Thorsten Michael Beitlich
Dr. med. Dipl.-Phys.

Fluorescence-Spectroscopic Investigation of Folding and Unfolding of Cytidine Monophosphate Kinase

Promotionsfach: Biochemie
Doktorvater: Professor Dr. med. Heiner Schirmer

Many proteins fold via partially folded protein species. Often these intermediates have molten globule (MG) conformation with significant amounts of secondary structures but no global tertiary structure interaction. Since these species are prone to aggregation they are likely to play an important role in numerous pathologies and human diseases. Knowledge about folding pathways and the characteristics of the involved intermediate structures is necessary to understand and control aggregation phenomena.
CMP kinase from *E. coli* (CMPK) is a protein of the large family of nucleoside monophosphate kinases which play an essential role in the phosphorylation process of nucleotides. Most of these members have a size of 20-25 kDa and share high structural and sequential similarity as well as the α/β fold, a very frequently found topology. I have addressed the question if folding in specific protein classes, e.g. proteins with the α/β fold, or subgroups, e.g. NMP kinases, share similar mechanisms in their folding pathways. The goal was to identify hierarchies or patterns within a protein class in order to suggest folding mechanisms based on secondary and tertiary structure information. I analyzed the refolding transition of cytidine monophosphate kinase (CMPK) from *E. coli* and compared the results to similar proteins like adenylate kinase from *E. coli*, UMP/CMP kinase from *Dictyostelium discoideum*, apoflavodoxin from *Azotobacter vinelandii* and Che Y from *E. coli*.
In this work I have created a new detection device by combining a BioLogic Stopped-flow setup with an Andor Newton charge-coupled device (CCD) system. After optimization of setup conditions and designing a new specialized adapter system, it was possible to achieve sufficiently high signal amplitude and a good signal-to-noise ratio with standard protein concentrations (4-5 μM in our specific case) to allow efficient and significant data analysis The advantage of the new setup is based on its capability to record multiple datasets at the same time, thereby significantly increasing the information-content of a single experiment and as a consequence reduce the amount of protein needed for analysis up to one order of magnitude. Using the spectral information collected with the CCD, it is now possible to distinguish optical shifts - a feature inherent to the frequently used fluorescence of tryptophan residues - from intensity changes during stopped-flow experiments and thereby detect changes that would otherwise remain unnoticed.
Besides studying these technical aspects, the folding of CMPK from *E. coli* was analyzed with a set of fluorescence techniques including FRET measurements, extensively using the new detection system. I found that folding of CMPK is organized in multiple steps. Within the first milliseconds the protein undergoes an initial collapse that cannot be resolved with the stopped flow system. This process leads to formation of significant amount (60 %) of secondary structures, most likely containing a stable nucleus around position 197. This initial folding intermediate is probably offpathway as judged from chevron plot analysis. From there the protein passes through a second intermediate state with an apparent rate constant of 3 s^{-1} and formation of additional secondary structures. Structural rearrangements can be observed for the NMP-domain around position 88 as well as for flexible peripheral regions like the last α-helix around position 208. Folding is finalized by a very slow transition including proline isomerization from *trans*-Pro124 to *cis*-Pro124 and formation of additional secondary structures with rate constants in the range of 0.006 s^{-1}, leading to the folded conformation.

This process is associated with major structural rearrangements of the protein structure as judged from FRET data, especially regarding the surrounding of Pro124 and flexible regions like the NMP domain. Double-jump experiments and data from labeled CMPK mutants have revealed an additional intermediate folding phase with rate constants in the range of 0.2 s^{-1} that most likely originates from folding intermediates with *cis*-Pro124 conformation that comprise at most 10 % of the unfolded CMPK molecules.
The unfolding process on the other hand is mainly described by a single transition with a rate constant of 0.01 s^{-1}, that includes dissolving of secondary structure elements, as well as a minor and spectroscopically silent transition in the range of 0.15 s^{-1} as judged from double jump experiments. The main unfolding rate of CMPK is two orders of magnitude slower than unfolding of other NMP kinases like AMPK and UMPK. This is probably due to the combination of two unique features of CMPK. Firstly, CMPK contains a specific insert of 40 amino acids that significantly increases the interface between NMP domain and CORE domain by a factor of two and acts as a stabilizing component for the folded protein conformation. Secondly, the single *cis*-proline Pro124 is located in a crucial hinge region between the CORE and the NMP domain. Isomerization of Pro124 should therefore lead to a displacement of the two domains and disrupt the interdomain interface. These two features show how minor adjustments in topology and primary structure of a protein can influence the energy landscape of folding within a protein family of similar topology. These aspects must be considered when studying protein misfolding and folding in pathology and pharmacology.