Summary

This study was designed to explore the feasibility of using CysK as a handle or facilitator protein for expression and purification of therapeutic proteins in heterologous expression systems. Learning from our experiences of working with CysK, we set out to design and create a CysK based "minimal protein-protein interaction" platform for expression of small and medium sized therapeutic proteins. CysK expresses well and has good solubility properties. The active site of CysK is a large tunnel which is also the binding cleft for last seven to eight amino acids of C-terminal of protein partners that have CysK binding motifs. Except the last invariant Isoleucine (ILE), the sequence pattern of binding motif is diffusive and could not be deciphered from sequence alignments. Our idea was to create a CysK and a C-terminal tag (C-tag) that would bind to CysK with high-affinity when is expressed as a fused C-tag of any protein or polypeptide sequence. The ultimate aim is to co-express target proteins fused with C-tag with CysK and purify the target protein in complex with CysK. We expect that expressing target proteins in complex with CysK would greatly enhance the solubility and allow us to purify through affinity step. In the absence of any exploratory studies, we planned this study to explore molecular features of CysK with respect to binding C-tags fused with different proteins.

As a first step, we framed the objective to gather experimental evidences that CysK is able to bind multiple proteins that have no structural and functional similarities. The prime idea was identifying natural CysK binding partners and decode "CysK binding" features. In this endeavor, we have created computational tools for identifying natural CysK binders from NCBI data base and validated all our predictions by solution and structural methods. As a case study, we demonstrated that CysK forms an enzyme•transcription factor complex with Rrf2 transcription factor from *Planctomyces limnophilus*. Computational tools created in this study will be used to further explore the protein-protein interaction space of CysK and it may be also used for identifying interaction space of other systems. In addition, the identification and validation of six new CysK binders is likely to benefit researchers in this field.

Next, we designed a different type of experimental step for testing CysK's ability to recognize "non-functional" protein sequences fused with C-tag (CysK binding motif). This attempt was to integrate our learnings into designing a robust protein-protein

interaction system, necessary to capture the heterologous proteins that are difficult to be expressed. The idea was to check if engineered CysK is able to bind a variety of "non-functional" polypeptides/proteins fused with C-tag. Often, heterologously expressed proteins do not fold to their native form unlike natural CysK binding partners. Therefore, it would be interesting to examine whether CysK may be able bind unfolded or misfolded polypeptides with the C-tag. While searching for a meaningful experimental system to test our hypothesis, we settled at an idea to link this goal with co-translational folding, a fundamental protein science problem. Structural biologists attempt to study and characterize partially folded or non-native structures of partially synthesized nascent polypeptide chains, still attached to the ribosomes. We hypothesized that if CysK can serve as a solid support, similar to ribosome, to which N-terminal truncates of proteins can be attached, we may undertake X-ray crystallography approach for characterizing the structural features of early folding intermediates. We demonstrated here that an engineered CysK, DMCysK, and a CysK binding high affinity tag, serve as protein-protein interaction partners. We demonstrate that DMCysK is able to bind to a variety of "non-functional, N-terminal truncates of RNaseA and Lysozyme with high affinity. Further, we were also able to crystallize and resolve crystal structures of multiple DMCysK•Truncate complexes. These results suggest that the engineered "minimal interaction CysK system" is suitable for using as the platform for co expression of heterologous proteins with CysK. These results also convince us that the C-tag when fused with any target sequence through glycine linker retains its high affinity towards CysK active site. The importance of this part of study relies on the feasibility of using CysK for studying protein folding. Lack of electron densities for the N-terminal truncates of target sequences implies that these sequence may have propensity to alternate between different confirmations and provides the first framework as well as template for studying structural features of early folding intermediates.

Finally, we execute design concepts and ideas learned from the first two objectives to engineer a CysK based "minimal interaction" protein-protein interaction system. We applied and implemented these ideas to create a duet expression system in which the engineered DMCysK and a C-tag were cloned at two independent expression sites. We identified expression conditions for optimizing protein-protein interactions between DMCysK and C-tagged IL-2. DMCysK expresses well under all conditions, but IL-2 expression was not significant. We systematically varied physical and induction

parameters, and showed some improvement in soluble IL-2 expression. This shows that this minimal interaction system can be further developed. In the future, we plan to find ways to scale up the expression of hIL-2 fused with tag so that DMCysK would be able to trap the hIL-2 through the C-tag and partition it into soluble fraction. The importance of this approach can be understood from the simplicity of design. We have applied principles of protein-protein engineering to engineer CysK based minimal protein-protein interaction system. We succeeded in creating a DMCysk and a C-tag "minimal interaction system" which is general in nature and can be used to bind "target protein" which will be expressed with a short tag or other purposes such as protein folding. Therefore, results presented here not only provide a general framework for optimizing this "minimal interaction system" for expression and purification of small and medium sized therapeutic proteins, it has also provided the first workable template CysK based interaction system for studying both fundamental and applied problems in biology.