown in van der Waals representation colored by restype. The PME plot is an electrostatic potential energy map of the proteasome's center. The color scale goes from red (negative) to green (neutral) to blue (positive). (If your computer permits loading this large structure, view it using the provided file proteasome.pdb.)

face through the center of the proteasome. This surface is a map of the potential in this area. In Figure 12c, we can see there are many charged residues, one common property of soluble proteins. We can also see the inside is highly positive while the outside is highly negative; the magnitudes of this property can be flexible, though, as protonation states of certain residues (histidine for example) can change depending on the environment.

The positive potential inside is due in large part to lysine residues spread throughout the interior. These lysines are key to the function of the threonine proteases that work to split the incoming-protein. The lysine aids the threonine in becoming more nucleophilic, i.e., wanting protons. Then the threonine is able to carry out a *nucleophilic attack* on incoming peptides [16]. By looking at the structure with VMD, you can see many Thr/Lys pairs exist in the proteasome.

Despite the fact that the specific interactions between ubiquitin and the proteasome are not fully understood, considering that the substrate protein is at least partially unfolded while ubiquitin is still attached, the stability of ubiquitin most likely plays an important role in this process. In order to investigate this idea further, both experiments and simulations have been employed. These will allow us to better understand what is precisely meant

by stability of a protein and where that stability arises from.

Exercise 3: 20S subunit. The proteasome is composed of two 19S subunits, in charge of collecting the proteins from solution, and the cylinder-shape 20S subunit, in charge of the proteolysis. The latter one has several proteases (active sites) that are responsible for chopping the proteins into small polypeptide pieces. Here we will explore the structure of the 20S subunit, the only one currently with an available structure.
(a) Describe the subunits of the proteasome. Be sure to include what names are given to them and what their general function is (this does not have to be very detailed). Which subunits are known to contain active proteolytic sites? (see Unno, *et al* [15], for example).
(b) Load the structure proteasome.psf/proteasome.pdb. Unno *et al.* describes a novel active site in the subunit β7. Find the residues involved (residues 1, 56, 59, 99, and 104 of chain N; residues 88 and 91 of chain H) and make a figure representing them using VMD.

5 Ubiquitin Stability: Experiments

To study the stability of ubiquitin, single-molecule atomic force microscopy has been used to pull one end of a ubiquitin chain at a constant velocity with the other end fixed while measuring the force during the process [9]. While the force required has some velocity dependence, interesting comparisons can be made between chains linked through different residues. Constant force pulling has also been used to study the unfolding and re-folding rates of polyubiquitin chains; this has the potential to allow one to determine specific events in the process [17].

In the case of constant velocity pulls, chains with two linkages have been studied. One involved two to seven ubiquitins linked from Lys48 of one moiety to the C terminus of the next. The other used nine repeats linked from the N terminus of one to the C terminus of the next. For the chain linked through Lys48-C, the average force required to pull at 300 nm/s was found to be 85 pN. However, for an N-C linked ubiquitin chain, the force required to pull at 400 nm/s was about 200 pN [9]. Clearly stability is not a singular idea when discussing a protein. Much like man-made structures have various points of greater or lesser stress, proteins also react differently when probed in different ways. Much of this difference likely comes from

the different geometries of the poly-ubiquitins. For example, the β-sheet will be more or less difficult to separate depending on the relative direction of the applied force; this is the difference between unzipping the β-strands one hydrogen bond at a time versus shearing them all at once. As suggested by some researchers, this could be how ubiquitin chains linked through different residues can function in different ways [9, 17].

6 Ubiquitin Stability: Simulations

Molecular dynamics simulations can act as a useful complement to experimental work. Using Steered Molecular Dynamics along with NAMD, we can analyze unfolding events and see at an atomic level what the limiting factors are in the process [18, 19]. We can simulate both constant velocity pulling and constant force, similar to the abilities of AFM experiments [9, 17]. There is a caveat; due to limitations in technology, we can currently only simulate on the nanosecond timescale. In order to see a protein unfold in such a short time, we have to use what would be unreasonable velocities and forces in experimental studies. For example, we will pull ubiquitin with a velocity on the order of 10^9 nm/s whereas in experiment the velocities used are on the order of 10^2 nm/s [9]. This does not invalidate the results obtained by this method, but one should be cautious in interpreting them.

6.1 Unfolding the Ubiquitin Monomer

As a prelude to unfolding ubiquitin chains, we begin by analyzing the unfolding pathway of a ubiquitin monomer. In this simulation, we can correlate specific events in the breaking of the secondary structure with peaks in the force necessary to do so. Because of the dependence on velocity though, the forces computed here will be larger than those seen in experiment. However, we can still make qualitative comparisons between different events in order to aid our investigations.

Here, we will explore the stability of ubiquitin by simulating a constant velocity pull on the N terminus and Lys48. This samples the force that the first ubiquitin monomer would feel in the tetramer simulation in the next section. The details of the simulation performed can be found at the end of this section.

Since this is a constant velocity simulation, we can easily measure the

force needed to maintain this velocity over the course of the run. By comparing it to the distance between LYS48 and MET1 (which actually should follow the time approximately linearly), we can see the unfolding of different secondary structures in the protein. We can also correlate specific points in the plot of force vs. distance (Fig. 13) with atomic-level events in the unfolding process.

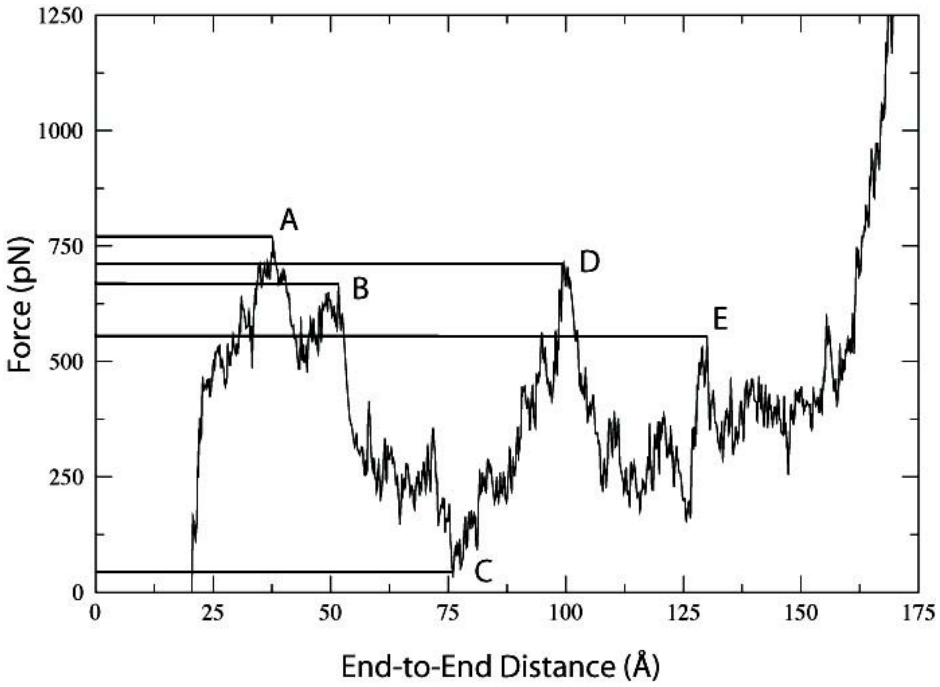


Figure 13: This plot compares the force and the separation during a constant velocity pull of .05 Å/ps of a ubiquitin monomer between the N terminus and Lys48. Points A, B, C, D, and E represent the breaking of different secondary structures. This is analogous to the stretching of a rubber band; the force is at a maximum immediately before it breaks. (View the trajectory corresponding to this graph using the provided files 1UBQ.psf and mon_sys.dcd. Monitor the extension of the K48-M1 distance and see if you can relate the plot to specific unfolding events.

The first structure to come apart are the β-strands; this occurs mostly as "unzipping" of one from the other, i.e., one or two hydrogen bonds at a time. At point A, the smallest strand separates from the group. At B, the middle strand of a triplet is almost completely pulled out. As the last few hydrogen bonds here are broken, the force comes to a minimum at point C.

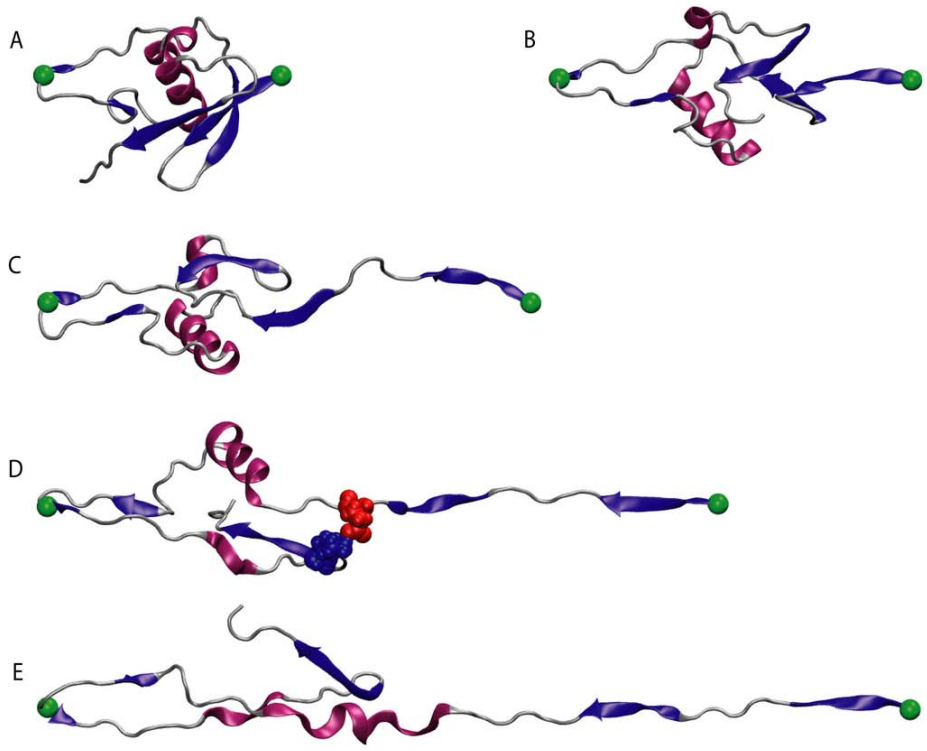


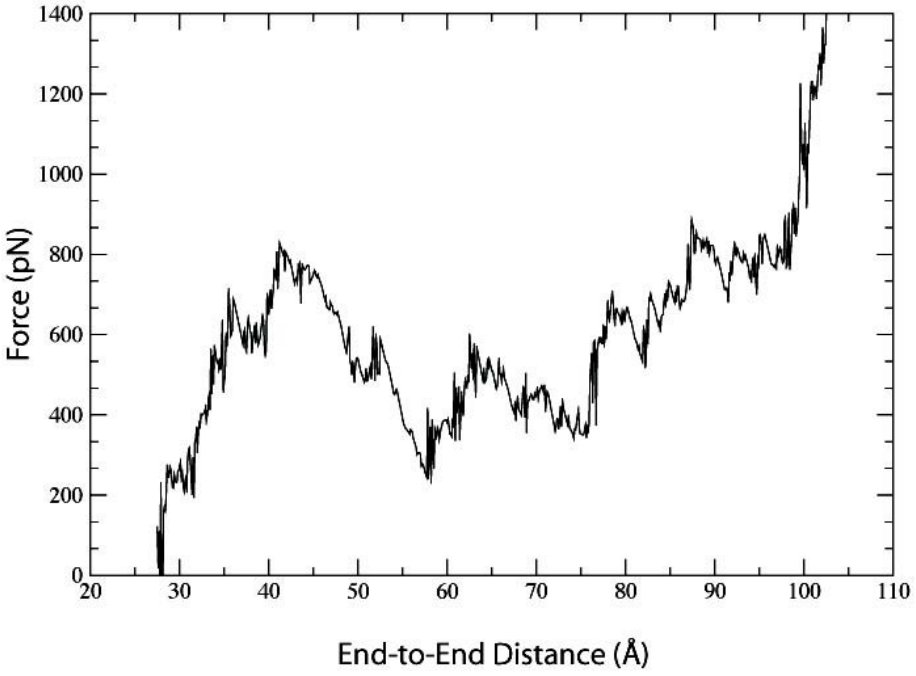
Figure 14: An illustration of points A-E in the plot shown in Figure 13. Each uses the same representation and coloring as in Figure 3 with the exception of D which also uses the van der Waals representation and restype coloring to illustrate the salt bridge between residues GLU18 and LYS63. (View the corresponding trajectory using the files 1UBQ.psf and mon_sys.dcd.)

Since at this point, much of the recent extension has come from the straightening of a coil, the hydrogen-bonded structure begins to come into play again. This increase is seen all the way to point D where a salt bridge between two charged residues (GLU18 and LYS63) is broken after which the force drops again temporarily. At E, we see the α-helix unraveling completely. Finally, past 160 Å the protein's full extension has been reached and stretching of the peptide bonds themselves occurs.

Simulation Details. The structure used for the monomer was obtained from the Protein Data Bank (1UBQ), and hydrogens were added using VMD along with the plugin tool PSFGen [20]; the result of this can be found in the provided files 1UBQ.psf and 1UBQ.pdb. In order to use as much water as reasonably possible, the protein was solvated in a water box of dimensions 63 Å x 65 Å x 180 Å bringing the system to a total of 70,000 atoms (ubiquitin itself is around 1,200 atoms). While water is necessary to mimic the natural cellular environment, it very quickly increases the computational difficulty of the simulation. For this simulation, we used periodic boundary conditions where the unit cell, our system, is effectively duplicated in all directions surrounding it. This may seem unreasonable at first, but we have enough water, at least 15 - 20 Å, to buffer ubiquitin from interacting with itself [21]. Periodic boundary conditions also allow us to use Particle Mesh Ewald (PME) electrostatics which is considered to be a more accurate method for long range interactions compared to merely "cutting off" the interaction at some finite distance [22, 23] (also see [24, 25] for a further discussion of this). The resulting system with water can be seen using the provided files mon_sys.psf and mon_sys.pdb.

Once the system was set up, it was equilibrated for .5 ns with the protein backbone fixed and .75 ns with everything free to move. Then for the SMD portion of the simulation, the α-carbon of the first residue, Met1, is pulled along the line between it and the α-carbon of Lys48 which is held fixed. For this simulation, a velocity of .05 Å/ps was used, and the force constant for the SMD pulling was 208.4 pN/Å. The total time is about 3.15 ns, and the full trajectory, with frames taken every 10 ps, is available in the form of the provided files 1UBQ.psf and mon_sys.dcd from which waters have been removed in order to reduce file size.

Exercise 4: Pulling from K48 to the C terminus. As a contrast to the NK48 pull, we will now analyze an SMD constant velocity pull between LYS48 and the C terminus, GLY76. To view this, load the file 1UBQ.psf and then the trajectory K48C.dcd into VMD. Recreate the graphical representation shown in Figure 14. Label the end points by selecting the specific α-carbons of the residues LYS48 and GLY76 and putting them in the vdW representation, then create a label for the distance between them. Discuss the features of the plot shown below and try to explain them by examining the corresponding trajectory in VMD. In particular, look for the following things.
(a) Identify the force peak near 45 Å and explain what happens in the trajectory to give this peak.
(b) Determine at what distance the first two β-strands separate from the last one (relative to the N terminus). Use both the plot as well the Hydrogen bonds in the protein to help.
(c) Make an image of your representation at both the beginning and the end of the trajectory.



A plot similar to that in Figure 13 showing the separation between LYS48 and GLY76 compared to the applied force needed to maintain a velocity of .1 Å/ps along the direction of separation. The entire pull is approximately 1.2 ns long.

Exercise 5: The Worm-like Chain Model for Proteins. The worm-like chain model has proven to be a useful polymer model for both DNA and proteins. It assumes a continuously flexible, uniform rod, but in its simplest form does not account for higher order structure. Nonetheless, it has been used successfully for interpreting experimental AFM data such as that described in a previous section [9]. In this model, the force required to extend the polymer can be expressed as

$$\frac{Fl_p}{k_BT} = \frac{z}{L_C} + \frac{1}{4(1 - z/L_C)^2} - \frac{1}{4} \tag{1}$$

where z is the extension, L_C is the contour length, and l_p is the persistance length.

We will now use this model to examine data provided from the last simulation of unfolding ubiquitin from the N terminus to K48.

(a) Determine the behavior of z in the high and low force limits ($F \rightarrow \infty$, $F \rightarrow 0$). Keep in mind physical limitations on z. (If you cannot do this using exact methods, you may also try numerically solving the equation for large and small values of F to determine the dependence of z).

(b) The simulation was performed at 300 K. In units of pN*Å, what is the thermal energy, k_BT?

(c) Open the file 1UBQ.pdb in VMD. In the Graphical Representations window, select "Trace" for the protein; it should resemble a jointed chain. Measure the length of a handful of segments, each one representing one peptide (using the "Labels $\rightarrow$ bonds" option and the Labels window). Use an average value to calculate L_C, the extended length, between the N terminus and K48.

(d) Since our model does not take into account the detailed interactions involved in unfolding a protein, we will only examine the last part of the extension, from 132 Å to 171 Å (see Figure 13). The data from this portion of the simulation is provided in the file SIMDATA.dat. The first column is extension in units of Å and the second column is force in units of pN. Fit this data to the model above using our computed values of k_BT and L_C. What is the fit value of the persistance length, l_p? (Note: while you may have to manipulate the form of the data or the equation, this problem does not require advanced programs to solve).

(e) Plot your equation now along with the data to visually examine your fit. A widely accepted value of the persistance length is 4 Å [26]. How does your computed value compare to this?

6.2 Unfolding the Ubiquitin Tetramer

The structure of the ubiquitin tetramer was thought to be known definitively when it was crystallized in 1994 [27]. However, it was crystallized again in 2001 and a somewhat different conformation was discovered [10]. One key difference is the location of the hydrophobic patch, known to be crucial for recognition at the proteasome (namely LEU8, ILE44 and VAL70); it is shown to be both exposed and buried [10]. This shows that tetramers may be flexible with different conformations, possibly playing an important functional role. Both structures are provided in 1TBE-4.pdb and 1TBE-4.psf for the first and 1F9J-4.pdb and 1F9J-4.psf for the latter.

For this simulation, the second crystal structure was used [10]. Constant force was used in this simulation to pull the N terminus of the first ubiquitin and LYS48 of the fourth ubiquitin; by the end of the simulation, the entire structure was unfolded. In order to look for specific unfolding events in the data, a plot was made of the distance between the two points versus time.

The simulation was run for a total of about 3.5 ns at which point the protein was completely unfolded, the results of which can be seen in tetr_sys.dcd (note that the waters have also been removed here).

There are some noticeable events revealed by this plot. The unfolding is pretty steady for the first 0.9 ns as the first and last ubiquitins unfold (the first completely but the last only partially since we are pulling at Lys48 and not the C terminus). Then there is a dip in the distance at point A. This represents the resistance of the hydrophobic contact between the second and third subunits. After this is broken, some unfolding occurs again normally until B. Here, the second and third subunits are beginning to unfold. Finally, at point C, all unfolding has occurred and the end-to-end distance levels off. These separate events are also represented in Figure 16.

Some conclusions can be drawn from this. First, the effect of the hydrophobic contacts between subunits is noticeable, but small, relative to, say, the unfolding of an entire subunit. While this does not preclude it from being important for stability, it does mean its role in preventing unfolding at the proteasome is likely small. Also, while the exterior subunits A and D unfold fairly easily, the interior ones resist unfolding for longer (on the order of 400 ps more). This could be why chains of ubiquitin are more functional than simple monomers or dimers in proteolysis. However, one must remember that the results of these simulations could be artifacts of the large force used and not actually representative of how ubiquitin chains behave.

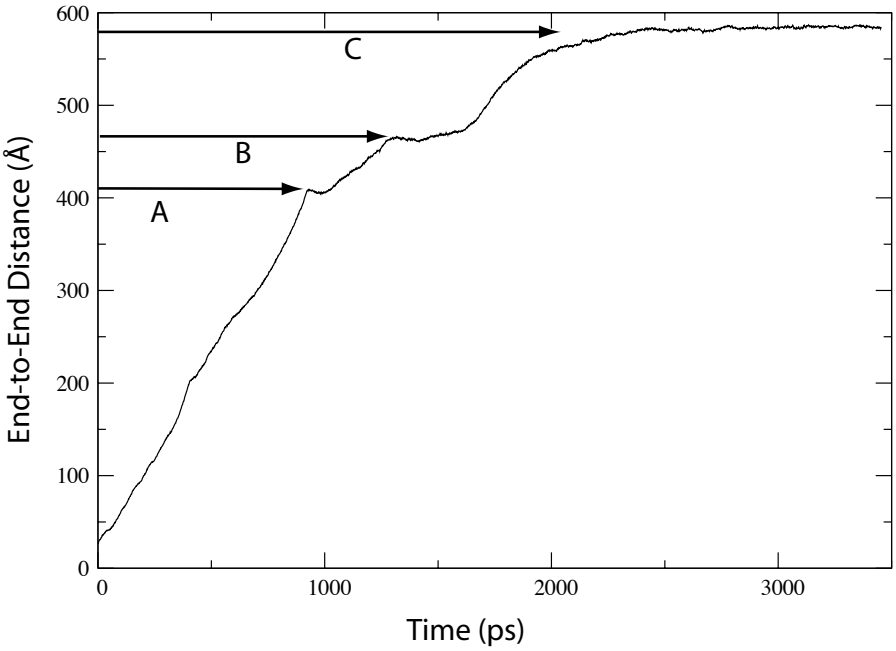


Figure 15: Results of a simulation pulling apart a ubiquitin tetramer at 1039 pN; A is the breaking of contact between the second and third subunits, B is the unfolding of the same subunits, and C is the point at which complete unfolding has occurred. (View the trajectory corresponding to this plot using the provided files 1F9J-4.psf and tetr_sys.dcd.)

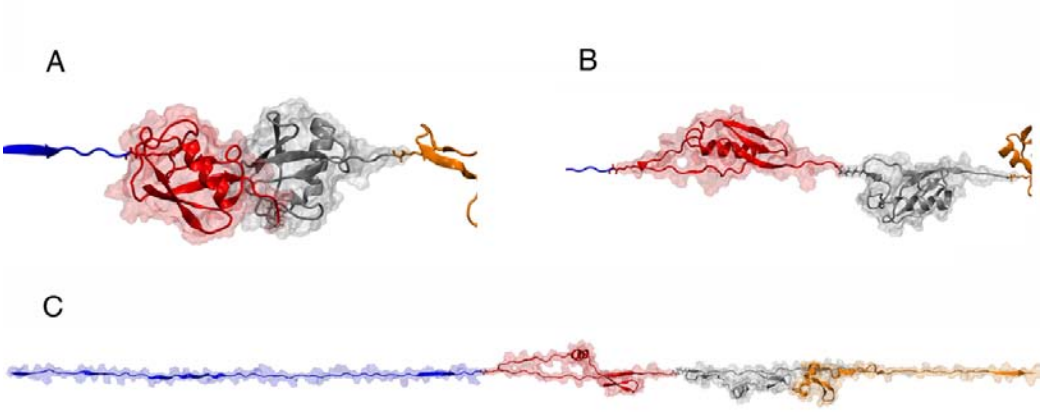


Figure 16: Unfolding events A, B, and C from Figure 15 shown in the same representation as Figure 9. (View the corresponding trajectory using the provided files 1F9J-4.psf and tetr_sys.dcd.)

Simulation Details. In order to explore the tetramer stability, a different approach from that for the monomer is taken. Because the size of the water box would be prohibitively large, instead a water sphere was used. The main difference in this is that since we can not include enough water to effectively buffer protein-protein interactions between cells, periodic boundary conditions will not be employed. Still, the amount of water is always an issue, and now we also have to contend with evaporation into the vacuum. Nonetheless, this simulation technique can strike a good balance between perfecting the electrostatics and minimizing the computational time. In this case, we are able to keep the system size down to 62000 atoms (5000 for the protein and 57000 for the water); the full setup can be inspected with VMD using the provided files tetr_sys.pdb and tetr_sys.psf. For the simulation, a constant force of 1039 pN was used to pull Lys48 of the fourth moiety while fixing the N terminus of the first moiety. The simulation was run for a total of about 3.5 ns, the results of which can be seen in 1F9J-4.psf and tetr_sys.dcd (note that the waters have also been removed here).

Exercise 6: (In)Stability of the 1F9J Closed Tetramer Configuration. It has been suggested that ubiquitin chains are inherently flexible due to the various conformations observed in different crystal structures [10]. In a previous exercise, you had to reconstruct the 1F9J tetramer, finding in fact that two unique conformations are possible. One, the closed conformation, is what we are examining in this case study, although another one was already completely open.

Since it is obvious the open one is flexible, we should attempt to determine if the closed one is also flexible. An equilibrium simulation of the 1F9J closed tetramer is provided in the files 1F9J-4.psf and tetr_eq.dcd. While just looking at the trajectory may tell us qualitatively the result, we can also examine it quantitatively. We will calculate the Root Mean-Square Deviation of the protein backbone of the ubiquitin tetramer. The RMSD effectively tells us how the structure changes over time by giving the average displacement of each protein residue from the initial structure after performing a least-squares fitting of the entire structure to the initial one.

(a) A script to calculate the RMSD is provided (rmsd-ubq.tcl). After loading the structure and trajectory into VMD, type "source rmsd-ubq.tcl" in the Tk Console. It will output five files containing the RMSD (measured in Å) for each of the subunits and for the whole chain. Plot all five sets of data on a graph and label them.

(b) How does the stability of the individual units compare to the stability of the entire conformation? What does this say about the difference between inter-domain and intra-domain contacts?

References

[1] A. Varshavsky. The ubiquitin system. *Trends Biochem. Sci.*, 22:383–387, 1997.

[2] S. Vijay-Kumar, C.E. Bugg, and W.J. Cook. Structure of ubiquitin refined at 1.8 å resolution. *J. Mol. Biol.*, 194:531–544, 1987.

[3] B. Ibarra-Molero, V.V. Loladze, G.I. Makhatadze, and J.M.Sanchez-Ruiz. Thermal versus guanidine-induced unfolding of ubiquitin. an analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry*, 38:8138–8149, 1999.

[4] A. Ciechanover and K. Iwai. The ubiquitin system: From basic mechanisms to the patient bed. *IUBMB Life*, 56:193–201, 2004.

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell 4th Ed.* Garland Science, 2002.

[6] R. Beal, Q. Deveraux, G. Xia, M. Rechsteiner, and C. Pickart. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl. Acad. Sci. USA*, 93:861–866, 1996.

[7] R. Varadan, O. Walker, C. Pickart, and D. Fushman. Structural properties of polyubiquitin chains in solution. *J. Mol. Biol.*, 324:637–647, 2002.

[8] B.K Ho, A. Thomas, and R. Brasseur. Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the alpha helix. *Protein Science*, 12:2508–2522, 2003.

[9] M. Carrion-Vazquez, H. Li, H. Lu, P.E. Marszalek, A.F. Oberhauser, and J.M. Fernandez. The mechanical stability of ubiquitin is linkage dependent. *Nature Struct. Biol.*, 10:738–743, 2003.

[10] C.L. Phillips, J. Thrower, C.M. Pickart, and C.P. Hill. Structure of a new crystal form of tetraubiquitin. *Acta Cryst. D*, 57:341–344, 2001.

[11] J. Thrower, L. Hoffman, M. Rechsteiner, and C. Pickart. Recognition of the polyubiquitin proteolytic signal. *EMBO J.*, 19:94–102, 2000.

[12] A. Varshavsky. The n-end rule: Functions, mysteries, uses. *Proceeding of the National Academy of Sciences*, 92:12142–12149, 1996.

[13] A. Hershko and A. Ciechanover. The ubiquitin system. *Ann. Rev. Biochem*, 67:425–479, 1998.

[14] M. Bochtler, L. Ditzel, M. Groll, C. Hartmann, and R. Huber. The proteasome. *Ann. Rev. Biophys. Biomol. Struct.*, 28:295–317, 1999.

[15] M. Unno, T. Mizushima, Y. Morimoto, Y. Tomisugi, K. Tanaka, N. Yasuoka, and T. Tsukihara. The structure of the mammalian 20S proteasome at 2.75 å resolution. *Structure*, 10:609–618, 2002.

[16] D. Voges, P. Zwickl, and W. Baumeister. The 26S proteasome: A molecular machine designed for controlled proteolysis. *Ann. Rev. Biochem*, 68:1015–1068, 1999.

[17] J.M. Fernandez and H. Li. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*, 303:1674–1678, 2004.

[18] Laxmikant Kalé, Robert Skeel, Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comp. Phys.*, 151:283–312, 1999.

[19] Barry Isralewitz, Mu Gao, and Klaus Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Op. Struct. Biol.*, 11:224–230, 2001.

[20] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[21] O.N. deSouza and R.L. Ornstein. Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophys. J.*, 72:2395–2397, 1997.

[22] O.N. deSouza and R.L. Ornstein. Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. *J. Biomol. Struct. Dyn.*, 16:1205, 1999.

[23] T. Fox and P.A. Kollman. The application of different solvation and electrostatic models in molecular dynamics simulations of ubiquitin: How well is the X-ray structure "maintained"? *Proteins: Struct., Func., Gen.*, 25:315–334, 1996.

[24] C. Sagui and T.A. Darden. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Ann. Rev. Biophys. Biomol. Struct.*, 28:155–179, 1999.

[25] P.H. Hünenberger and J.A. McCammon. Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study. *Biophys. Chem.*, 78:69–88, 1999.

[26] H. X. Zhou. Polymer models of protein stability, folding, and interactions. *Biochemistry*, 43:2141–2154, 2004.

[27] W.J. Cook, L.C. Jeffrey, E. Kasperek, and C.M. Pickart. Structure of tetraubiquitin shows how multiubiquitin chains can be formed. *J. Mol. Biol.*, 236:601–609, 1994.