Published in final edited form as:

Protein Expr Purif. 2010 December ; 74(2): 223–230. doi:10.1016/j.pep.2010.06.018.

Over-expression and purification strategies for recombinant multi-protein oligomers: A case study of *Mycobacterium tuberculosis* σ /anti- σ factor protein complexes

Krishan Gopal Thakur¹, Ravi Kumar Jaiswal¹, Jinal K. Shukla, T. Praveena, and B. Gopal^{*}
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Abstract

The function of a protein in a cell often involves coordinated interactions with one or several regulatory partners. It is thus imperative to characterize a protein both in isolation as well as in the context of its complex with an interacting partner. High resolution structural information determined by X-Ray crystallography and Nuclear Magnetic Resonance offer the best route to characterize protein complexes. These techniques, however, require highly purified and homogenous protein samples at high concentration. This requirement often presents a major hurdle for structural studies. Here we present a strategy based on co-expression and co-purification to obtain recombinant multi-protein complexes in the quantity and concentration range that can enable hitherto intractable structural projects. This feasibility of this strategy was examined using the σ factor/anti- σ factor protein complexes from *Mycobacterium tuberculosis*. The approach was successful across a wide range of σ factors and their cognate interacting partners. It thus appears likely that the analysis of these complexes based on variations in expression constructs and procedures for the purification and characterization of these recombinant protein samples would be widely applicable for other multi-protein systems.

Introduction

Structural studies on multi-protein complexes have often entailed elaborate procedures to obtain protein samples that are suitable for subsequent *in vitro* characterization. *In vitro* reconstitution experiments, often involving the addition of different protein constituents at stoichiometric concentrations of a protein complex, were noted to be partially successful primarily as individual components of a protein complex are often unfolded or poorly soluble. Co-expression and co-purification strategies were thus pursued as a likely avenue to obtain multi-protein complexes on a large scale. A successful demonstration of the feasibility of this procedure was based on the biophysical characterization and crystal structure of the *M. tuberculosis* PE/PPE protein complex [1]. This example prompted an evaluation of co-expression as a potential route to incorporate high-throughput expression and purification procedures optimized for single, globular proteins to multi-protein complexes [2, 3]. A straightforward application of this procedure has, however, proved more demanding due to the variability imposed by the size of a recombinant protein construct that can substantially influence expression, solubility and the ability of the recombinant protein complex to form diffraction quality crystals.

^{*} Corresponding Author Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012 Phone : (+91) 80 2293 3219 Fax: (+91) 80 23600535 bgopal@mbu.iisc.ernet.in.

¹These authors contributed equally to the work reported in this manuscript

Most cellular and developmental processes involve the coordinated action of different proteins and multi-protein complexes. In the case of prokaryotic transcription, complexes formed by the transcription initiation proteins, called the sigma (σ) factors, dictate the transcription profile of a bacterial cell. The holo-enzyme formed by the complex of a σ factor with the RNA polymerase is transcriptionally active with the ability to recognize promoter elements in the DNA. There are several σ factors in a bacterial cell and the relative concentration of these σ factors is believed to substantially influence the expression profile. In general, σ factors exist in two forms. The inactive form of a σ factor is achieved either by conformational changes as in the case of the primary σ factor, σ^A , or by interactions with an inhibitory protein in the case of the Extracytoplasmic Function (ECF) σ factors. The active form, on the other hand, is the free σ factor that can bind the RNA polymerase to form the holo-enzyme complex. σ factors are released from the inactive complex by cellular or environmental cues or by the selective proteolysis of the anti- σ factor [4]. The structural characterization of a σ factor/anti- σ factor complex is thus crucial to understand the features that govern the relative cellular concentration of a given σ factor in a bacterial cell under specific environmental conditions.

This report describes observations on *M. tuberculosis* σ factor/anti- σ factor complexes. The experimental strategy adopted to obtain these multi-protein complexes is a variation of the co-expression and co-purification strategy that has been demonstrated in the case of the PE-PPE protein complex [1]. The experimental variations in the co-expression of σ factor/anti- σ factor complexes incorporate the complexity due to the multi-domain nature of the σ factors as well as the membrane bound anti- σ factor proteins. Here we describe the experimental design and biochemical characterization of four σ factor/anti- σ factor complexes. While it may yet not be possible to generalize this procedure, these studies nevertheless demonstrate the applicability of this procedure across a wide range of σ factors with poor sequence similarity. The methods adopted for this model system are thus likely to be adaptable to other multi-protein complexes using freely available bioinformatics software and standard *E. coli* cloning vectors.

Materials and methods

Design of expression constructs

The length of the expression constructs containing the recombinant proteins were varied for both the σ factor as well the anti- σ factor proteins. In a few cases, the anti- σ factors are associated with the cell membrane. These proteins thus have a three domain composition- a cytosolic domain that binds a σ factor, a transmembrane segment that anchors the protein to the membrane and the extracytoplasmic domain that senses an environmental signal. Since the cytosolic region alone interacts with a σ factor, only the cytosolic domain of a membrane-associated anti- σ factor was examined. The expression constructs were designed based on the prediction of cytosolic and transmembrane regions in the membrane proteins. The transmembrane segments were predicted using TMPRED [5]. The length of the recombinant protein was further optimized based on their secondary structure. This analysis was performed using PSIPRED [6].

Co-expression and co-purification strategy for multi-protein complexes

The pETDuet-1 vector (Novagen, Inc.) having two multiple cloning sites (MCS) was used for cloning and co-expression of σ factor/anti- σ factor proteins. In this expression vector, the gene inserted in MCS-I yielded a recombinant protein with a poly-histidine tag at the N-terminus. The other recombinant protein corresponding to the gene located in MCS-II did not have a fused affinity tag. Another hybrid cloning vector was used incorporating features of pET-22b (yields a recombinant protein with a poly-histidine tag at the C-terminus). In this

variant, the pET-22b vector was modified having two MCSs with independent promoters and ribosomal binding sites. This modified vector led to a recombinant protein that had no tag associated with the protein corresponding to the MCS-1 insert whereas the protein corresponding to the MCS-2 insert had a poly-histidine tag at the C-terminal end. Co-transformation experiments were performed by transforming the two vectors with two different antibiotic selection markers in *E. coli* expression strains.

Fusion protein as an alternative to a multi-protein complex

A flexible linker comprising of repeating units of the tripeptide Threonine-Glycine-Serine (TGS) was incorporated using a 66-base primer
5'ACGCATATGAGACCCGGTGGAGCCTGTGGACCCTGTGGACCCTGTGCGCATAGT
CACCTGCCGCAAC 3'. This strategy was based on the structure of the RAD51-BRCA2 complex [7]. The example reported in this manuscript is based on a (TGS)₃ linker in the case of the σ^D /RsdA complex. The steps involved in the cloning strategy of the σ^D -RsdA fusion construct are shown in supplementary figure 1.

Purification of recombinant proteins

Briefly, cells transformed with the expression vectors were grown in LB broth with antibiotics (100 μ g/ml ampicillin and 30 μ g/ml chloramphenicol) to an A_{600 nm} of 0.4–0.6. The cells were induced with 0.3 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Subsequently, the growth temperature was lowered to 17 °C and cells were grown for further 12–18 hrs before they were spun down. The cells were lysed by sonication in lysis buffer (20 mM Tris-HCl pH 7.5 containing 250 mM NaCl). The cell free lysate was incubated with Co²⁺-NTA or Ni²⁺-NTA beads and were eluted by a gradient of elution buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 200 mM imidazole). The recombinant proteins were further purified by size exclusion chromatography using a Superdex S-200 column (GE Healthcare) after the affinity chromatography purification step. The molecular weights of the recombinant proteins were verified by mass spectrometry both on a MALDI-TOF as well as LC-ESI-MS (Bruker Daltonics, Inc.) mass spectrometers. Samples at various stages of purification were analyzed on a 12% SDS-PAGE.

Thermostability of the σ factor/anti- σ factor complexes

The thermostability of the σ factor/anti- σ factor complex was monitored by CD spectroscopy. The ellipticity measurements were recorded at 222 nm on a JASCO-J715 instrument. All spectra were recorded in 5 mM sodium phosphate buffer adjusted to pH 7.5, using a 1 mm path-length cuvette. The protein concentrations were maintained in the range of 5–15 μ M for the spectropolarimetry experiments. The temperature was varied from 10–90 °C at the rate of 1 °C per minute.

Analytical size exclusion chromatography

For the size exclusion chromatography experiments, ca 0.2–0.3 mg protein samples were passed through a Superdex 75 10/300 GL column (column volume 24 ml) equilibrated in 20 mM Tris-HCl (pH 7.5), 200 mM NaCl at 6 °C using an AKTA FPLC system (GE Healthcare). The elution profile was monitored by measuring the absorbance at 280 nm. The flow rate was maintained at 0.3 ml/min.

LC-ESI-MS analysis

LC-ESI-MS (Bruker Daltonics, Inc.) was carried out using a C8-analytical column (Agilent Eclipse Plus; 4.6 $\times$ 150 mm). The mass differences obtained for the protein complexes were in the error range of $\pm$ 2 Da.

Dynamic Light Scattering (DLS) experiments

DLS experiments were performed on a Viscotek 802 DLS instrument in 12 μ l volume quartz cuvette at 22 °C. The protein concentrations for the scattering experiments were in the range of about ~1 mg/ml and data were analyzed using the software (Omnisize 3.0) associated with the Viscotek instrument. Proteins were dialyzed in 20 mM Tris-HCl, pH 7.5 and 200 mM NaCl for these experiments.

Results

Operon analysis

Classical σ factor/anti- σ factor pairs can often be recognized based on the location of their genes. The genes corresponding to both a σ factor as well as an anti- σ factor lie in the same operon. Furthermore, the gene corresponding to a σ factor is followed by the anti- σ factor in a majority of cases. A search for the locations of genes corresponding to a σ factor and its immediate neighbors is thus a quick test to identify potential anti- σ factors (Figure 1). In *M. tuberculosis*, several σ factor/anti- σ factor pairs can be identified in this way. These include σ^E /Rv1222, σ^H /RshA, σ^L /RslA and σ^F /RsbW. The interactions between these σ factor/anti- σ factor pairs have been experimentally verified in a few cases (reviewed in [8]). Here we report the identification of σ factor/anti- σ factor pair, σ^D /RsdA, based on the location of genes in an operon. The experimental data presented here also demonstrate that the σ^D /RsdA and σ^K /RskA protein pairs form compact complexes *in vitro*. We note that the functional link between σ factor/anti- σ factor pairs, as annotated in the STRING database, is conserved across actinobacteria, in particular within the mycobacterial species [9].

Expression of soluble multi-protein complexes and fusion proteins

An anti- σ factor serves to sequester a σ factor in an inactive conformation. The variations seen in the interactions between these two proteins are consistent with the low sequence similarity between σ factors or anti- σ factors. The σ factor/anti- σ factor complexes can thus result in the localization of a σ factor to a membrane or in some cases lead to the occlusion of RNAP and/or DNA promoter binding sites. It should be noted in this context that although many σ factor/anti- σ factor pairs have been predicted based on a bio-informatics approach such as the one described earlier, very few of these have been experimentally confirmed. Two strategies of protein expression and purification were adapted to solubilize the *M. tuberculosis* σ factor/anti- σ factor pairs. In one approach, a single vector with two multiple cloning sites (pETDuet-1 vector) incorporated both interacting partners whereas the other strategy involved the co-transformation of two vectors having different antibiotic resistance markers. A variant to the pETDuet-1 strategy involved using a hybrid vector that had the poly-histidine tag located at the C-terminus of one of the proteins in the complex. In addition to these variants, a fusion protein incorporating the two interacting partners attached by a flexible linker was also examined (Figure 2). We note that σ factors when expressed alone were either insoluble as in the case of σ^D and σ^K or were very feebly soluble (σ^L , σ^F). On the other hand anti- σ factors alone (RsbW, RslA and RskA) are soluble except anti- σ^D (RsdA). Surprisingly, although intact σ factors were not soluble or poorly soluble, the domains of these proteins when cloned separately (domain 2 or domain 4) are soluble (data not shown). A list of the expression constructs used in this study is compiled in Table1. The co-transformation strategy was examined in the case of the σ^F /RsbW and σ^D /RsdA complexes. This approach yielded contrasting results in these two cases. In the case of the σ^F /RsbW complex, σ^F was cloned in the pET-15b vector without an affinity tag and RsbW was cloned in the pET-28a vector that provided a recombinant protein with a poly-histidine tag at the C-terminus (Figure 3G). The complex could be obtained albeit with sub-stoichiometric amounts of σ^F . In the case of the σ^D /RsdA complex, however, the yield of the resultant complex was very poor. The approach using the pETDuet-1 expression vector

was uniformly successful- protein complexes could easily be obtained in a scale suitable for extensive crystallization trials. This was used to purify the σ^D /RsdA, σ^L /RslA, σ^K /RskA and the σ^F /RsbW complexes (Figure 3).

Stoichiometric analysis of σ factor/anti- σ factor complexes

We note that the presence of the poly-histidine tag substantially influences the stoichiometry of a protein complex (i.e. one of the interacting partner is present in excess). This was most prominent in the case of the σ^L /RslA and the σ^F /RsbW complexes (Figures 3A, B, F and G). This imbalance in the stoichiometry of the protein complex resulted in heterogeneity in the protein sample. Additional steps of purification such as gel filtration and/or ion-exchange chromatography were needed to get homogenous samples. However, this is not a general feature of σ factor/anti- σ factor pairs. For example, in the case of the σ^D /RsdA complex both proteins are insoluble when over-expressed separately in *E. coli*. The complex of the recombinant proteins is, however, quite soluble and we note that the presence of a poly-histidine tag in either protein (σ^D or RsdA) did not affect the stoichiometry of the purified complex (Figures 3C and 3D).

The apparent molecular weight and structural integrity of σ factor/anti- σ factor complexes were assessed by analytical size exclusion chromatography and Dynamic Light Scattering (DLS) experiments (Figure 4 panel ii and iii). Symmetric elution profiles are characteristic of homogeneous proteins, whereas asymmetric profiles reflect heterogeneous samples (Figure 4 panel ii). Based on the size exclusion profiles, we conclude that the σ^L /RslA, σ^D /RsdA and σ^K /RskA form complexes with 1:1 stoichiometry. The σ^F /RsbW complex, on the other hand, appeared to have a 1:2 stoichiometry. While the σ^D /RsdA and the σ^K /RskA complexes had a mono modal elution profile, the σ^L /RslA and σ^F /RsbW complexes elute with a bimodal profile in a size exclusion chromatography experiment. In the case of the σ^L /RslA complex, the major peak corresponds to the 1:1 complex while a substantial proportion of the protein elutes as a higher order oligomer. While aggregation was more prominent in the σ^F /RsbW complex the other, minor peak corresponds to a complex with 1:2 (σ^F :RsbW) stoichiometry. The co-transformation strategy in the case of this complex led to additional heterogeneity. In this case, the σ^F /RsbW complex showed three elution peaks. Two elution fractions were same as in the case of σ^F /RsbW complex purified from pETDuet-1 vector over-expression. The additional peak corresponds to that of the dimer of RsbW alone (Supplementary Figure 2A). Indeed the polydispersity of this sample is higher in a DLS experiment when compared to the complex purified using the pETDuet-1 vector (Supplementary Figure 2B and Figure 4D panel iii). It is likely that the presence of the poly-histidine tag at the C-terminus of RsbW leads to a higher yield of this protein after the affinity chromatography step. In the case of the σ^D /RsdA fusion protein, the elution profile also showed two additional elution peaks compared to the σ^D /RsdA co-expressed in pETDuet-1. These peaks suggest the formation of higher oligomeric species in the fusion protein (Figures 4B and 4E). It is likely that the formation of higher oligomers in the σ^D /RsdA fusion protein could be a consequence of domain swapping. Co-expression and size exclusion experiments were also performed to identify σ factor domain(s) that specifically interact with an anti- σ factor. These experiments were often very informative. For example in the case of the interactions between σ^L and RslA, the elution profile of the full length σ^L /RslA complex and the σ^L_4 /RslA complex (containing domain 4 alone) are similar. The σ^L_2 domain, however, did not show any interaction with RslA in these experiments (Figure 5A and 5B).

The solution behavior as well as the stoichiometry of the σ factor/anti- σ factor proteins were also examined by DLS [10, 11]. DLS measures the translational diffusion coefficients (D_T) of proteins by monitoring the scattered light produced when a monochromatic light is directed at a protein in solution. In this experiment, the hydrodynamic radius (R_h) of the

protein was calculated using the Stokes–Einstein equation: $R_h = k_B T / 6\pi\eta D_t$. The data were analyzed to obtain the hydrodynamic radius (R_h) value for the protein complexes. The reported R_h values correspond to the dominant peak. The DLS data suggest that σ factor/anti- σ factor complexes exist as homogenous species in solution (Figure 4, panel iii). The molecular weight calculated from different σ factor/anti- σ factor complexes based on the R_h values are similar to the molecular weight derived from the size exclusion chromatography profiles. The higher order oligomers or aggregates noted in the size exclusion profiles also corroborate well with the DLS observations. A higher degree of poly-dispersity was observed in the σ^F /RsbW and σ^L /RslA complexes as well as in the σ^D -RsdA fusion protein samples.

Thermal stability of σ factor/anti- σ complexes

The σ factor/anti- σ factor complexes purified in the course of this study were fairly stable as judged by their melting temperatures (T_m) which range from 55-73 °C (Figure 4, panel i). Although the data presented here are rather preliminary, a cursory examination of the thermal denaturation data for all the σ factor/anti- σ factor complexes suggest that these multi-domain proteins unfold together. This observation is consistent with the finding that either one or both of the interacting proteins are often insoluble in isolation.

Discussion

The wealth of structural information that can be obtained from multi-protein complexes can often be correlated with the difficulty in obtaining these complexes in a pure, homogenous form at high concentration. Protein expression for structural and biophysical studies is best achieved using *E. coli* expression hosts and compatible vectors [12, 13]. These also provide a level of flexibility in terms of the level of expression of the recombinant protein by adjusting the concentration of the inducer. Earlier experimental design for structural studies on protein-protein complexes involved reconstitution of interacting partners where one or more partners were insoluble and hence required refolding [14, 15]. This approach in recent years has been all but replaced by co-expression strategies in *E. coli* that allow interacting proteins to fold together resulting in better yield of recombinant protein complexes [1, 16]. There are three main strategies for heterologous production of binary protein complexes. The first involves cloning of target genes in a bi-cistronic vector, where genes are present in the same order as in an operon and the expression of the recombinant protein is controlled by a single promoter. A drawback of this approach is that the expression of second gene is usually lower, resulting in a sub-stoichiometric arrangement of the resultant complex [10, 17]. The second strategy involves co-transformation wherein the interacting partners are cloned in two different plasmids with different selection markers [18, 19]. Although this approach circumvents the problem of differential expression of the genes, it is open to the possibility that one vector dominates over the other depending on its copy number. The third possible route for the production of interacting proteins involves expression of heterologous proteins under the control of two different promoters of same vector. It has been suggested that this approach results in a four to nine-fold increase in the yield of protein-protein complexes compared to the other two strategies [20, 21].

The σ factors examined in this study were either feebly soluble (< 0.3-0.5 mg/ml) or insoluble when expressed alone. In case of σ^D , even the GST fusion protein was insoluble. In the previous studied cases of σ factor/anti- σ factor interactions both the interacting partners could be purified separately and the complexes could be reconstituted *in vitro* [22]. This was not possible in the case of *M. tuberculosis* ECF σ factors. An interesting observation in this context is that a sub-set of σ factors (σ^A , σ^C and σ^J) which do not have a predicted anti- σ factor could be over-expressed in *E. coli* and were soluble (data not shown).

The steps involved in predicting interacting partners to the purification of complexes suitable for crystallization trials or any other biochemical or biophysical studies have been shown schematically in Figure 6. In principle, if the two interacting partners in a σ factor/anti- σ factor complex were to be expressed separately, any of the following outcomes are likely. Either both proteins are soluble or one is soluble and the other less soluble or both interacting partners are insoluble. If a poly-histidine affinity tag is attached to the protein that is more soluble, this results in an increased quantity of this protein in the purified complex, in effect leading to additional heterogeneity in the sample. On the other hand, if the affinity tag is attached to the less soluble or insoluble protein partner, the resultant complex has stoichiometric amounts of the two interacting proteins. This empirical finding also holds true in the case where both the interacting proteins are insoluble when expressed separately. In this case, attaching an affinity tag to either partner results in a stoichiometric complex. It thus appears likely that attaching an affinity tag to a relatively less soluble or insoluble protein in a complex helps in obtaining stoichiometric amounts of proteins in the resultant complex. We were successful in getting soluble protein-protein complexes in all of the four σ factor/anti- σ factor complexes using the pETDuet-1 expression system. The co-transformation strategy, on the other hand, worked in only one of the cases. A potential application of the protein co-expression/co-purification strategies is in the identification of interacting domain(s) between the two protein partners. This was especially fruitful in the case of the σ^L /RslA complex [23].

The fusion protein strategy was successful in obtaining the σ^D /RsdA complex in a soluble form although both the proteins were insoluble when expressed in isolation. However, this construct tended to form higher order oligomers leading to heterogeneity in the protein sample. In the present study, a flexible (TGS)₃ polypeptide linker was examined. Indeed, it is likely that varying the length of the (TGS) linker could substantially alter the solution behavior of the resultant fusion protein. A major limitation, however, is that this strategy can only be applied to those complexes where interacting partners interact in 1:1 stoichiometry.

To summarize, we present a strategy that proceeds from the identification of a potential interacting partner to the production of a recombinant protein complex that is suited for structural studies. The feasibility of this approach was demonstrated in the case of four σ factors from *M. tuberculosis*. This strategy can be applied to any sequenced genome to predict the anti- σ factors of an ECF σ factor and subsequently analyze the protein complex. The co-expression strategy that we present suggests that attaching an affinity tag to a less soluble or insoluble protein component substantially improves the yield and stoichiometry of the purified recombinant protein complex. This observation could easily be adapted for other interacting proteins perhaps influencing the solution properties of the resulting protein complexes or the ability of the complex to form diffraction quality crystals to enable structure determination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- [1]. Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*. 2006; 103:8060–8065. [PubMed: 16690741]
- [2]. Romier C, Ben Jelloul M, Albeck S, Buchwald G, Busso D, Celie PH, Christodoulou E, De Marco V, van Gerwen S, Knipscheer P, Lebbink JH, Notenboom V, Poterszman A, Rochel N, Cohen SX, Unger T, Sussman JL, Moras D, Sixma TK, Perrakis A. Co-expression of protein complexes

- in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Cryst. D*. 2006; 62:1232–1242. [PubMed: 17001100]
- [3]. Graslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, Prilusky J, Sussman JL, Stevens RC, Lesley SA, Wilson IA, Joachimiak A, Collart F, Dementieva I, Donnelly MI, Eschenfeldt WH, Kim Y, Stols L, Wu R, Zhou M, Burley SK, Emtage JS, Sauder JM, Thompson D, Bain K, Luz J, Gheyi T, Zhang F, Atwell S, Almo SC, Bonanno JB, Fiser A, Swaminathan S, Studier FW, Chance MR, Sali A, Acton TB, Xiao R, Zhao L, Ma LC, Hunt JF, Tong L, Cunningham K, Inouye M, Anderson S, Janjua H, Shastry R, Ho CK, Wang D, Wang H, Jiang M, Montelione GT, Stuart DI, Owens RJ, Daenke S, Schutz A, Heinemann U, Yokoyama S, Bussow K, Gunsalus KC. Protein production and purification. *Nat. Methods*. 2008; 5:135–146. [PubMed: 18235434]
  - [4]. Helmann JD. Anti-sigma factors. *Curr. Opin. Microbiol.* 1999; 2:135–141. [PubMed: 10322161]
  - [5]. Hofmann K, Stoffel W. TMBASE - A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler*. 1993; 374:166.
  - [6]. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000; 16:404–405. [PubMed: 10869041]
  - [7]. Pellegrini L, Yu DS, Lo T, Anand S, Lee M, Blundell TL, Venkitaraman AR. Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature*. 2002; 420:287–293. [PubMed: 12442171]
  - [8]. Sachdeva P, Misra R, Tyagi AK, Singh Y. The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS J*. 2010; 277:605–626. [PubMed: 19951358]
  - [9]. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 2009; 37:412–416. [PubMed: 18940858]
  - [10]. Singh VK, Nurmohamed S, Davey SK, Jia Z. Tri-cistronic cloning, overexpression and purification of human Rad9, Rad1, Hus1 protein complex. *Protein Expr. Purif*. 2007; 54:204–211. [PubMed: 17493829]
  - [11]. Skouri M, Delsanti M, Munch JP, Lorber B, Giege R. Dynamic light scattering studies of the aggregation of lysozyme under crystallization conditions. *FEBS Lett*. 1991; 295:84–88. [PubMed: 1765173]
  - [12]. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 1986; 189:113–130. [PubMed: 3537305]
  - [13]. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*. 1990; 185:60–89. [PubMed: 2199796]
  - [14]. Tan S, Hunziker Y, Sargent DF, Richmond TJ. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature*. 1996; 381:127–151. [PubMed: 8610010]
  - [15]. Guise AD, West SM, Chaudhuri JB. Protein folding in vivo and renaturation of recombinant proteins from inclusion bodies. *Mol. Biotechnol*. 1996; 6:53–64. [PubMed: 8887361]
  - [16]. Zeng J, Zhang L, Li Y, Wang Y, Wang M, Duan X, He ZG. Over-producing soluble protein complex and validating protein-protein interaction through a new bacterial co-expression system. *Protein Expr. Purif*. 2010; 69:47–53. [PubMed: 19747546]
  - [17]. Rucker P, Torti FM, Torti SV. Recombinant ferritin: modulation of subunit stoichiometry in bacterial expression systems. *Protein Eng*. 1997; 10:967–973. [PubMed: 9415447]
  - [18]. Dzivenu OK, Park HH, Wu H. General co-expression vectors for the overexpression of heterodimeric protein complexes in *Escherichia coli*. *Protein Expr. Purif*. 2004; 38:1–8. [PubMed: 15477075]
  - [19]. Johnston K, Clements A, Venkataramani RN, Trievel RC, Marmorstein R. Coexpression of proteins in bacteria using T7-based expression plasmids: expression of heteromeric cell-cycle and transcriptional regulatory complexes. *Protein Expr. Purif*. 2000; 20:435–443. [PubMed: 11087683]

- [20]. Kim KJ, Kim HE, Lee KH, Han W, Yi MJ, Jeong J, Oh BH. Two-promoter vector is highly efficient for overproduction of protein complexes. *Protein Sci.* 2004; 13:1698–1703. [PubMed: 15133160]
- [21]. Hunt I. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr. Purif.* 2005; 40:1–22. [PubMed: 15721767]
- [22]. Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, Gross CA, Darst SA. Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol. Cell.* 2003; 11:1067–1078. [PubMed: 12718891]
- [23]. Thakur KG, Praveena T, Gopal B. Structural and biochemical bases for the redox sensitivity of *Mycobacterium tuberculosis* RslA. *J. Mol. Biol.* 2010; 397:1199–1208. [PubMed: 20184899]

Figure 1. Bioinformatics approaches to predict a cognate anti- σ factor

Classical σ factor/anti- σ factor pairs occur as adjacent genes on the same operon. **A.** Arrangement of σ factor/anti- σ factor genes in their respective operons. The intergenic distances are also shown in these cases. **B.** Bioinformatics analyses performed to predict probable σ factor/anti- σ factor pairs. This data was compiled using the STRING database [9] and available literature on these proteins.

1. Co-expression strategies

A. Single vector multiple promoters

B. Co-transformation

2. Fusion protein strategy

Figure 2. Schematic of the co-expression and co-transformation strategies for the expression of σ factor/anti- σ factor complexes in *E. coli*

Among these two strategies, the co-expression strategy was successful in all the cases examined, whereas the co-transformation approach had limited success.

Figure 3. SDS PAGE analysis of purified σ factor/anti- σ factor complexes

Protein complexes were purified either using the co-expression or co-purification strategies. These were analyzed on a 12 % SDS PAGE. **A**, σ^L /RslA (poly-histidine tag at the N-terminus in σ^L); **B**, σ^L /RslA (poly-histidine tag at the N-terminus of RslA); **C**, σ^D /RsdA (poly-histidine tag at the N-terminus in σ^D); **D**, σ^D /RsdA (poly-histidine tag at the C-terminus of RsdA); **E**, σ^K /RskA (poly-histidine tag at the N-terminus in σ^K); **F**, σ^F /RsbW (poly-histidine tag at the N-terminus in σ^F); **G**, σ^F /RsbW (poly-histidine tag at the N-terminus in RsbW); **H**, σ^D -RsdA fusion protein with (TGS)₃ linker (poly-histidine tag at the N-terminus in the fusion protein). The lanes labeled E1, E2, E3 and E4 correspond to the elution fractions of the fusion protein of σ^D and RsdA. In panels A to G, the upper and lower bands correspond to σ factor (arrow) and the anti- σ factor (asterisk) respectively.

Figure 4. Biophysical characterization of the σ factor /anti- σ complexes
 Panels A, B, C, D and E represent σ^L /RslA, σ^D /RsdA, σ^K /RskA, σ^F /RsbW and the σ^D -RsdA fusion proteins. Subpanels (i), (ii) and (iii) represent thermal denaturation (T_m), analytical gel filtration and DLS profiles, respectively. The inset in subpanel (ii) shows the calibration curve for the size exclusion chromatography column with standard molecular weight markers. The blue or/red triangles on the calibration scale correspond to the molecular weight(s) inferred based on the peak(s) in the elution profile. The statistics given in subpanel (iii) correspond to the peak with the highest area under curve in the DLS distribution.

Figure 5. Protein co-expression as a strategy to identify interacting domains

The σ^L /RslA complex was used as a model system. **A.** Analytical size exclusion experiments suggest that σ^L_2 is a monomer whereas the σ^L_4 /RslA and σ^L /RslA complexes exist in two distinct oligomeric states in solution. **B.** A representative 12% SDS PAGE analyses of the σ^L_2 /RslA and σ^L_4 /RslA proteins obtained after the Co^{2+} -NTA affinity chromatography step of purification. Co-expression and co-purification strategies suggest that σ^L_4 interacts with RslA whereas the σ^L_2 /RslA interactions are much weaker.

Figure 6. Schematic of the strategy employed in the identification and purification of protein complexes

This protocol was validated using four *M. tuberculosis* σ factor /anti- σ factor pairs.

Table 1

List of the Expression constructs

S.No.	Vector	MCS 1/MCS 2	Expression host	Solubility (Single protein)	Solubility (Complex)	Mol. wt. (kDa)
1	pETDuet-1	σ^D_{1-212}	<i>E. coli</i> /Rosetta DE3 pLysS	NS	–	22.9
2	pET22b	RsdA ₁₋₆₇	<i>E. coli</i> /Rosetta DE3 pLysS	NS	–	7.5
3	pETDuet-1	σ^D_{1-212} /RsdA ₁₋₆₇	<i>E. coli</i> /Rosetta DE3 pLysS	–	~50 mg/ml	22.9/7.5
4	Modified pET22b	σ^D_{1-212} /RsdA ₁₋₆₇	<i>E. coli</i> /Rosetta DE3 pLysS	–	~50 mg/ml	22.9/7.5
5	pETDuet-1	σ^D_{1-212} -RsdA ₁₋₆₇ fusion protein	<i>E. coli</i> /Rosetta DE3 pLysS	~20 mg/ml	–	33.5
6	pET28a + pETDuet-1	σ^D_{1-212} + RsdA ₁₋₆₇ coexpression	<i>E. coli</i> /Rosetta DE3 pLysS		~0.5 mg/ml	22.9 + 7.5
6	pETDuet-1	σ^L_{1-177}	<i>E. coli</i> /Rosetta DE3	<0.3 mg/ml	–	19.5
7	pETDuet-1	RslA ₁₋₁₀₈	<i>E. coli</i> /Rosetta DE3	~25 mg/ml	–	11.3
8	pETDuet-1	σ^L_{1-177} /RslA ₁₋₁₀₈	<i>E. coli</i> /Rosetta DE3	–	~7 mg/ml	19.5/11.3
9	pETDuet-1	RslA ₁₋₁₀₈ / σ^L_{1-177}	<i>E. coli</i> /Rosetta DE3	–	~15 mg/ml	11.3/19.5
10	pETDuet-1	$\sigma^L_{2(1-95)}$ /RslA ₁₋₁₀₈	<i>E. coli</i> /Rosetta DE3	~1mg/ml	–	10.5/11.3
11	pETDuet-1	$\sigma^L_{4(100-177)}$ /RslA ₁₋₁₀₈	<i>E. coli</i> /Rosetta DE3	–	~15mg/ml	8.6/11.3
12	pET15b	σ^F_{1-261}	<i>E. coli</i> /Rosetta DE3	<0.5 mg/ml	–	28.7
13	pET15b	RsbW ₁₋₁₄₅	<i>E. coli</i> /Rosetta DE3	~5 mg/ml	–	15.5
14	pETDuet-1	σ^F_{1-261} /RsbW ₁₋₁₄₅	<i>E. coli</i> /Rosetta DE3	–	~4 mg/ml	28.7/15.5
15	pET15b + pET28a	σ^F_{1-261} + RsbW coexpression	<i>E. coli</i> /Rosetta DE3	–	~4 mg/ml	28.7 + 15.5
16	pETDuet-1	σ^K	<i>E. coli</i> /Rosetta DE3	NS	–	21.0
17	pETDuet-1	RskA ₁₋₈₀	<i>E. coli</i> /Rosetta DE3	~10mg/ml	–	8.5
18	pETDuet-1	σ^K /RskA ₁₋₈₀	<i>E. coli</i> BL21DE3	–	~40 mg/ml	21.0/8.5

Footnote: NS = Not soluble. The pET vectors and *E. coli* expression hosts were from Novagen Inc.