Preparation and characterization of bacterial protein complexes for structural analysis

Matte, Allan; Kozlov, Guennadi; Trempe, Jean-François; Currie, Mark A.; Burk, David; Jia, Zongchao; Gehring, Kalle; Ekiel, Irena; Berghuis, Albert M.; Cygler, Miroslaw

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/S1876-1623(08)76001-2
Advances in Protein Chemistry and Structural Biology, 76, pp. 1-42, 2009-11-27

PREPARATION AND CHARACTERIZATION OF BACTERIAL PROTEIN COMPLEXES FOR STRUCTURAL ANALYSIS

By ALLAN MATTE,* GUENNADI KOZLOV,† JEAN-FRANÇOIS TREMPE,† MARK A. CURRIE,‡ DAVID BURK,† ZONGCHAO JIA,‡ KALLE GEHRING,† IRENA EKIEL,*,§ ALBERT M. BERGHUIS,† AND MIROSLAW CYGLER*,†

*Biotechnology Research Institute, National Research Council Canada, Montreal, Quebec, Canada
†Department of Biochemistry, McGill University, Montreal, Quebec, Canada
‡Department of Biochemistry, Queen's University, Kingston, Ontario, Canada
§Department of Chemistry and Biochemistry, Concordia University, Montreal, Quebec, Canada

ADVANCES IN PROTEIN CHEMISTRY, Vol. 76
DOI: 10.1016/S1876-1623(08)76001-2

1876-1623/08 $35.00

Abstract

Bacteria mediate a large variety of biological processes using protein complexes. These complexes range from simple binary heterodimeric enzymes to more complex multi-subunit complexes that can be described as macromolecular machines. A key to understanding how these complexes function is obtaining structural information using methods that include electron microscopy, small-angle X-ray scattering, NMR spectroscopy, and X-ray crystallography. Here we describe a variety of approaches to the expression, purification, and biophysical characterization of bacterial protein complexes as a prerequisite to structural analysis. We also give several examples of the kinds of information these different biophysical approaches can provide and various experimental approaches to obtaining structure information for a given system. Further, we describe several examples of protein complexes where we have obtained structural data that have led to new biological insights.

I. Introduction

Cellular processes involve interactions between multiple proteins. There are many well-documented examples of this in microbial systems, including the co-localization of enzymes associated with sequential steps within an enzymatic pathway, sometimes found as a hetero-oligomeric complex (Klem and Davisson, 1993); the association of proteins to effect cell division (Gamba *et al*., 2009); the assembly of proteins into complexes to create pores or channels in order to import or export molecules through one or more membranes (Collins *et al*., 2007); and the control of gene expression through the modulation of both transcription (Rutherford *et al*., 2009) and translation (Yu *et al*., 2009). These protein–protein interactions span a wide range of timescales and affinities adapted to their specific biological roles.

We can differentiate between relatively weak and transient protein–protein interactions, which exist as part of many cellular functions, versus stronger interactions resulting in longer lived and more stable protein complexes, in which proteins come together to effect function as an unit. On one end of this spectrum are hetero-oligomeric enzymes that contain two or more subunits usually tightly associated together,

while on the other end are proteins that exist independently and associate into meta-stable complexes in order to perform a specific task. Structural analysis of protein complexes, using methods such as X-ray crystallography and NMR spectroscopy, offers a way to visualize at the molecular level protein–protein interfaces and analyze their properties. Many important questions in this area need to be addressed. What dictates the specificity of association? How does one protein interact with multiple partners? What are the structural adjustments at protein interfaces upon association? What features of the surface define interaction "hot spots"? What are the important molecular interactions that govern association and dissociation? What role does conformational change play in interaction? Answers to these questions will yield a new level of understanding at the molecular level and provide a basis to modulate the strength of protein association and allow engineering of new interacting proteins that exert their function *in vivo*.

There is a significant and rapidly growing number of protein complexes deposited in the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). Several derivative databases have been created in recent years to navigate and curate such complexes. Some examples of these databases are PROTCOM (http://www.ces.clemson.edu/compbio/protcom (Kundrotas and Alexov, 2007)), 3D Complex (http://supfam.mrc-lmb.cam.ac.uk/elevy/3dcomplex/Home.cgi (Levy *et al.*, 2006)), and SNAPPI-DB (http://www.compbio.dundee.ac.uk/SNAPPI/snappidb.jsp (Jefferson *et al.*, 2007)). These databases can serve as useful resources for biologists in various disciplines in order to correlate structural information with *in vivo* and *in vitro* data on particular systems of interest. Analysis of such data at the molecular level also helps to shed light on fundamental questions in protein biochemistry, such as the differences between specific and non-specific protein–protein interfaces (Bahadur *et al.*, 2004; Kobe *et al.*, 2008). A number of other databases accumulate experimental information on protein complexes and protein–protein interactions, many of which are listed at http://www.imb-jena.de/jcb/ppi/jcb_ppi_databases.html.

In this chapter we will summarize our experiences regarding the preparation of bacterial protein complexes for structural analysis, with a specific emphasis on protein complexes from the model bacterium *Escherichia coli*. The structural characterization of protein complexes

poses special experimental challenges and requires attention to details that to some extent are unique to each system and in some cases different from those for individual proteins.

II. Protein Complexes of Bacteria

Bacteria contain a wealth of both soluble and membrane-bound protein complexes, as is clearly evident from the enormous literature on this topic and numerous databases devoted to cataloging these molecules. These complexes range from those that can truly be described as "molecular machines" such as the ribosome (Steitz, 2008), the RNA degradosome (Carpousis, 2007), and the protein translocation machinery (Driessen and Nouwen, 2008) to much more modest pairs of relatively small proteins. Systematic large-scale studies of bacterial protein interaction networks at the level of whole proteomes have been initiated using the yeast 2-hybrid (Y2H) method as applied to a subset of the genome of *Helicobacter pylori* (Rain *et al*., 2001), *Synechocystis* sp. PCC6803 (Sato *et al*., 2007), and *Campylobacter jejuni* (Parrish *et al*., 2007). The agreement between different methods applied to the same genome is not particularly good, although in a favorable case of 31 proteins from the *H. pylori* type IV secretion system (Terradot *et al*., 2004) co-purification experiments authenticated 76% of the interactions predicted from large-scale Y2H experiments (Rain *et al*., 2001). Blue native/SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) (Wittig and Schagger, 2008) has been used to identify additional protein complexes from *H. pylori* (Lasserre *et al*., 2006).

As with many aspects relating to bacterial biochemistry and physiology, a wealth of data relating to protein complexes is available for *E. coli*. This data has been generated over many years from a large number of investigator-driven studies, as well as large-scale interactome analysis of the *E. coli* K-12 genome. These large-scale studies include tandem-affinity purification (TAP-tagging (Collins and Choudhary, 2008)) to isolate protein complexes, followed by their identification by mass spectrometry (MS)–MS analysis (Butland *et al*., 2005) and pull downs of His-tagged proteins from the ASKA library (Kitagawa *et al*., 2005), followed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) analysis (Arifuzzaman *et al*., 2006). Although these two techniques

rely on similar principle, the overlap of bona fide protein complexes between these two independent datasets is rather low, making it difficult to use as a source of information for selecting protein complexes for structural studies. Recent studies on the *E. coli* protein interaction network suggest that essential proteins make up a significant "core" of the experimentally identified interactions (Lin *et al*., 2009). It has also been shown that proteins that participate in the same interaction share similar mRNA half-lives, a relationship not previously noted (Janga and Babu, 2009). An alternative approach, the use of blue native/SDS-PAGE, followed by the identification of the protein complex by LC–MS/MS, has been used to identify hetero-oligomeric complexes from *E. coli* (Lasserre *et al*., 2006). Other studies have focused on characterizing another, although more experimentally challenging set of protein complexes, those found in the membrane (Stenberg *et al*., 2005).

Collectively, the experimentally determined protein–protein interaction data for *E. coli* have been captured within various databases, such as the Bacteriome.org portal (http://www.compsysbio.org/bacteriome/ (Su *et al*., 2008)) and the microbial protein interaction database, MPIDB (http://www.jcvi.org/mpidb/about.php (Goll *et al*., 2008), and are available for browsing using a variety of criteria (Su *et al*., 2008). A number of computational approaches have also been applied to predict protein interactions in *E. coli*, including methods relying on gene fusions (Enright *et al*., 1999), functional linkages (Yellaboina *et al*., 2007), and the interacting domain profile pair method (Wojcik *et al*., 2002). The extent to which these predictive methods correlate with the high-throughput data on physical interactions is yet to be established. A promising new approach for the discovery and validation of protein interaction networks in bacteria is the application of one-step inactivation of chromosomal genes in *E. coli* (Datsenko and Wanner, 2000). This method permits the high-throughput construction of double knockouts that allows analysis of synthetic lethality of gene product pairs (Butland *et al*., 2008). It is important to keep in mind that physical and genetic interactions are highly complementary. As was shown in yeast, less than 1% of synthetic lethal genetic interactions were also observed physically (Tong and McIntosh, 2004). A new and rapidly developing branch of bioinformatics deals with integrating and "cleaning" high-throughput experimental data (Beyer *et al*., 2007; Liu *et al*., 2008; Patil and Nakamura, 2005). It is clear from the comparison of results obtained by different methods and

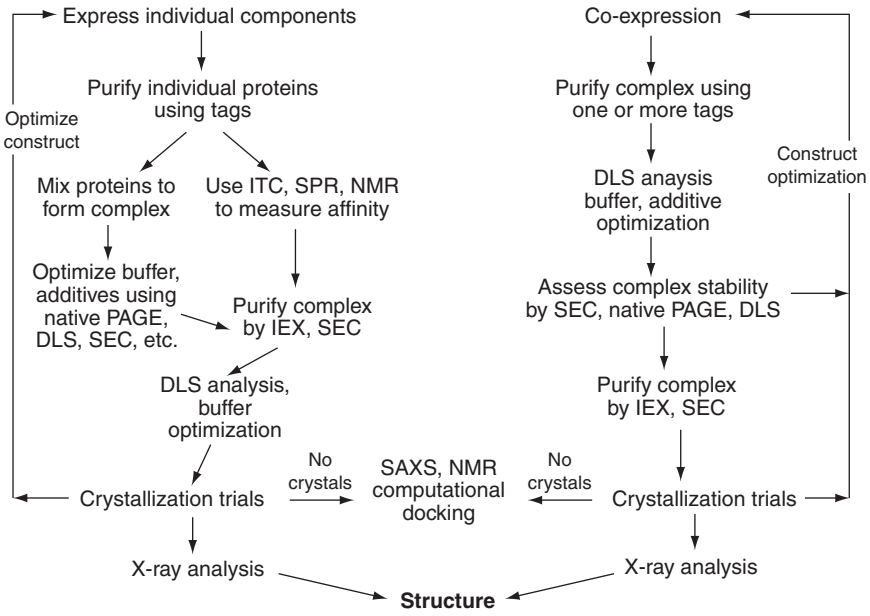

FIG. 1. Summary of experimental strategies used to prepare and characterize a bacterial protein complex.

different investigators that the high-throughput data indicating pair-wise interactions contain a significant fraction of false positives. Therefore, structural investigation of bacterial protein complexes initially requires the verification of protein complex formation *in vitro*. A summary of the experimental strategies to prepare and characterize protein complexes from bacteria is presented in Fig. 1

III. PREPARATION OF BACTERIAL PROTEIN COMPLEXES

A. General Expression and Affinity Purification Strategies

The *in vitro* structure–function studies of any protein complex require its isolation in a form amenable to study. The isolation of protein complexes direct from bacterial cells through successive fractionation is a

possible approach, with advantages including the isolation of relatively stable complexes amenable to electron microscopy (EM) or crystallization and the ability to identify new groups of interacting proteins. However, the use of recombinant methods offers significant advantages, especially for proteins of low abundance and for weakly associated protein complexes, including a higher level of protein expression and the potential to introduce fusion tags, either for detection or for affinity purification. The size and location of tags, however, must be kept in mind, in as much as they could negatively impact protein complex formation or stability.

With regard to the application of strategies for the expression of protein complexes, there are several specific issues that require careful consideration. When the individual protein components are well expressed and soluble, one can reconstitute the complex using individual protein components. There are different strategies that can be adopted, and it may be necessary to work through several of these in order to find the most appropriate one for the particular system of interest. The most straightforward is to mix together the purified proteins and isolate the complex by size exclusion chromatography (SEC). Another approach is to mix the cells isolated from individual bacterial cultures that were used to express the two (or more) proteins that form the complex. The cells can be combined, lysed by a variety of means, and the complex purified, for example, using an affinity tag. This approach can be especially powerful if the individual protein components are expressed with different affinity tags, allowing different, sequential affinity chromatographic steps to be employed in isolation of the complex. This method at the same time allows purification of the complex from any excess of one of the partners. Other variations on this approach can also be adopted, for example, combining individual bacterial lysates for the proteins being expressed or purification of one of the components by affinity chromatography, followed by incubation of the resin containing the purified protein with a lysate of the other protein (with or without a tag or with a different affinity tag) in order to capture one or more additional proteins.

In many situations, the above described approaches are not plausible. This is the case when one or more of the individually expressed protein components of a complex are insoluble or poorly soluble and unstable and have a tendency to precipitate. In this case, co-expression of the appropriate protein-binding partner can serve to stabilize and in some

cases improve the solubility behavior of the protein of interest, providing the only way to obtain a viable complex (Tolia and Joshua-Tor, 2006). There are two approaches commonly used for co-expression: insertion of multiple targets in the same vector or co-transformation with compatible plasmids with different resistance genes and origins of replication, each carrying a single construct. Both approaches have certain advantages and disadvantages.

Several bi- or multicistronic vectors for co-expression of proteins have been described (Scheich *et al*., 2007; Selleck and Tan, 2008). We have been using the bicistronic pET-Duet vector (Novagen) or a combination of up to four compatible versions of pET-Duet plasmids that permit simultaneous expression of up to eight different proteins (Kim *et al*., 2006). One advantage of the co-expression strategy from a single vector is that the stoichiometry of a pair of interacting proteins should be easier to control, in as much as both transcripts and corresponding proteins are from a single plasmid, abrogating any effect of plasmid copy number. A disadvantage, however, is that for each new variant of the same protein pair (domains, tags), a new clone must be independently generated.

Using two or more compatible plasmids to express individual proteins of a complex provides a powerful advantage when several combinations of constructs with or without different tags have to be matched together in co-expression trials in order to screen for those that give the best-behaved complex. However, since the various plasmids differ in their copy numbers, and therefore protein expression levels, the desired stoichiometry may not be achieved and an excess of one of the partners is very likely. *A priori*, it is difficult to predict whether co-expression is the best strategy, although if the sequence of interest is predicted to have significant unfolded regions, this would serve as an important indicator favoring co-expression. The unfolded regions can be quite reliably predicted with programs such as DisProt (Sickmeier *et al*., 2007), FoldIndex (Prilusky *et al*., 2005), or GlobPlot (Linding *et al*., 2003).

For bacterial protein complex for which the genes are located adjacent to each other within an operon (polycistronic mRNA unit), another strategy is to clone the entire operon (or an appropriate segment) in one vector for co-expression. A catalog of transcriptional units in *E. coli* has been organized within the RegulonDB database (http://regulondb.ccg.unam.mx/ (Gama-Castro *et al*., 2008)) and EcoCyc (http://ecocyc.org/ (Keseler *et al*., 2009)).

IV. Characterization of Protein Complexes

A. Size Exclusion Chromatography

SEC provides several advantages in the preparation of a protein complex. It is often a powerful purification step prior to crystallization screening, removing protein aggregates, and polishing the protein sample, while at the same time permitting buffer exchange to the final buffer used for crystallization. As a means of characterizing protein complexes, it offers a variety of information. Under conditions that the column is properly calibrated, it gives an apparent molecular mass and, therefore, an approximate stoichiometry for the complex under study. When used sequentially, that is, through re-injection of the main elution peak or analysis of the same sample at different points in time, it offers information on the stability of the protein complex. It also offers some level of dynamic information, in that high-affinity, strongly interacting protein complexes are expected to elute with sharper, better defined peaks than those complexes in which there are fast on–off kinetics among the interacting proteins. In the later case, the peaks from SEC experiments are often observed to be broad or contain distinct leading or lagging shoulders, indicative of multiple components in solution. It is necessary, however, to exercise some care in the analysis of data for protein complexes from SEC. One issue is the effect of total protein concentration on complex formation and apparent stoichiometry – it is often desirable to analyze a protein complex at both higher (15–20 mg/ml) and lower (2–5 mg/ml) protein concentrations, in order to detect differences in behavior. Weaker complexes may only be stable in SEC experiments when the starting protein concentrations are relatively high. The second consideration is the need to pay careful attention to the individual masses of the component proteins; proteins that form dimers or larger homo-oligomers may mask, or be misinterpreted, as protein complexes themselves. Careful analysis of elution fractions by native PAGE and SDS-PAGE can clarify this issue.

B. Dynamic Light Scattering

Dynamic light scattering (DLS) offers a complementary approach to SEC for the characterization of protein complexes. Like SEC, it can also be used to determine the apparent molecular masses of protein

complexes and is a useful tool for establishing complex stability. Compared to SEC, DLS uses relatively little protein, depending on the specific instrument. Using the Wyatt's DynaPro Plate Reader (Wyatt Technology Corp., Santa Barbara, CA; http://www.lightscattering.com), a volume of 40 μl at a concentration of 1 mg/ml is normally sufficient and can be recovered following the experiment. Using such an instrument, it is possible to screen a variety of buffers, pH values, ionic strength conditions, and ligands or other additives in order to establish under what conditions the complex is stable or not. This information can in turn be used to design alternative lysis buffers for purification or protein buffers for crystallization screening. It is beneficial to determine the solution behavior of the protein complex by DLS at relatively high protein concentration, compatible with that to be used for crystallization, in order to evaluate its propensity for aggregation.

C. Native PAGE

Like SEC and DLS, native PAGE can be used to assess the behavior and stability of a protein complex. Native gels can be especially powerful, as they require even less protein than does DLS (as little as a few micrograms), and many conditions, for example, as many as 30, can be tested and evaluated simultaneously. As with DLS, various buffers and additives can be incubated with the protein complex in order to determine (a) if both proteins co-migrate as a complex and (b) to visualize formation of aggregates or other forms of protein heterogeneity that may negatively impact crystallization screening. Once the "best" conditions have been established, an optimized buffer for purification and crystallization screening can be formulated, in part, based on these results.

D. Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a technique that allows one to quantify the affinity of protein–protein interactions. Unlike surface plasmon resonance (SPR), the measurements involve both interacting proteins in solution, thereby avoiding potential artifacts arising from surface immobilization and protein labeling. Also, the stoichiometry of the binding reaction is measured directly during the experiment. A specific

limitation of ITC is that the useful range of accurately measured affinities is somewhat narrower, typically in the range $\sim K_d$ 0.1–10 μM, as compared with SPR. As well, typical ITC experiments require larger quantities of protein, though this problem is somewhat alleviated with the recently released iTC200 instrument from GE/MicroCal (GE Healthcare, Piscataway, NJ; http://www.microcal.com/), which uses a smaller, 200 μl sample cell than previous ITC instruments.

An ITC experiment is performed at a constant temperature while titrating one of the proteins (the "titrand") in the sample cell with the other protein (the "titrant") loaded within a syringe. After each addition of a small aliquot of the titrant, the heat released or absorbed in the sample cell is measured with respect to a reference cell filled with buffer. As the number of injections increases, the quantity of free protein available progressively decreases until the available binding sites become saturated. The binding constant (K_a), molar binding stoichiometry (N), molar binding entropy (ΔS^0), and molar binding enthalpy (ΔH^0) are determined directly from fitting the data, permitting calculation of the Gibbs free energy of binding. By repeating a titration at different temperatures, it is possible to determine the heat capacity change (ΔC_p) associated with the binding reaction, $\Delta C_p = d\Delta H/dT$.

The parameters ΔG, ΔH, ΔS, and ΔC_p are global properties of the system being studied and reflect contributions from the protein–protein binding reaction, conformational changes of the component molecules during association, as well as changes in molecule/solvent interactions and in the state of protonation. The relative magnitude of the change in binding enthalpy, ΔH, primarily reflects the strength of the interactions of the ligand with the target proteins, including van der Waals, salt bridges, hydrogen bonds, and electrostatic interactions, whereas the magnitude of the change in entropy, ΔS, is associated with solvent reorganization and other entropic contributions to binding (Leavitt and Freire, 2001). Typically, hydrophobic interactions involve the favorable burial of non-polar groups from contact with water and are considered to be entropic in nature. The change in heat capacity, ΔC_p, is the parameter in the ITC experiment with the most straightforward structural interpretation. This quantity is directly proportional to the change in the estimated amount of polar and apolar solvent accessible surface area buried on formation of the complex and, to a lesser extent, from changes in molecular vibrations (Freire, 1993). A negative

value for ΔC_p indicates an increase in hydrophobic interactions upon binding.

Ideally, the protein concentration in the sample cell should be at least 10-fold higher than the expected K_d. In practice, titration experiments are often performed with 30–60 μM protein solution in the sample cell and a 10- to 20-fold higher concentration of titrant in the syringe to ensure a final titrant:titrand ratio of (2–4):1 in the reaction cell. Practically, this means that the more soluble and better behaved of the two proteins will usually be used in the syringe. Both protein samples are prepared in or dialyzed against the same buffer to minimize artifacts due to any differences in buffer composition, that is, heats of dilution. Maintaining an identical pH for the titrand and titrant solutions is particularly important. When studying protein–protein interactions, the best way to achieve an identical buffer composition is by passing both proteins through a size exclusion column using the same buffer prior to ITC measurements. While there are no special requirements for the buffers used, the high protein concentrations in ITC experiments make it desirable to use higher capacity (50–100 mM) buffers. This is especially important when using lyophilized peptides that are chemically synthesized, as they usually contain traces of strong acid. When one or both proteins require the presence of a reducing agent, it is recommended to use tris(2-carboxyethyl)phosphine, which is more stable than dithiothreitol (DTT). DTT should be avoided in ITC buffers, as it often results in erratic baselines.

When using fusion proteins for ITC measurements, one should be aware of the possible interference of tags with the interactions being studied, such as occlusion of a binding site and non-specific binding. Special attention should be paid when using tags, which are not monomeric in solution, including the glutathione-*S*-transferase (GST) tag, as this may affect the stoichiometry of the binding event.

E. Surface Plasmon Resonance

SPR is a technique that yields real-time data on protein–ligand interactions and which can be used to guide or validate protein complexes for further structural studies (Masson *et al.*, 2000; McDonnell, 2001; Piliarik *et al.*, 2009). Typically, one binding partner called the ligand is immobilized on the surface of a microfluidic cartridge coated

with an inert organic polymer matrix as sold by GE/Biacore (http://www.biacore.com). Then, a solution of the binding partner (the analyte) is injected at a constant flow rate over the surface. The SPR response is proportional to the mass at the surface, a feature that is used to monitor interactions between molecules in real time. An SPR sensorgram typically consists of a baseline phase (buffer flowing before injection of the analyte), a binding phase (during the injection), and a dissociation phase (after injection), when again buffer flows over the surface. The binding phase depends on both the association and dissociation kinetics, whereas the dissociation phase depends only on the dissociation kinetics. By injecting different concentrations of the analyte, typically over three orders of magnitude around the K_d, the equilibrium, and kinetic rate constants can be determined. SPR chips typically have four independently monitored flow cells on which different ligands can be immobilized and over which a single analyte solution can be flowed. One flow cell should be used as a control to account for bulk refractive index changes and non-specific adsorption of the analyte on the surface.

Although less than 1 μg of protein ligand is normally needed for one immobilization, identifying conditions for the immobilization of the protein remains one of the most crucial steps when designing an SPR experiment. The protein should be immobilized in a way that does not impede its interaction with a potential binding partner. Several immobilization procedures have been proposed. Proteins can be covalently coupled to a surface using thiol or amine-based chemistry. The former often requires engineering of a solvent-exposed cysteine in the protein ligand, whereas the latter involves EDC/NHS-mediated coupling through the side-chain ϵ-amino group of lysines or amino-terminal groups. The protein should be dissolved in a non-reactive buffer (no thiols or amines such as DTT or Tris, respectively) at a pH and salt concentration that allow the protein to interact electrostatically with the surface. However, the amine-coupling procedure can inactivate a protein and can lead to heterogeneity in the mode of ligand immobilization if the ligand has many exposed lysines, which can in turn affect determination of both stoichiometry and affinity. Alternatively, affinity tags such as hexa/deca-histidine, biotin, or GST can be used to immobilize the ligand on Ni-NTA, streptavidin, or anti-GST antibody-coated SPR chips, respectively. Because anti-GST antibody surfaces can be

easily regenerated and the immobilization of GST-fusion proteins does not require covalent modification of the ligand, it remains one of the most popular SPR immobilization methods for investigating protein–protein interactions. The downside of using GST is that this protein is a dimer, which can lead to artificially enhanced affinities due to avidity effects.

The main advantages of SPR for the characterization of protein complexes are the wide range of affinities that can be determined ($K_d < 10^{-3}$ M) and the small amount of protein required for conducting an experiment. The highest concentration of analyte required to accurately measure a K_d should be 10–20 times greater than the dissociation constant in order to reach saturation of the ligand during the binding phase. A single injection requires 20–300 μl depending on the flow rate (10–50 μl/min) and the time required for the binding phase to reach a steady state (30–300 s). High flow rates ($> 30\,\mu$l/min) are recommended to avoid rebinding and mass transport effects. In principle, any chemically inert buffer can be used, although it is common to use a neutral HEPES-buffered saline solution supplemented with a mild non-ionic detergent such as Tween-20 to prevent non-specific hydrophobic interactions. The availability of multiple flow cells on a single chip makes SPR especially suitable for rapidly evaluating the binding activity of multiple single-site mutant proteins with a second partner protein. For example, in addition to the control lane with GST or biotin alone, one can immobilize a wild-type protein and two mutant proteins and test them simultaneously with a single analyte concentration series.

F. Characterization of Protein Complexes by NMR Spectroscopy

Solution NMR spectroscopy has proven to be one of the most powerful techniques for characterizing protein–protein interactions, owing to the development of sensitive NMR instrumentation (higher magnetic fields, cryoprobes, etc.), NMR pulse sequences for assignments, and methods for the isotopic labeling of proteins (Clarkson and Campbell, 2003; Takeuchi and Wagner, 2006). Although solution NMR can be used to determine *de novo* structures of small protein complexes (<30 kDa), its scope is much wider as it can be used to map binding interfaces, monitor conformational changes, or determine kinetic or equilibrium binding

constants. The range of K_d's that is typically measurable by NMR spectroscopy is between 10 μM and 10 mM.

The simplest NMR application consists of recording a one-dimensional proton NMR spectrum of a protein in an aqueous buffer. A good dispersion of relatively narrow resonances in the aliphatic (–2 to 4 ppm) and aromatic/amide (6–11 ppm) regions indicates that the protein is folded and not aggregated. However, a good deal of more information can be obtained through the use of isotopic enrichment. Typically, NMR-detectable isotopes such as ^{15}N and ^{13}C are incorporated into recombinant proteins by growing plasmid-bearing bacteria in a minimal medium supplemented with uniformly labeled $^{15}NH_4Cl$ and/or $^{13}C_6$-D-glucose. The protein can then be purified as usual and concentrated to greater than 100 μM in a low-salt acidic buffer (pH $<$ 7.0) to record multidimensional heteronuclear experiments such as the ^{1}H-^{15}N HSQC.

Kinetic and mechanistic information can be gained through the use of NMR titrations. When a ligand is added in a stepwise fashion to a labeled protein, the position and intensity of a resonance at different molar ratios indicate the exchange rate regime (fast or slow) between the bound and unbound state of the labeled protein. These NMR titrations can also be used to quantify equilibrium binding constants between interaction partners (if the system is in a fast exchange regime) and to determine the stoichiometry of a complex. NMR offers the unique potential for characterizing the affinities of separate binding sites on a protein in a single experiment without recourse to mutagenesis studies, with every resonances acting as a "probe" of a local binding event. For larger complexes, additional protein deuteration is required. TROSY experiments for amide signal assignment in combination with the cross-saturation technique (Nakanishi *et al.*, 2002; Takahashi *et al.*, 2000) provide high-quality definition of protein–protein interfaces.

NMR is particularly useful for studies of transient and ultraweak complexes, which are usually hard to crystallize. In addition to the ^{1}H-^{15}N HSQC experiment, which allows detection of weak yet specific interactions, an NMR method, transferred nuclear Overhauser spectroscopy, is particularly suitable for studies of weak complexes as it relies on fast exchange between free and bound states. This approach is applicable to systems in which one partner is small (e.g., a peptide). NMR methodology as applied to study weak interactions has been described in recent reviews (Prudencio and Ubbink, 2004; Vaynberg and Qin, 2006).

G. Amide Proton/Deuterium Exchange

This method allows mapping protein–protein interaction interfaces (for reviews, see Lanman and Prevelige, 2004; Mandell *et al*., 2005) and is especially useful for systems which have solubility or aggregation problems and which are difficult to crystallize. Amide $^{1}H/^{2}H$ exchange is followed by proteolytic digestion and detection by MS. Both MALDI and electrospray ionization techniques have been used successfully for such experiments. Since fast exchanging amides at the protein–protein interface are most useful for mapping the interfaces, the method is applicable for strong complexes (K_d below 100 nM; (Mandell *et al*., 1998)). Alternatively, NMR spectroscopy can be used as a read-out technique. The advantage of NMR involves the detection of individual amide signals, but it requires isotope labeling (at least ^{15}N). Simpler exchange experiments, in which H_2O is replaced by $^{2}H_2O$ in the buffer, are too slow to capture surface amides. However, recent promising approaches based on using the aprotic solvent DMSO allow detection of protected surface amides through analysis of unfolded samples (Dyson *et al*., 2008; Hoshino *et al*., 2002). Such experiments can be particularly suitable for large oligomeric complexes made up of many small subunits.

V. Experimental Determination of the Structure of Protein Complexes

Due to the efforts of many individual research laboratories and more recently, structural genomics programs worldwide, the number of structures of individual proteins has increased very rapidly. Often, one or more of the components of a complex have a homolog with a known structure. The availability of this information permits, in addition to "high-resolution" methods, such as crystallography or NMR spectroscopy, the utilization of "low-resolution" methods including small-angle X-ray scattering (SAXS), EM, and computational molecular docking.

A. Crystallography of Protein Complexes

As with their preparation and characterization, the crystallization of protein complexes also presents some special challenges. The influence of the variety of ionic strength, pH, and chemical environments

experienced by the protein complex during crystallization screening can mean practically that weaker complexes (those having high micromolar affinity) have a relatively low probability of yielding crystals. Crystallization trials should include, in the case of binary complexes, various molar ratios of the two proteins forming the complex. Protein complexes that form *via* extensive interactions with one another, that have high affinity, and that are relatively stable in solution, would appear to have the best chance of crystallizing. One of the practical outcomes of crystallization is that it is not unusual to obtain crystals of only one member of a protein complex included in the initial screening, and so it is necessary to check any crystals of the putative complex by careful washing and SDS-PAGE in order to confirm the crystal's content. There are few, if any, practical methods so far to evaluate the formation or stability of a given protein complex under conditions of sparse matrix crystallization screening. Efforts have been made, however, to design specific crystallization screens for protein complexes based on the published crystallization conditions of protein complexes contained within the RCSB Protein Data Bank (Radaev *et al*., 2006) and have led to the commercially available ProPlex crystallization screen (http://www.moleculardimensions.com/). We have found the Protein Complex Suite from Qiagen (http://www1.qiagen.com/) to be especially useful. In those cases where the complex is found to be recalcitrant to efforts to crystallize it, the use of mutant proteins for one or more members of the complex, either with the goal of enhancing stability, for example, by the mutation of specific Cys residues, or using surface entropy mutagenesis (Cooper *et al*., 2007) should be tried. As with individual proteins, attempting crystallization of the corresponding complex from various bacterial orthologs is another potentially useful strategy.

Structure determination involves the same procedures as for individual proteins. When the structure of a homolog of one of the proteins in the complex is known, the molecular replacement method has a good chance to lead to a full structure determination without resorting to SeMet substitution and using anomalous dispersion to solve the structure. In such cases, even low-resolution data (3–3.5 Å) may provide sufficient insight into complex formation if the model(s) of the protein (s) from a higher resolution study already exist. In such a case, the crystal of a complex diffracting to a resolution that would be too low for *de novo* structure determination can still be valuable. This is an important factor to keep in mind when assessing crystal usability.

B. NMR Spectroscopy of Protein Complexes

Because the amide ^{1}H and ^{15}N chemical shifts are influenced by different factors in their local environment, a ^{1}H-^{15}N HSQC correlation spectrum effectively provides a fingerprint of the folded state of a protein. Backbone assignments of these spectra can be obtained for proteins up to 25 kDa (or even greater size if the protein is deuterated) by recording triple-resonance spectra that correlate sequential and intra-residue $^{1}H^{N}$, ^{15}N, and $^{13}C^{\alpha}/^{13}C^{\beta}$ signals that can be assigned to specific amino acids in the protein sequence (Muhandiram and Kay, 1994). Once assignments are obtained, the ^{1}H-^{15}N HSQC spectra of a ^{15}N-labeled protein, which shows one peak for every backbone amide group in the protein, can be used to monitor its interaction with any unlabeled ligand that is added to the labeled protein solution. An example of this method as applied to the interaction of a yeast protein with a peptide of its cognate binding partner is shown in Fig. 2. Chemical shift perturbations (peak displacement or broadening) arise from changes in the environment of the NMR nucleus and can be

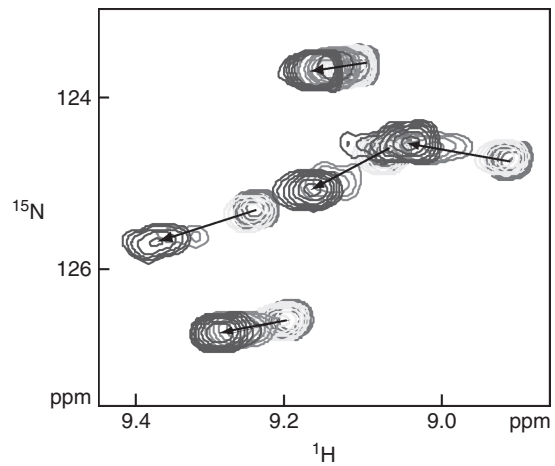

FIG. 2. NMR analysis of Ste50_RA/Opy2 peptide interactions. Overlay of regions of ^{1}H-^{15}N HSQC spectra for solutions containing a free RA domain (red) and increasing ratios of unlabeled Opy2 peptide titrated into ^{15}N-RA domain (orange, blue, and green, respectively). The arrows indicate displacement of selected ^{15}N-RA domain peaks. (See Color Insert.)

caused by direct protein–protein interactions or conformational changes induced by the binding event. These perturbations can then be mapped onto a protein structure to reveal interaction sites. Once the binding sites are known, the relative orientation of proteins in a complex can be readily determined using residual dipolar couplings (RDCs) measured in dilute liquid crystals (Bax *et al*., 2001). RDCs can generally be fitted to structures determined in solution or to crystal structures, and can be used to derive more accurate structural models, as exemplified by the NMR-based docking of an acyl carrier protein to the acyl transferase enzyme LpxA (Jain *et al*., 2004).

NMR spectroscopy can also provide dynamic information that can be used to further guide structural studies. Flexible termini or long loops can hinder crystallization and their removal can increase the likelihood of generating diffracting crystals (Page, 2008), but it is sometimes difficult to identify these regions with certainty through sequence analysis alone. The ^{15}N-^{1}H heteronuclear NOE experiment (Kay *et al*., 1989) can be used to determine domain boundaries and flexible regions of proteins, as it depends on the rotational correlation time, which is itself a function of the folded state of the amide group. Fast-tumbling amide groups have negative heteronuclear NOE values whereas amide located in folded regions have NOE values above 0.6, a feature that can be used to delineate protein domains and to engineer recombinant protein constructs amenable to crystallization.

C. Small-Angle X-ray Scattering

In recent years there has been a revival in the use of SAXS to analyze protein shapes and conformations in solution (Koch *et al*., 2003; Putnam *et al*., 2007). The basic experimental design consists of exposing a protein solution to a collimated beam of hard X-rays and detecting X-rays scattered at a small angle. Data can be acquired at various synchrotron radiation facilities (SIBYLS beamline at the ALS, X33 at DESY, etc.) or using in-house instruments, such as the SAXSess Kratky camera (Anton Paar USA Inc, Ashland, VA; http://www.anton-paar.com/), the NanoStar U (Bruker AXS Inc., Madison, WI; http://www.bruker-axs.de/nanostar.html), and the PSAXS (Rigaku Corp., The Woodlands, TX; http://www.rigaku.com/index_en.html). The resulting rotationally averaged scattering curve is a scattering vector (q)-dependent intensity profile $I(q)$ that

depends on the electron density contrast between the protein and the buffer, the shape and size of the protein, and intermolecular interferences (Glatter and Kratky, 1982). By recording SAXS profiles at different protein concentrations and by subtracting the buffer contribution with a blank measurement, it is possible to extrapolate the scattering curve corresponding to that of an infinitely diluted single particle. Any buffer can be used, but the salt concentration should be kept below 0.5 M to reduce background scattering. Glycerol may be added (up to 5%) to reduce radiation damage from synchrotron radiation. It is especially important that the buffer is the same in the blank and protein samples: ideally, the protein should be dialyzed and the dialyzate can be used as a blank. Gel filtration is a useful last step for SAXS analysis as it removes aggregates that can compromise data quality.

Various mathematical transformations can be applied to a scattering curve to evaluate the state of a protein complex in solution. The Guinier plot (q^2 vs. ln(I)) should be linear at very low angles ($q^*R_g<1.3$) for a monodisperse solution of globular particles (Guinier and Fournet, 1955). Deviation from linearity in a Guinier plot indicates the presence of aggregates, which compromise the scattering curve, thus preventing further analysis (Fig. 3A). This quick and easy diagnostic can be run prior to embarking on crystallization trials with a concentrated protein solution, as the absence of aggregates correlates strongly with the probability of generating diffracting protein crystals (Jancarik *et al*., 2004). If the Guinier plot is linear, the radius of gyration (R_g) and the forward scattering I_0 can be extracted from the slope and intercept of the plot, respectively (Fig. 3A). The latter can be used to calculate the molecular weight of the protein in solution as it depends solely on the mass and concentration of the scattering particle. This information can be used to determine the stoichiometry of a protein complex or oligomer. Finally, if the Guinier plot obtained at different protein concentrations shows a marked change in R_g and I_0/c values, it reveals the formation of concentration-dependent oligomers that can compromise SAXS analysis and crystallization experiments. Another useful diagnostic tool is the Kratky plot (q vs. I^*q^2), which is an indicator of the folded state of a protein (Putnam *et al*., 2007). Folded globular proteins show a typical bell curve whose q value at the maximum can be used to determine an approximate molecular weight, whereas unfolded proteins show a plateau-shaped curve with no distinct peak.

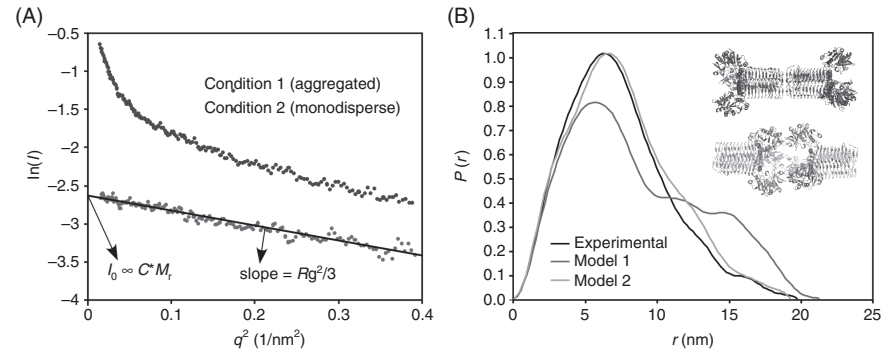

FIG. 3. Use of SAXS in structural genomics of protein complexes. (A) The Guinier plot, which is the square of scattering vector q ($4\pi \sin\theta/\lambda$) versus the natural logarithm of the intensity, can be used as a diagnostic of protein aggregation and size in solution. An example is shown here for a protein complex concentrated to 4 mg/ml in two different buffer conditions. Monodisperse globular particle solution should give a linear Guinier within the range q_{min} to $q_{max}{*}R_g < 1.3$, as observed in condition 2. The slope enables calculation of R_g whereas the intercept enables the molecular weight to be determined if the concentration is known accurately. (B) The pair-density distribution function $P(r)$ allows the identification of a biological unit from a crystal structure that gives two possible hexameric forms (models 1 and 2). The experimental $P(r)$ function is much more similar to the hexamer #2, which implies that the protein forms hexamer of this type in solution. (See Color Insert.)

Although the information content of a scattering curve is relatively low, the combination of SAXS with other types of structural data can yield valuable information on the conformation of protein complexes in solution. The group of Dmitri Svergun at the EMBL-Hamburg has developed several data processing and analysis programs for SAXS (http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html). A one-dimensional distance distribution function, $P(r)$, can be calculated from a scattering curve by inverse Fourier transformation using programs such as GNOM (Svergun *et al.*, 2001) or GIFT (Bergmann *et al.*, 2000). The $P(r)$ function provides a first glance at global structural features of a protein complex, that is, its maximum diameter (D_{max}), the distance between globular domains, and whether it is spherical or elongated (Fig. 3B). Scattering curves and the $P(r)$ function can be computed directly from PDB coordinates and compared with experimental curves using the program CRYSOL (Svergun *et al.*, 1995). This

procedure is useful to determine the oligomeric state of a protein complex in solution and to identify physiologically relevant binding interfaces in protein crystal structures (Fig. 3B). SAXS can also be used to ascertain conformational changes in a protein induced upon addition of small molecule inhibitors or cofactors. If the structures of individual components forming a complex are available, it is possible to perform a rigid-body docking of these structures to fit the experimental data. Finally, *ab initio* shape reconstruction can also be performed to model the shape of the scattering particle using an ensemble of dummy spheres with restraints that best represent the compactness of protein structures (Svergun *et al*., 2001).

D. Computational Molecular Docking with Experimental Restraints

In a number of cases, the individual structures of interacting proteins are available, although efforts to co-crystallize the corresponding protein complex prove to be elusive. This is rather typical for transient interactions, where proteins do not possess high binding affinity. One of the possible ways to overcome this problem is the use of molecular docking to produce low-resolution complex models. The docking methods are classified into global methods based on geometric matching, Monte Carlo methods, and restraint-based High Ambiguity Driven biomolecular DOCKing (HADDOCK) approaches (Vajda and Kozakov, 2009). While the accuracy and reliability of predictions have improved significantly in recent years, the results of the first community-wide critical assessment of predicted interactions (CAPRI) experiment suggest that the addition of experimental information is critical to improve the accuracy of predicted models (Janin *et al*., 2003).

The HADDOCK program makes ready use of experimental data to drive docking, unlike other approaches based on a combination of energetics and shape complementarity (Dominguez *et al*., 2003). A variety of data, including NMR measurements, site-directed mutagenesis, and sequence conservation analysis, can be introduced as restraints for docking calculations. In particular, NMR is a valuable tool in obtaining HADDOCK restraints, including ambiguous interaction restraints (AIRs) from chemical shift mapping of interacting surfaces, unambiguous distance restraints in the form of intermolecular NOEs, and orientational restraints as RDCs. While NMR has limitations in terms of the size of

proteins that can be easily studied, binding data can often be obtained for individual domains of multi-domain proteins.

The predicted structural model of the complex can be further confirmed using site-directed mutagenesis. Though the computational models lack the atomic resolution details of binding determinants, binding-deficient mutants can be designed and used by biologists to test the physiological importance of the protein complex in question, thereby providing important new information.

VI. Examples of Characterization of Bacterial Protein Complexes

In the following we provide several examples from our own work, illustrating various applications of the methodologies described in the preceding sections. These examples illustrate both the characterization of protein complexes and their structure determination by a variety of methods.

A. TtdA–TtdB: The L-tartrate Dehydratase of E. coli

E. coli utilizes L-tartrate as a fermentable carbon source under anaerobic conditions, *via* conversion to oxaloacetate. A hetero-tetrameric $\alpha_2\beta_2$ enzyme made up of two copies of TtdA and TtdB constitutes the *E. coli* L-tartrate dehydratase enzyme (Reaney *et al*., 1993). The TtdA subunit contains a 4Fe–4S cluster, giving the protein a brown color and making the enzyme sensitive to oxygen. This enzyme is specific for the L-isomer of tartrate, with D-tartrate metabolized via the enzyme fumarase (Kim *et al*., 2007a). Crystal structures are available for D-tartrate dehydratase (Yew *et al*., 2006), but not for the L-specific enzyme. In order to obtain a structure for *E. coli* L-tartrate dehydratase, we have cloned, expressed, and purified the TtdA and TtdB subunits and attempted re-constitution of the enzyme by mixing and analysis by gel filtration chromatography. In this experiment, we did not succeed in obtaining the hetero-tetrameric complex. As an alternative strategy, we PCR-amplified the *ttdA-ttdb* genes in tandem and cloned them into several expression vectors. Co-expression of the two proteins together readily resulted in a well-behaved TtdA-TtdB complex. Efforts to crystallize this enzyme in the presence and absence

of L-tartrate are ongoing. Our work on this complex illustrates the advantage, in this instance, of cloning and expression of both contiguous genes as one unit, as opposed to expressing, purifying, and mixing purified subunits individually.

B. MnmG–MnmE: An Enzyme Complex Involved in tRNA Modification

Bacteria extensively process and modify tRNA molecules, in part, to enhance codon–anticodon recognition, necessary for fidelity during translation (Bregeon *et al*., 2001). One such modification involves the attachment of a 5-carboxymethylaminomethyl (cmnm) group onto the uridine base of U34 of selected tRNAs. A pathway responsible for synthesizing the cmnm group on U34, consisting of several enzymes, has been elucidated in *E. coli*. A FAD-dependent oxidoreductase, MnmG, GTPase, and MnmE function together to generate an intermediate form of the cmnm group subsequent to its methylation by MnmC. Using gel filtration chromatography, it has been shown that MnmG and MnmE form a 2:2 $\alpha_2\beta_2$ hetero-tetrameric complex *in vitro* (Yim *et al*., 2006). Co-expression and purification of MnmG and MnmE in pET-Duet revealed that MnmE is able to co-purify with a His-tagged form of MnmG, although the quantity of MnmE appeared to be sub-stoichiometric, illustrating one of the limitations of co-expression in the case of weaker protein complexes. The MnmG interacting surface involves the C-terminal ~50–70 residues, as deletion of this region results in loss of complex formation with MnmE (Meyer *et al*., 2008, M. Cygler, unpublished data). Additional studies, including alanine-scanning mutagenesis of the C-terminal region of MnmG, combined with protein–protein binding analysis by SPR or ITC, will be required to more precisely map the MnmG–MnmE interface.

C. HypE–HypF and HypC–HypD: Protein Complexes Involved in [NiFe] Hydrogenase Maturation

The *E. coli* genome encodes three hydrogenase systems in which the large hydrogenase subunit incorporates a [NiFe] metallocenter for its function (Forzi *et al*., 2007). Hydrogenases make use of a number of maturation enzymes to synthesize the CO and CN ligands which are incorporated into the metal center. The two proteins required for

synthesis of the cyano moiety are HypE and HypF (Jacobi *et al*., 1992). HypF is an ATP-dependent enzyme, utilizing carbamoyl phosphate to generate a thiocarabamate moiety, which is transferred to the C-terminus of HypE (Reissmann *et al*., 2003). This thiocarbamate then undergoes dehydration to a cyano moiety in an ATP-dependent reaction catalyzed by HypE (Blokesch *et al*., 2004b). Previously, we have determined the crystal structure of *E. coli* HypE and characterized its interaction with HypF using several approaches (Rangarajan *et al*., 2008). The affinity of the two proteins for one another was measured using both SPR and ITC, giving a K_d of approximately 400 nM. Gel filtration and scanning densitometry of SDS-PAGE gels were used to determine that the proteins form a 2:2 hetero-oligomeric complex. Depending on protein concentration, a 1:1 HypE–F complex also was observed. The SPR sensorgrams were consistent with a model in which a conformational change occurs in HypE–F upon complex formation. In order to define more precisely the binding site between HypF and HypE, a truncated version of HypF lacking residues 1–191 was PCR-amplified, expressed in *E. coli*, purified, and mixed with purified HypE protein. Analysis of the protein mixture revealed that this shorter version of HypF still forms a complex with HypE, and that the HypF deletion in the presence of HypE is apparently better behaved than the truncated version of the protein alone (Fig. 4A). The apparent mass of the complex by DLS, 100 kDa, is most consistent with a 1:1 stoichiometry of the two proteins (Fig. 4B). This apparent stoichiometry if further validated by SDS-PAGE analysis of fractions from SEC analysis of the complex (Fig. 4C). Efforts to further delineate the HypE–F interaction surface are ongoing.

In addition to complex formation between HypE and HypF, two other proteins, HypC and HypD, have been shown to form a complex together, as well possibly also with bound HypE (Blokesch *et al*., 2004a). The function of the HypC–D complex is to presumably deliver Fe to the hydrogenase active site, as HypC binds Fe as a $[4Fe–3S]^{2+}$ cluster (Blokesch *et al*., 2004a) and also forms a specific interaction with the precursor form of the hydrogenase III large subunit (Drapal and Bock, 1998). The molecular details of how HypC–D interact and donate the Fe atom to the large hydrogenase subunit remains unknown, although a model for the HypC–D complex has been proposed, based on the crystal structures of both proteins from *Thermococcus kodakaraen*sis (Watanabe *et al*., 2007). We have focused on determining the co-crystal structure of *E. coli* HypC–

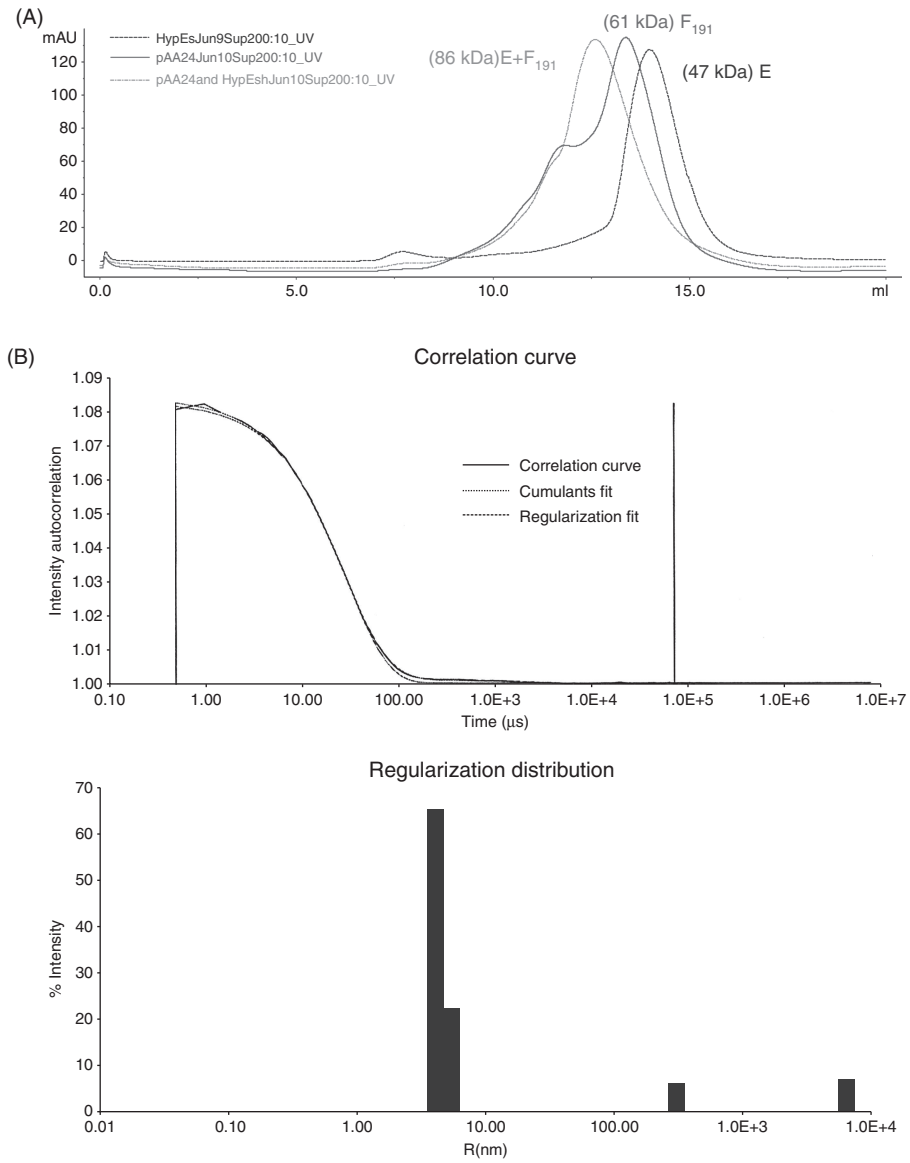

FIG. 4. (Continued)

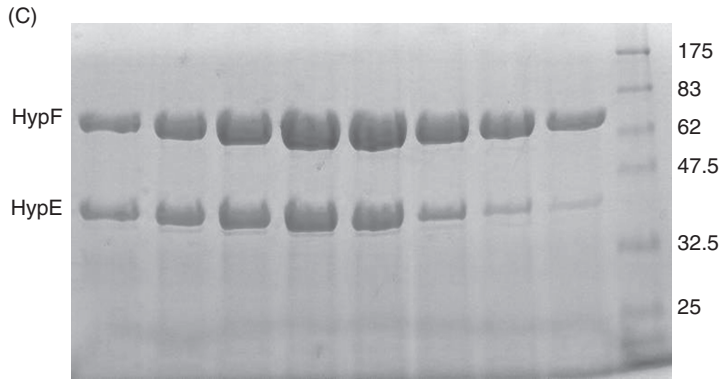

FIG. 4. Protein–protein complex between *E. coli* HypE and a deletion construct (Δ1–191) of *E. coli* HypF. (See Color Insert.)

D in order to address these questions. We have demonstrated complex formation between HypC–D by SEC, as well as by ITC. Titration of HypC into HypD yielded an association constant, K_a, of 200 nM, with an apparent 1:1 stoichiometry of the two proteins (Fig. 5). Based on the thermodynamic data, the interaction between the two proteins is primarily entropy driven, with the overall binding reaction being endothermic.

D. YaeO–Rho: Inhibition of Rho-Dependent Transcription Termination

Transcription termination is the process where a nascent RNA is released from its complex with RNA polymerase and the DNA template. In bacteria, two main mechanisms of transcription termination have been described. These mechanisms, commonly referred to as Rho-independent and Rho-dependent termination, are essential for the regulation of bacterial gene expression (Richardson and Greenblatt, 1996). Rho-dependent termination requires the presence of a hexameric helicase, Rho (Brown *et al.*, 1981), an essential transcription factor that binds nucleic acids at specific termination sites (rut), and translocates along the RNA until it reaches the transcription complex (Geiselmann *et al.*, 1993; Platt, 1994; Richardson, 1996). One of the Rho-specific inhibitors of transcription is the product of the *yaeO* gene, which reduces termination

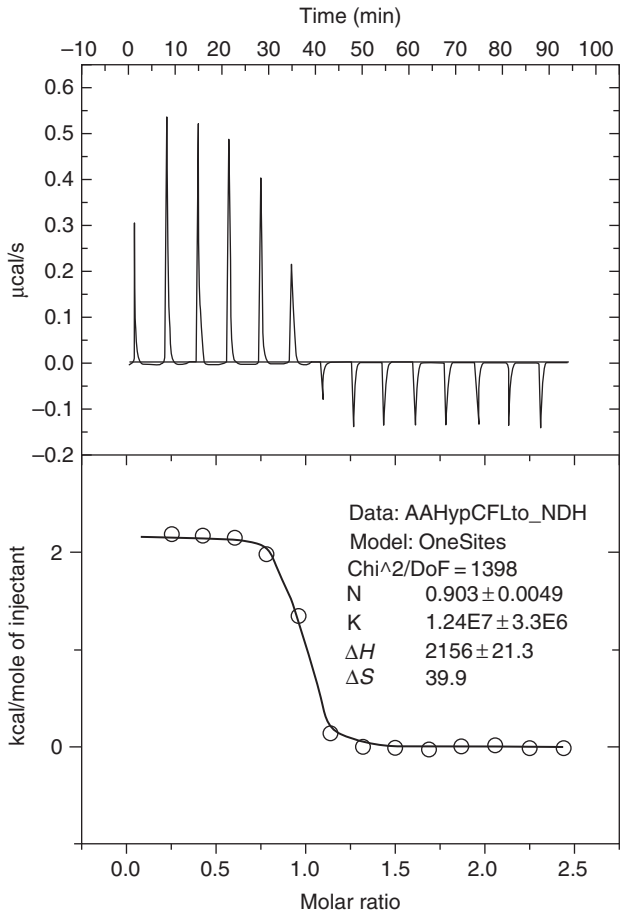

FIG. 5. ITC experiment to determine the association constant between the hydrogenase maturation proteins, HypC and HypD, from *E. coli*. The calculated association constant is 200 nM, with an apparent *n* value of 1.

in the Rho-dependent bacteriophage terminator tL1, and upstream the autogenously regulated gene *rho* (Pichoff *et al*., 1998). Overexpression of YaeO can cause the pleiotropic suppression of conditional lethal mutations in cell division and heat shock genes, such as *ftsQ, ftsA, grpE, groEL*, and *groES* (Pichoff *et al*., 1998).

We first determined the NMR solution structure of YaeO that revealed a topologically similar fold to that of the RNA-binding domain of small ribonucleoproteins (Sm-fold) (Gutierrez *et al*., 2007). In order to understand the mechanism of transcription termination inhibition by YaeO, NMR experiments were used to characterize the interaction of YaeO with Rho *in vitro*. We used the N-terminal fragment (residues 1–130), referred as Rho130, that corresponds to the primary RNA binding site of Rho and has been shown to be a good model of Rho-oligonucleotide interactions (Briercheck *et al*., 1998). The titration resulted in mapping the binding site of Rho130 on YaeO, which consists of the N- and C-termini, helix α1, and strands β3, β4, β5, and β7. These regions localize to one edge of the β-sandwich with clustered acidic residues. As the structure of Rho130 has been also determined (Briercheck *et al*., 1998), we mapped the YaeO binding site on Rho130 that partially overlaps with the RNA binding surface, suggesting a mechanism of transcription termination inhibition.

As NMR titration data for the YaeO–Rho interaction was obtained for both proteins, a docking model of the complex was calculated using HADDOCK (Dominguez *et al*., 2003). AIRs were derived from the NMR titration data by selecting residues with both the biggest chemical shifts and solvent accessibility. The resulting model is compatible with the hexameric Rho structure and reflects the charge complementarities of the interacting protein surfaces (Fig. 6B). This is consistent with *in vitro* binding results that show that the YaeO–Rho interaction is salt dependent and can be disrupted at high ionic strengths (Pichoff *et al*., 1998). Importantly, the structural model was used to design the D14K, E19K, and E52K YaeO mutants that prevent inhibition of Rho activity using an *in vivo* β-galactosidase assay (Gutierrez *et al*., 2007).

E. SufBCD: A Protein Complex Involved in Bacterial Fe–S Cluster Synthesis Under Stress Conditions

Iron–sulfur clusters are important metal cofactors in enzymes involved in a wide range of biological processes including respiration and the regulation of gene expression (Johnson *et al*., 2005). Under most conditions, *E. coli* uses a general ISC pathway for the assembly of Fe–S proteins, but under conditions of oxidative stress bacteria can employ an alternative, Suf pathway (Nachin *et al*., 2001). This pathway is well conserved in microorganisms and may play a role in bacterial pathogenesis

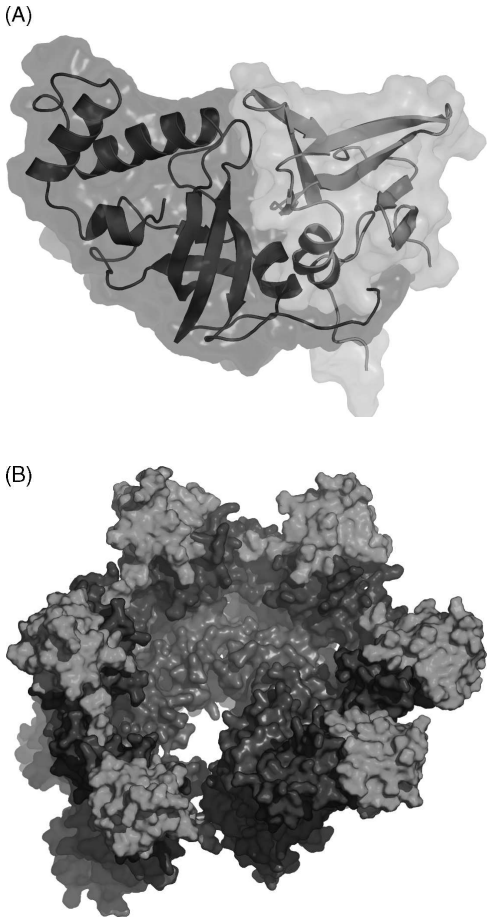

FIG. 6. (A) Model of the YaeO–Rho130 complex calculated with the program HADDOCK. YaeO is colored green and Rho130 is purple. (B) The YaeO/Rho130 model is compatible with the open-ring, hexameric form of Rho, accommodating six molecules of YaeO (in green) per Rho hexamer. The RNA-binding domain of Rho is colored purple and the ATP-hydrolysis domain is blue. (See Color Insert.)

by helping bacteria deal with the host immune response. The Suf system consists of a cysteine desulfurase (SufS) and five accessory proteins (Suf A, B, C, D, and E). SufBCD has been found to associate as a stable complex and to act synergistically with SufE to enhance the activity of the

desulfurase SufS (Outten *et al*., 2003). Our goal is to determine the three-dimensional structure of the SufBCD complex in order to examine the protein–protein interactions among the components of the cysteine desulfurase activator complex.

Initially the SufB, SufC, and SufD proteins were expressed separately in order to assemble the complex from purified components. While it proved possible to obtain pure samples of SufC and SufD using Ni-affinity chromatography, the final yield of SufC was low. We were able to improve the yield by combining the cell lysates from separately expressed SufC and SufD and co-purifying the complex. Unfortunately, expression of SufB on its own resulted in protein accumulating as insoluble aggregates, thwarting efforts to assemble the SufBCD complex from purified components. In an attempt to circumvent this problem, we decided to co-express all three components and purify the SufBCD complex. To accomplish this we cloned the entire *sufABCDSE* operon using an existing protocol (Outten *et al*., 2003). Purification of the SufBCD complex was then carried out using a combination of anion exchange and SEC. Preliminary SDS-PAGE gels of these samples indicate the presence of proteins with molecular weights of 29, 47, and 55 kDa, values consistent with those of SufC, SufD, and SufB, respectively. By expressing the *suf* operon as a unit, we have been able to achieve the partial purification of the *E. coli* SufBCD complex. Additional purification steps will be required prior to crystallization trials.

F. SdbA–CipA: The Mechanism of Cellulosome Cell-Surface Attachment

The cellulosome is a large cell-surface bound multi-enzyme complex that synergistically degrades plant cell wall polysaccharides. First discovered in the thermophilic anaerobe *Clostridium thermocellum*, cellulosomes have since been identified in a variety of other anaerobic bacteria, ruminal bacteria, and anaerobic fungi (Bayer *et al*., 2004). In general, the cellulosome can be divided into three modular protein components: (1) cell-surface proteins, (2) scaffold proteins, and (3) enzymes (Fig. 7A). Cellulosome assembly is mediated by conserved calcium-dependent protein–protein interaction modules called cohesins (Coh) and dockerins (Doc) (Bayer *et al*., 2004). At least three types of Coh–Doc pairings exist, although they do not exhibit any cross specificity for other types (Ding

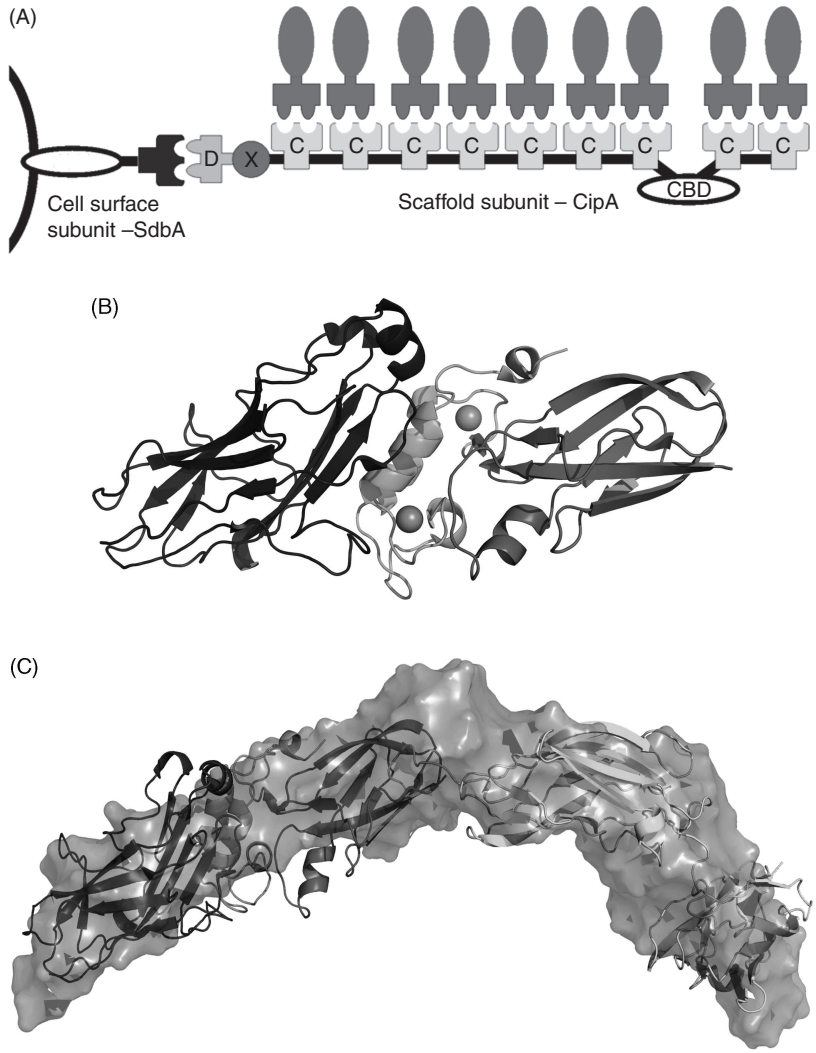

FIG. 7. (A) Cellulosome architecture: dark blue, type II Coh; green, type II Doc; pink, X-module; yellow, type I Coh; orange, type I Doc-containing enzymes. (B) Coh–DocX crystal structure: dark blue, type II Coh; green, type II Doc; pink, X-module; orange, calcium ions. (C) Crystal structures of Coh–DocX and two type I Cohs aligned with the SAXS structure of CohII–DocXCohICohI: gray, SAXS envelope; dark blue, type II Coh; green, type II Doc; pink, X-module; yellow, type I Coh; orange, calcium ions. (See Color Insert.)

et al., 2001). The scaffold subunit of *C. thermocellum*, CipA contains a C-terminal type II Doc module that anchors it to type II Coh-containing cell-surface proteins, nine type I Coh modules that bind to type I Doc-containing hydrolases, an X-module of unknown function, and a cellulose-binding domain (CBD) (Fujino *et al*., 1993) (Fig. 7A). To date, three type II Coh-containing cell-surface proteins and more than 70 type I Doc-containing cellulases and hemicellulases have been identified in *C. thermocellum*, which offers a large degree of subunit plasticity (Bayer *et al*., 2004).

We have characterized the mechanism of cellulosome cell-surface attachment in *C. thermocellum*, both biochemically and structurally, using several methods including ITC, DSC, SAXS, and X-ray crystallography. The type II Coh module from SdbA and a C-terminal fragment of CipA containing both the type II Doc and the X-module (DocX) were cloned, expressed, purified from inclusion bodies under denaturing conditions, and refolded (Adams *et al*., 2004). ITC and DSC data indicate 1:1 binding and an ultra-high-affinity association constant (K_a) of 1.44×10^{10} M^{-1} (Adams *et al*., 2006). We solved the crystal structure of this complex to 2.1 Å resolution using X-ray crystallography (Adams *et al*., 2006) (Fig. 7B). The complex is elongated, with a highly hydrophobic interface and extensive hydrogen bonding between the X-module and both Doc and Coh modules (Adams *et al*., 2006). The elongated structure allows the cellulosome to extend away from the surface of the cell to contact its extracellular substrate. Based on the observed X-module interactions, we propose a role for the X-module in Doc structure stability as well as in enhanced Coh recognition. The structure also provides a rationale for type I and type II Coh–Doc specificity, based on differences in Doc orientations, interface physicochemical properties, and X-module involvement. SAXS was used to gain further insight into the arrangement of neighboring modules. We solved the structure of the type II Coh of SdbA in complex with a fragment of CipA containing the type II Doc, the X-module, and two of the neighboring type I Cohs using SAXS (Fig. 7C). The SAXS envelope reveals an elongated, bent structure with no interactions evident between either of the type I Cohs and the other modules. This work illustrates a unique mode of cell-surface attachment, delineates a putative function for the previously uncharacterized X-module, and provides rationale for the specificity of type I and type II Coh–Doc modules.

VII. Future Perspectives

The expression, purification, and crystallization of protein complexes pose challenging problems, both technically and scientifically. Tremendous effort will be required to validate high-throughput proteomics data and expand the list of tractable bacterial protein complexes that will be amenable to structural analysis. We can be certain that a great many protein complexes remain to be discovered. New methodologies, such as the use of whole-genome protein arrays (Oishi *et al*., 2006), and *in vivo* split-protein complementation assays (Remy *et al*., 2007) will contribute to discovering new protein–protein interactions. One approach that is likely to identify several new complexes is the systematic expression of bacterial operons, incorporating affinity tags at different points within the operon in order to "fish-out" interacting proteins. There is good reason to believe that functional linkages among proteins, including genetic linkage, can be used as a means to predict proteins that will interact (Kim *et al*., 2007b). Knowledge of both predicted and experimentally observed disorder of individual proteins may offer an alternative avenue to identify new interacting proteins, as the presence of disorder correlates with some extent to a propensity to form a complex (Hegyi *et al*., 2007). Despite the hurdles, many new and ultimately important biological insights will emerge from the structural analysis of protein complexes in the years ahead.

Acknowledgments

The authors wish to thank many of their colleagues whose work is described in this chapter, especially Eunice Ajamian Abdalin Asinas, Wayne Miller, Christine Munger, Ariane Proteau, Anna Rog, Rong Shi, as well many others for their contributions to refining the various methods described herein. This research was supported by a grant (GSP48370) from the Canadian Institutes of Health Research and by the National Research Council Canada. A.M.B. is a Canada Research Chair in structural biology. This is National Research Council of Canada publication no. 50671.

References

Adams, J. J., Jang, C. J., Spencer, H. L., Elliott, M., and Smith, S. P. (2004). Expression, purification and structural characterization of the scaffoldin hydrophilic X-module from the cellulosome of *Clostridium thermocellum*. *Protein Expr. Purif*. **38**, 258–263.

Adams, J. J., Pal, G., Jia, Z., and Smith, S. P. (2006). Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc. Natl. Acad. Sci. USA*. **103**, 305–310.

Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H. C., Hirai, A., Tsuzuki, K., Nakamura, S., taf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006). Large-scale identification of protein–protein interaction of *Escherichia coli* K-12. *Genome Res.* **16**, 686–691.

Bahadur, R. P., Chakrabarti, P., Rodier, F., and Janin, J. (2004). A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.* **336**, 943–955.

Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. *Methods Enzymol.* **339**, 127–174.

Bayer, E. A., Belaich, J. P., Shoham, Y., and Lamed, R. (2004). The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* **58**, 521–54. 521–554.

Bergmann, A., Fritz, G., and Glatter, O. (2000). Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing (BSSA). *J. Appl. Crystal.* **33**, 1212–1216.

Beyer, A., Bandyopadhyay, S., and Ideker, T. (2007). Integrating physical and genetic maps: From genomes to interaction networks. *Nat. Rev. Genet.* **8**, 699–710.

Blokesch, M., Albracht, S. P., Matzanke, B. F., Drapal, N. M., Jacobi, A., and Bock, A. (2004a). The complex between hydrogenase-maturation proteins HypC and HypD is an intermediate in the supply of cyanide to the active site iron of [NiFe]-hydrogenases. *J. Mol. Biol.* **344**, 155–167.

Blokesch, M., Paschos, A., Bauer, A., Reissmann, S., Drapal, N., and Bock, A. (2004b). Analysis of the transcarbamoylation-dehydration reaction catalyzed by the hydrogenase maturation proteins HypF and HypE. *Eur. J. Biochem.* **271**, 3428–3436.

Bregeon, D., Colot, V., Radman, M., and Taddei, F. (2001). Translational misreading: A tRNA modification counteracts a +2 ribosomal frameshift. *Genes Dev.* **15**, 2295–2306.

Briercheck, D. M., Wood, T. C., Allison, T. J., Richardson, J. P., and Rule, G. S. (1998). The NMR structure of the RNA binding domain of *E. coli* rho factor suggests possible RNA-protein interactions. *Nat. Struct. Biol.* **5**, 393–399.

Brown, S., Brickman, E. R., and Beckwith, J. (1981). Blue ghosts: A new method for isolating amber mutants defective in essential genes of *Escherichia coli*. *J. Bacteriol.* **146**, 422–425.

Butland, G., Babu, M., Diaz-Mejia, J. J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A. G., Pogoutse, O., Mori, H., Wanner, B. L., Lo, H., Wasniewski, J., Christopolous, C., Ali, M., Venn, P., Safavi-Naini, A., Sourour, N., Caron, S., Choi, J. Y., Laigle, L., Nazarians-Armavil, A., Deshpande, A., Joe, S., Datsenko, K. A., Yamamoto, N., Andrews, B. J., Boone, C., Ding, H., Sheikh, B., Moreno-Hagelseib, G., Greenblatt, J. F., and Emili, A. (2008). eSGA: *E. coli* synthetic genetic array analysis. *Nat. Methods* **5**, 789–795.

Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., and Emili, A. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537.

Carpousis, A. J. (2007). The RNA degradosome of *Escherichia coli*: An mRNA-degrading machine assembled on RNase E. *Ann. Rev. Microbiol.* **61**, 71–87.

Clarkson, J., and Campbell, I. D. (2003). Studies of protein-ligand interactions by NMR. *Biochem. Soc. Trans.* **31**, 1006–1009.

Collins, M. O., and Choudhary, J. S. (2008). Mapping multiprotein complexes by affinity purification and mass spectrometry. *Curr. Opin. Biotechnol.* **19**, 324–330.

Collins, R. F., Beis, K., Dong, C., Botting, C. H., McDonnell, C., Ford, R. C., Clarke, B. R., Whitfield, C., and Naismith, J. H. (2007). The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **104**, 2390–2395.

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007). Protein crystallization by surface entropy reduction: Optimization of the SER strategy. *Acta Crystallogr. D* **63**, 636–645.

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645.

Ding, S. Y., Rincon, M. T., Lamed, R., Martin, J. C., McCrae, S. I., Aurilia, V., Shoham, Y., Bayer, E. A., and Flint, H. J. (2001). Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* **183**, 1945–1953.

Dominguez, C., Boelens, R., and Bonvin, A. M. (2003). HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737.

Drapal, N., and Bock, A. (1998). Interaction of the hydrogenase accessory protein HypC with HycE, the large subunit of *Escherichia coli* hydrogenase 3 during enzyme maturation. *Biochemistry* **37**, 2941–2948.

Driessen, A. J.M., and Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. *Ann. Rev. Biochem.* **77**, 643–667.

Dyson, H. J., Kostic, M., Liu, J., and Martinez-Yamout, M. A. (2008). Hydrogen–deuterium exchange strategy for delineation of contact sites in protein complexes. *FEBS Lett.* **582**, 1495–1500.

Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402**, 86–90.

Forzi, L., Hellwig, P., Thauer, R. K., and Sawers, R. G. (2007). The CO and CN(–) ligands to the active site Fe in [NiFe]-hydrogenase of *Escherichia coli* have different metabolic origins. *FEBS Lett.* **581**, 3317–3321.

Freire, E. (1993). Structural thermodynamics: Prediction of protein stability and protein binding affinities. *Arch. Biochem. Biophys.* **303**, 181–184.

Fujino, T., Beguin, P., and Aubert, J. P. (1993). Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. *J. Bacteriol.* **175**, 1891–1899.

Gama-Castro, S., Jimenez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Penaloza-Spinola, M. I., Contreras-Moreira, B., Segura-Salazar, J., Muniz-Rascado, L., Martinez-Flores, I., Salgado, H., Bonavides-Martinez, C., Abreu-Goodger, C.,

Rodriguez-Penagos, C., Miranda-Rios, J., Morett, E., Merino, E., Huerta, A. M., Trevino-Quintanilla, L., and Collado-Vides, J. (2008). RegulonDB (version 6.0): Gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.* **36**, D120–D124.

Gamba, P., Veening, J. W., Saunders, N. J., Hamoen, L. W., and Daniel, R. A. (2009). Two-step assembly dynamics of the *Bacillus subtilis* divisome. *J. Bacteriol.* **191**, 4186–4194.

Geiselmann, J., Wang, Y., Seifried, S. E., and von Hippel, P. H. (1993). A physical model for the translocation and helicase activities of *Escherichia coli* transcription termination protein Rho. *Proc. Natl. Acad. Sci. USA* **90**, 7754–7758.

Glatter, O., and Kratky, C. (1982). "Small Angle X-ray Scattering." Academic Press, London.

Goll, J., Rajagopala, S. V., Shiau, S. C., Wu, H., Lamb, B. T., and Uetz, P. (2008). MPIDB: The microbial protein interaction database. *Bioinformatics* **24**, 1743–1744.

Guinier, A., and Fournet, F. (1955). "Small Angle Scattering of X-rays." Wiley Interscience, New York.

Gutierrez, P., Kozlov, G., Gabrielli, L., Elias, D., Osborne, M. J., Gallouzi, I. E., and Gehring, K. (2007). Solution structure of YaeO, a Rho-specific inhibitor of transcription termination. *J. Biol. Chem.* **282**, 23348–23353.

Hegyi, H., Schad, E., and Tompa, P. (2007). Structural disorder promotes assembly of protein complexes. *BMC Struct. Biol.* **7**, 65.

Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H., and Goto, Y. (2002). Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange. *Nat. Struct. Biol.* **9**, 332–336.

Jacobi, A., Rossmann, R., and Bock, A. (1992). The hyp operon gene products are required for the maturation of catalytically active hydrogenase isoenzymes in *Escherichia coli*. *Arch. Microbiol.* **158**, 444–451.

Jain, N. U., Wyckoff, T. J., Raetz, C. R., and Prestegard, J. H. (2004). Rapid analysis of large protein-protein complexes using NMR-derived orientational constraints: The 95 kDa complex of LpxA with acyl carrier protein. *J. Mol. Biol.* **343**, 1379–1389.

Jancarik, J., Pufan, R., Hong, C., Kim, S. H., and Kim, R. (2004). Optimum solubility (OS) screening: An efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. *Acta Crystallogr. D* **60**, 1670–1673.

Janga, S. C., and Babu, M. M. (2009). Transcript stability in the protein interaction network of *Escherichia coli*. *Mol. Biosyst.* **5**, 154–162.

Janin, J., Henrick, K., Moult, J., Eyck, L. T., Sternberg, M. J., Vajda, S., Vakser, I., and Wodak, S. J. (2003). CAPRI: A critical assessment of predicted interactions. *Proteins* **52**, 2–9.

Jefferson, E. R., Walsh, T. P., Roberts, T. J., and Barton, G. J. (2007). SNAPPI-DB: A database and API of structures, interfaces and alignments for protein–protein interactions. *Nucleic Acids Res.* **35**, D580–D589.

Johnson, D. C., Dean, D. R., Smith, A. D., and Johnson, M. K. (2005). Structure, function, and formation of biological iron-sulfur clusters. *Annu. Rev. Biochem.* **74**, 247–281.

Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979.

Keseler, I. M., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S., Paulsen, I. T., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A. G., and Karp, P. D. (2009). EcoCyc: A comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res.* **37**, D464–D470.

Kim, O. B., Lux, S., and Unden, G. (2007a). Anaerobic growth of *Escherichia coli* on D-tartrate depends on the fumarate carrier DcuB and fumarase, rather than the L-tartrate carrier TtdT and L-tartrate dehydratase. *Arch. Microbiol.* **188**, 583–589.

Kim, S. M., Bowers, P. M., Pal, D., Strong, M., Terwilliger, T. C., Kaufmann, M., and Eisenberg, D. (2007b). Functional linkages can reveal protein complexes for structure determination. *Structure* **15**(9), 1079–1089. September 11, Ref Type: Abstract.

Kim, Y. G., Raunser, S., Munger, C., Wagner, J., Song, Y. L., Cygler, M., Walz, T., Oh, B. H., and Sacher, M. (2006). The architecture of the multisubunit TRAPP I complex suggests a model for vesicle tethering. *Cell* **127**, 817–830.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF Archive): Unique Resources for Biological Research. *DNA Res.* **12**, 291–299.

Klem, T. J., and Davisson, V. J. (1993). Imidazole glycerol phosphate synthase: The glutamine amidotransferase in histidine biosynthesis. *Biochemistry* **32**, 5177–5186.

Kobe, B., Guncar, G., Buchholz, R., Huber, T., Maco, B., Cowieson, N., Martin, J. L., Marfori, M., and Forwood, J. K. (2008). Crystallography and protein–protein interactions: Biological interfaces and crystal contacts. *Biochem. Soc. Trans.* **36**, 1438–1441.

Koch, M. H., Vachette, P., and Svergun, D. I. (2003). Small-angle scattering: A view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **36**, 147–227.

Kundrotas, P. J., and Alexov, E. (2007). PROTCOM: Searchable database of protein complexes enhanced with domain–domain structures. *Nucleic Acids Res.* **35**, D575–D579.

Lanman, J., and Prevelige, P. E. (2004). High-sensitivity mass spectrometry for imaging subunit interactions: Hydrogen/deuterium exchange. *Curr. Opin. Struct. Biol.* **14**, 181–188.

Lasserre, J. P., Beyne, E., Pyndiah, S., Lapaillerie, D., Claverol, S., and Bonneu, M. (2006). A complexomic study of *Escherichia coli* using two-dimensional blue native/SDS polyacrylamide gel electrophoresis. *Electrophoresis* **27**, 3306–3321.

Leavitt, S., and Freire, E. (2001). Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11**, 560–566.

Levy, E. D., Pereira-Leal, J. B., Chothia, C., and Teichmann, S. A. (2006). 3D complex: A structural classification of protein complexes. *PLoS. Comput. Biol.* **2**, e155.

Lin, C. C., Juan, H. F., Hsiang, J. T., Hwang, Y. C., Mori, H., and Huang, H. C. (2009). Essential core of protein–protein interaction network in *Escherichia coli*. *J. Proteome Res.* **8**, 1925–1931.

Linding, R., Russell, R. B., Neduva, V., and Gibson, T. J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**, 3701–3708.

Liu, G., Li, J., and Wong, L. (2008). Assessing and predicting protein interactions using both local and global network topological metrics. *Genome Inform.* **21**, 138–149.

Mandell, J. G., Baerga-Ortiz, A., Croy, C. H., Falick, A. M., and Komives, E. A. (2005). Application of amide proton exchange mass spectrometry for the study of protein-protein interactions. *Curr. Protoc. Protein Sci.* Chapter 20:Unit20.9.

Mandell, J. G., Falick, A. M., and Komives, E. A. (1998). Identification of protein-protein interfaces by decreased amide proton solvent accessibility. *Proc. Natl. Acad. Sci. USA* **95**, 14705–14710.

Masson, L., Mazza, A., and De, C. G. (2000). Determination of affinity and kinetic rate constants using surface plasmon resonance. *Methods Mol. Biol.* **145**, 189–201.

McDonnell, J. M. (2001). Surface plasmon resonance: Towards an understanding of the mechanisms of biological molecular recognition. *Curr. Opin. Chem. Biol.* **5**, 572–577.

Meyer, S., Scrima, A., Versées, W., and Wittinghofer, A. (2008). Crystal structures of the conserved tRNA-modifying enzyme GidA: Implications for its interaction with MnmE and substrate. *J. Mol. Biol.* **380**, 532–547.

Muhandiram, D. R., and Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. B* **103**, 203–216.

Nachin, L., El, H. M., Loiseau, L., Expert, D., and Barras, F. (2001). SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: The key role of SufC, an orphan ABC ATPase. *Mol. Microbiol.* **39**, 960–972.

Nakanishi, T., Miyazawa, M., Sakakura, M., Terasawa, H., Takahashi, H., and Shimada, I. (2002). Determination of the interface of a large protein complex by transferred cross-saturation measurements. *J. Mol. Biol.* **318**, 245–249.

Oishi, Y., Yunomura, S., Kawahashi, Y., Doi, N., Takashima, H., Baba, T., Mori, H., and Yanagawa, H. (2006). *Escherichia coli* proteome chips for detecting protein–protein interactions. *Proteomics* **6**, 6433–6436.

Outten, F. W., Wood, M. J., Munoz, F. M., and Storz, G. (2003). The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *Escherichia coli*. *J. Biol. Chem.* **278**, 45713–45719.

Page, R. (2008). Strategies for improving crystallization success rates. *Methods Mol. Biol.* **426**, 345–362.

Parrish, J. R., Yu, J., Liu, G., Hines, J. A., Chan, J. E., Mangiola, B. A., Zhang, H., Pacifico, S., Fotouhi, F., DiRita, V. J., Ideker, T., Andrews, P., and Finley, R. L., Jr. (2007). A proteome-wide protein interaction map for *Campylobacter jejuni*. *Genome Biol.* **8**, R130.

Patil, A., and Nakamura, H. (2005). Filtering high-throughput protein–protein interaction data using a combination of genomic features. *BMC Bioinformatics* **6**, 100.

Pichoff, S., Alibaud, L., Guedant, A., Castanie, M. P., and Bouche, J. P. (1998). An *Escherichia coli* gene (yaeO) suppresses temperature-sensitive mutations in

essential genes by modulating Rho-dependent transcription termination. *Mol. Microbiol.* **29**, 859–869.

Piliarik, M., Vaisocherova, H., and Homola, J. (2009). Surface plasmon resonance biosensing. *Methods Mol. Biol.* **503**, 65–88.

Platt, T. (1994). Rho and RNA: Models for recognition and response. *Mol. Microbiol.* **11**, 983–990.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005). FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438.

Prudencio, M., and Ubbink, M. (2004). Transient complexes of redox proteins: Structural and dynamic details from NMR studies. *J. Mol. Recognit.* **17**, 524–539.

Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285.

Radaev, S., Li, S., and Sun, P. D. (2006). A survey of protein-protein complex crystallizations. *Acta Crystallogr. D* **62**, 605–612.

Rain, J. C., Selig, L., De, R. H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., Chemama, Y., Labigne, A., and Legrain, P. (2001). The protein–protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211–215.

Rangarajan, E. S., Asinas, A., Proteau, A., Munger, C., Baardsnes, J., Iannuzzi, P., Matte, A., and Cygler, M. (2008). Structure of [NiFe] hydrogenase maturation protein HypE from *Escherichia coli* and its interaction with HypF. *J. Bacteriol.* **190**, 1447–1458.

Reaney, S. K., Begg, C., Bungard, S. J., and Guest, J. R. (1993). Identification of the L-tartrate dehydratase genes (ttdA and ttdB) of *Escherichia coli* and evolutionary relationship with the class I fumarase genes. *J. Gen. Microbiol.* **139**, 1523–1530.

Reissmann, S., Hochleitner, E., Wang, H., Paschos, A., Lottspeich, F., Glass, R. S., and Bock, A. (2003). Taming of a poison: Biosynthesis of the NiFe-hydrogenase cyanide ligands. *Science* **299**, 1067–1070.

Remy, I., Campbell-Valois, F. X., and Michnick, S. W. (2007). Detection of protein–protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat. Protoc.* **2**, 2120–2125.

Richardson, J. P. (1996). Structural organization of transcription termination factor Rho. *J. Biol. Chem.* **271**, 1251–1254.

Richardson, J. P., and Greenblatt, J. (1996). "*Escherichia Coli* and *Salmonella Thyphimurium*: Cellular and Molecular Biology," pp. 822–848. Ed FC Neidhardt. Washington, DC: American Society for Microbiology Press.

Rutherford, S. T., Villers, C. L., Lee, J. H., Ross, W., and Gourse, R. L. (2009). Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA. *Genes Dev.* **23**, 236–248.

Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y., and Tabata, S. (2007). A large-scale protein–protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* **14**, 207–216.

Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U., and Bussow, K. (2007). Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* **35**, e43.

Selleck, W., and Tan, S. (2008). Recombinant protein complex expression in *E. coli*. *Curr. Protoc. Protein Sci.* Chapter 5:Unit 5.21.

Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z., and Dunker, A. K. (2007). DisProt: The database of disordered proteins. *Nucleic Acids Res.* **35**, D786–D793.

Steitz, T. A. (2008). A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell. Biol.* **9**, 242–253.

Stenberg, F., Chovanec, P., Maslen, S. L., Robinson, C. V., Ilag, L. L., von, H. G., and Daley, D. O. (2005). Protein complexes of the *Escherichia coli* cell envelope. *J. Biol. Chem.* **280**, 34409–34419.

Su, C., Peregrin-Alvarez, J. M., Butland, G., Phanse, S., Fong, V., Emili, A., and Parkinson, J. (2008). Bacteriome.org an integrated protein interaction database for *E. coli*. *Nucleic Acids Res.* **36**, D632–D636.

Svergun, D., Barberato, C., and Koch, M. H.J. (1995). CRYSOL – A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystal.* **28**, 768–773.

Svergun, D. I., Petoukhov, M. V., and Koch, M. H.J. (2001). Determination of domain structure of proteins from x-ray solution scattering. *Biophys. J.* **80**(6), 2946–2953. June 1, Ref Type: Abstract

Takahashi, H., Nakanishi, T., Kami, K., Arata, Y., and Shimada, I. (2000). A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat. Struct. Biol.* **7**, 220–223.

Takeuchi, K., and Wagner, G. (2006). NMR studies of protein interactions. *Curr. Opin. Struct. Biol.* **16**, 109–117.

Terradot, L., Durnell, N., Li, M., Li, M., Ory, J., Labigne, A., Legrain, P., Colland, F., and Waksman, G. (2004). Biochemical characterization of protein complexes from the *Helicobacter pylori* protein interaction map: Strategies for complex formation and evidence for novel interactions within type IV secretion systems. *Mol. Cell Proteomics* **3**, 809–819.

Tolia, N. H., and Joshua-Tor, L. (2006). Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods* **3**, 55–64.

Tong, J., and McIntosh, T. J. (2004). Structure of supported bilayers composed of lipopolysaccharides and bacterial phospholipids: Raft formation and implications for bacterial resistance. *Biophys. J.* **86**, 3759–3771.

Vajda, S., and Kozakov, D. (2009). Convergence and combination of methods in protein-protein docking. *Curr. Opin. Struct. Biol.* **19**, 164–170.

Vaynberg, J., and Qin, J. (2006). Weak protein-protein interactions as probed by NMR spectroscopy. *Trends Biotechnol.* **24**, 22–27.

Watanabe, S., Matsumi, R., Arai, T., Atomi, H., Imanaka, T., and Miki, K. (2007). Crystal structures of [NiFe] hydrogenase maturation proteins HypC, HypD, and HypE: Insights into cyanation reaction by thiol redox signaling. *Mol. Cell* **27**, 29–40.

Wittig, I., and Schagger, H. (2008). Features and applications of blue-native and clear-native electrophoresis. *Proteomics* **8**, 3974–3990.

Wojcik, J., Boneca, I. G., and Legrain, P. (2002). Prediction, assessment and validation of protein interaction maps in bacteria. *J. Mol. Biol.* **323**, 763–770.

Yellaboina, S., Goyal, K., and Mande, S. C. (2007). Inferring genome-wide functional linkages in *E. coli* by combining improved genome context methods: Comparison with high-throughput experimental data. *Genome Res.* **17**, 527–535.

Yew, W. S., Fedorov, A. A., Fedorov, E. V., Wood, B. M., Almo, S. C., and Gerlt, J. A. (2006). Evolution of enzymatic activities in the enolase superfamily: D-tartrate dehydratase from *Bradyrhizobium japonicum*. *Biochemistry* **45**, 14598–14608.

Yim, L., Moukadiri, I., Björk, G. R., and Armengod, M. E. (2006). Further insights into the tRNA modification process controlled by proteins MnmE and GidA of *Escherichia coli*. *Nucleic Acids Res.* **34**, 5892–5905.

Yu, H., Chan, Y. L., and Wool, I. G. (2009). The identification of the determinants of the cyclic, sequential binding of elongation factors tu and g to the ribosome. *J. Mol. Biol.* **386**, 802–813.

Chapter No: 1

Query No	Contents
AU1	Please suggest whether "range ~ K_d 0.1–10 μM" can be changed to "range of approximately $K_d = 0.1–10$ μM" in the sentence "A specific limitation of ITC . . ."
AU2	Please confirm the deletion of "for example" in the sentence "For example, in . . ."
AU3	Please confirm the change of "K_D" to "K_d" in the sentence "Since fast exchanging . . ."
AU4	(Comment by reviewer): NMR technique in a mass spectrometry section? The section should be renamed "Amide proton/deuterium exchange," in order to accommodate both NMR and mass spec.
AU5	Please confirm the edit made to the sentence "In order to obtain a structure . . .".
AU6	Please provide better quality image for figure 4B.