Molecular dynamics investigation of the structure-function relationships in proteins with examples from Hsp70 molecular chaperones, αA-crystallin, and sericin

PhD Thesis

Ewa Irena Gołaś

SUMMARY

Structure function relationships play a central role in protein science—they define the functionally relevant behaviors and interactions that govern biological systems. Relationships can be thought to span a spectrum, where on one end, the function of a protein is to provide structure, and on the other end, the protein's structure lends to its function. Unraveling the structure function mystery therefore entails an investigation of the interplay between the notions of structure and function. My work provides insight into the structure function relationships operational in three representative proteins: the Hsp70 (Heat shock protein 70kDa) molecular chaperone, whose function is to promote the proper structure of client proteins; αA-crystallin, which stands at the middle of the structure-function dipole, expressing both chaperone functionality and purpose as a structural protein in the lens of the vertebrate eye; and finally, a sericin-based biopolymer, which puts structural relationships to the test by engineering a novel biomaterial based solely on an understanding of the structural behavior of protein motifs in silk. Moreover, occurring at the micro scale level, the study of such relationships renders itself readily to investigation by molecular dynamics (computational) techniques, which can provide in-depth perspective and detail to inherent protein mobility and structural change.

The Hsp70 molecular chaperone consists of two subdomains—the nucleotide binding domain (NBD), which functions as an ATP-ase, and the substrate binding domain (SBD), which captures client substrate. A complex allosteric network communicates both the presence of nucleotide and substrate between subdomains. I took an all-atom perspective on the NBD (from *Bos Taurus*, pdb 3C7N:B), in an effort to identify the behavior of the domain in response to bound nucleotide. Canonical molecular

dynamics simulations were performed using the AMBER all-atom model and force field, after which an Essential Dynamics analysis (a Principal Component Analysis-based method) enabled a decomposition of the behavior of the domain into a series of eigenvectors, i.e. trend motions. As such, the domain was shown to express a set of basal dominant trend motions, which occurred regardless of the binding state. These focused on the module-like rotation of individual subunits; indeed, this trend is concordant with experimental evidence suggesting the latter. Clearly, re-orientation of individual subunits changes the internal in-plane (δ) and out-out-plane (τ) angle between the lobe-like halves that constitute the NBD. Concomitant to this motion are secondary vectors that are specific to the binding state. Instead of concentrating on rotation, these manifest themselves in the form of hot-spots—small regions on the surface and at subunit interstices that exhibit enhanced mobility. These were identified as key sites in the allosteric network of the NBD. Depending on the binding state, a distinct set of hotspots was triggered. In fact, many of the hotspots pin-pointed herein overlap with those proposed experimentally. Moreover, for the ATP-bound state, the enhanced contribution of state-specific vectors to the variance of the system was found to restrict the global trend of subunit rotation, and so this state explored a more conservative and correlated range of δ and τ angle than did the ADP-bound state. The identity of the nucleotide therefore influences subdomain rotation, and is especially transduced into the mobility of hotspots. Thus, the protein dynamics of the NBD reveals a complex nucleotide-dependent structure function relationship that is based on subdomain rotation and on the instigation of an internal allosteric network.

I simulated the whole chaperone (from *E. coli*, pdb 2KHO) employing the UNRES coarse-grained model and force field. In this case, the various binding states of the NBD were modeled by restraining its conformation to a form that reflected the subdomain when bound to a particular nucleotide. Langevin molecular dynamics were then performed, which enabled the observation and characterization of three types of binding between the SBD and NBD. Accordingly, Type I binding constituted an opening of the SBD into its component α and β subunits, which then bound to the I and II subdomain lobes of the NBD, respectively. Type II binding entailed spatial rotation of the former α and β subunit orientation by 180º,

with concomitant placement of the β subunit over the top of the lobe interstice and binding of the α unit (a long helix bundle) to the opposite face of the domain. Lastly, Closed-binding involved the binding of the SBD in its closed form to the NBD. Although closed binding occurred with the greatest probability across all states, the state of the NBD was shown to influence the opening frequency of the SBD and its binding behavior to the NBD. In particular, the ATP-bound state was demonstrated as most flexible—existing for extended periods with an open, unbound SBD before subsequent binding to the NBD. This agrees well with the experimental behavior of the protein, which displays low affinity for substrate in its ATP-bound form; indeed, an open SBD would encourage substrate dissociation. In contrast, all SBD-open conformations for the ADP-bound species were captured by the NBD via binding. Moreover, the ATP-bound state exhibited the highest occurrence of open binding—in the case of Type I binding, this resulted in an interdomain conformation that reproduced experimentally obtained structures of the ATP-bound chaperone. Thus, simulation enabled a detailed description of the interdomain communication between subdomains, describing subdomain behavior as the product of a structure-function relationship reliant on the state of the ATP-ase.

I studied αA-crystallin (from *B. Taurus*, pdb 3L1E) employing the AMBER all-atom model and force field. In this case, the effect of amino-acid racemization—that is, the substitution of an L-amino acid for its D-analogue—was investigated with respect to the mechanical and structural properties of the protein. Racemization is a common phenomenon that occurs as an adverse Post-Translational Modification (PMT) in the lens of the eye, and has been implicated in the formation of cataract and in lens hardening (presbyopia). Steered Molecular Dynamics (SMD) was the primary tool utilized, which introduces an external force in a molecular dynamics simulation of the system. In this case, the small heat-shock (sHsp) protein αA-crystallin was stretched along its long axis; three systems bearing a single D-amino acid substitution and a control system of the native protein were simulated. An Essential Dynamics-like method was employed as the analysis methodology—instead of the manipulating the coordinates of the protein, 'structural links' were utilized; these describe interactions in the protein between C^{α}-C^{α}

residue pairs, typically reflecting the presence of elements of secondary structure such as β-sheet, α-helix, or other chain-chain contacts. Eigenvectors therefore described trends correlating the pulling force to the existence or lack thereof of structural links. The nature and extent of the mechanical and structural effects rendered upon D-isomerization depended on the location of the point-substitution. A dominant feature was a change in the stiffness of structural elements within the protein. Increased stiffness was associated with a re-organization of the structural links relative to the native system. Decreased stiffness was correlated to a decoupling of the force response from the typical structural-link pattern of αA-crystallin. As a corollary to either case, the unfolding pathway was modified. Clearly, the unfolding facility/stiffness and the presence of non-native unfolding quasi-intermediates present the potential for unwanted protein-protein interactions and thus agglomeration. For αA-crystallin, the structure-function relationship relating amino-acid chirality to the protein's mechanical and structural properties is indeed highly sensitive and residue-specific.

I designed A sericin-based biopolymer and tested it mechanically by SMD simulation, using the AMBER all-atom model and force field. Sericin is obtained as a byproduct of the silk processing industry. Unlike silk, it assumes a non-uniform structure dominated by random coil; incorporation of sericin into synthetic materials has endowed them with biodegradability and enhanced mechanical and/or other tunable features such as water absorption, efficacy as an antioxidant, and UV-protection. Nonetheless, pure-sericin materials have suffered from their apparent brittleness. On the other hand, the features of spider silk have been ascribed to the organization of structural motifs in the silk fiber; in particular, the elasticity of *flagelliform* silk has been attributed to an elastin-like motif, whose spiral-like structure resembles that of a spring. In the present study I propose a fully silk-based biopolymer comprised of sericin and elastin-like monomers, whose mechanical properties were tested by stretching through SMD; a sericin-only biopolymer was likewise generated and served as a control. The elasticity of each system was investigated with a secondary stretching experiment, wherein the stretched structures were allowed to relax by MD and were subsequently stretched again. Both systems initially responded to the application of force by unraveling the random-coil of sericin, which occurred via a low-force operation. As a monomer

cross-link region was encroached, the sericin-only system manifested a steep increase in force, and thus reflected entry into its bond-failure window. In contrast, the dual sericin and elastin-like motif system began to unfold its elastin-like motif, enabling elongation to continue. Secondary stretching presented the sericin-only system as one of diminished elasticity, and so suffering from permanent deformation. Conversely, the dual biopolymer exhibited almost ideal elasticity, reproducing its initial stretch behavior. Thus, based on the structure function relationships known to silk, a novel resilient and sericin-based biopolymer was engineered, expressing the desired properties of increased strength and elasticity.

In the present study I have therefore examined the structure function relationships active in three proteins that span the structure-function spectrum, concomitantly shedding light on the interplay between the concepts of structure and function. From Hsp70, αA-crystallin, to a sericin-based biopolymer, these relationships serve as elements of a molecular toolbox—employed by biological systems or harnessed accordingly for their engineering potential.