Studying protein complexes by the yeast two-hybrid system

Seesandra V. Rajagopala, Patricia Sikorski, J. Harry Caufield, Andrey Tovchigrechko, Peter Uetz

1. Introduction

Proteins act by either interacting with other proteins or small molecules such as sugars. In fact, many enzymes that interact with their low molecular weight substrates also form protein complexes, such as fatty acid synthase, a complex of 12 chains [1]. In the past 15 years high-throughput methods have been developed to map both protein–protein interactions on a large scale, using either the yeast two-hybrid system for binary interactions [2–4] or affinity purification and mass spectrometry for protein complexes [5–8]. In principle, a two-hybrid screen using all the proteins of a complex should yield all the binary interactions within that complex, but this is rarely the case; in most cases only a few interactions are discovered [9,10]. On the other hand, the two-hybrid system has a certain preference for transient interactions that are lost during complex purification because of the necessary washing steps [4]. In addition, two-hybrid screening data derived from genome-scale projects is not complete or/and comprehensive. Thus, little overlap is observed between the protein interaction datasets generated by protein-complex purification and two-hybrid studies [4].

In order to infer direct interactions from complex purification data, either the matrix or spoke model has been applied to lists of co-purified proteins. More recent protein complex purification studies used the socio-affinity index (SAI) to infer the direct interactions between complex members [5–8]. A related strategy uses data from complex purification data sets to identify interacting proteins [11]. These strategies are based on the observation that certain pairs of proteins are more frequently found in multiple purifications than others, and are thus predicted to be closer or even directly associated in the complex. Similarly, solutions have been presented to computationally identify complexes from Y2H interaction networks [12–15].

While these experimental and computational attempts to map protein–protein interactions have produced a massive amount of data, structural analysis of complexes and interacting proteins has lagged behind. One of the goals of PPI studies is thus to identify complexes that are amenable for crystallization. Often, several strategies need to be combined to reconstruct the topology of complexes that cannot be crystallized, including proteomics, cryo-electron microscopy and others.

Recently, we have developed multiple variants of the yeast two-hybrid system and shown that different two-hybrid systems detect markedly different subsets of interactions in the same interactome [16]. Ten different configurations of bait-prey fusions were required to detect up to 67% of a set of gold-standard interactions, whereas individual vector pairs detected only 25% on average [17].

Here we describe and review a similar strategy, using yeast two-hybrid assays to map the interactions within several complexes. In addition, we analyze several well-characterized protein...
complexes, ranging from ribonucleotide reductase (four subunits) to spliceosomes (>200 proteins), and compare structural data to published Y2H data. Our data shows that a majority of interactions in a complex can be identified by systematic Y2H screening and that Y2H assays often detect subcomplexes within a larger complex that may be amenable to crystallography while the whole complex may be not crystalizable.

2. Materials and methods

2.1. Yeast two-hybrid (Y2H) assays

All full-length ORF clones of two protein complexes (Tables 1 and 2) were selected from the *Escherichia coli* ORFeome collection [18]. The entry clones were transferred into prey and bait yeast two-hybrid expression vectors using Gateway LR reactions (namely pGADT7g and pGBKT7g for N-terminal fusions and pGBKCG-att for C-terminal fusions; the latter vector having an in-frame start codon for ORFs that do not have one) [19]. After bacterial transformation, miniprep plasmid DNA of all two-hybrid prey and bait clones were transformed into yeast two-hybrid strains *MATa Y187* (prey clones) and *MATα AH109* (bait clones). Yeast two-hybrid matrix screening was conducted as described in [20]. In brief, a yeast strain expressing a single protein as a bait fusion was mated to individual preys of all other protein complex members (Fig. 1A and B). After mating, the colonies were transferred to the Y2H protein–protein interaction selective medium, and the interacting protein pairs were identified by the resulting positive yeast colony. The positive interactions show a clear colony growth at a certain level of 3-Amino-1,2,4-triazole, whereas no growth was usually seen in the negative control (auto-activation), i.e. the bait mated with the empty prey vector strain (Fig. 1B).

Three Dimensional Modeling: Homology models of the VZV R1 (ORF19) and R2 (ORF18) proteins were produced with I-TASSER server [21] and superimposed onto the 2BQ1 asymmetric holostructure from PDB (http://www.rcsb.org). The final model was rendered in Pymol (http://www.pymol.org).

### Table 1

DNA Polymerase III complex [7]. See Fig. 1 for interactions.

<table>
<thead>
<tr>
<th>Name</th>
<th>Locus</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AsnB</td>
<td>JW0660</td>
<td>Asparagin synthetase B</td>
</tr>
<tr>
<td>DnaE</td>
<td>JW0179</td>
<td>DNA polymerase III alpha subunit</td>
</tr>
<tr>
<td>DnaX</td>
<td>JW0459</td>
<td>DNA polymerase III/ DNA elongation factor III, tau and gamma subunits</td>
</tr>
<tr>
<td>HolC</td>
<td>JW4216</td>
<td>DNA polymerase III, chi subunit</td>
</tr>
<tr>
<td>HolD</td>
<td>JW1334</td>
<td>DNA polymerase III, psi subunit</td>
</tr>
<tr>
<td>HolE</td>
<td>JW1831</td>
<td>DNA polymerase III, theta subunit</td>
</tr>
<tr>
<td>TrmD</td>
<td>JW2588</td>
<td>tRNA (guanine-1-) -methyltransferase</td>
</tr>
</tbody>
</table>

### Table 2

Protein–protein interactions of the *Escherichia coli* DNA polymerase III protein complex. Interactions between protein A and protein B have been curated from the literature ("References") or are the results of our own Y2H studies (vectors used are given after "Y2H:").

<table>
<thead>
<tr>
<th>Protein_A</th>
<th>Protein_B</th>
<th>Interaction detection method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DnaE</td>
<td>AsnB</td>
<td>Tandem affinity purification</td>
<td>[5]</td>
</tr>
<tr>
<td>DnaE</td>
<td>DnaE</td>
<td>Biochemical</td>
<td>[61]</td>
</tr>
<tr>
<td>HolD</td>
<td>DnaE</td>
<td>Tandem affinity purification</td>
<td>[5]</td>
</tr>
<tr>
<td>HolC</td>
<td>DnaX</td>
<td>Molecular sieving and biochemical</td>
<td>[62,63]</td>
</tr>
<tr>
<td>HolD</td>
<td>DnaX</td>
<td>Molecular sieving</td>
<td>[62,64]</td>
</tr>
<tr>
<td>HolD</td>
<td>HolC</td>
<td>X-ray crystallography, molecular sieving and biochemical</td>
<td>[54,54]</td>
</tr>
<tr>
<td>DnaX</td>
<td>DnaX</td>
<td>Y2H: GBG-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>HolD</td>
<td>DnaX</td>
<td>Y2H: GBG-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>HolD</td>
<td>HolC</td>
<td>Y2H: GBG-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>HolC</td>
<td>HolD</td>
<td>Y2H: GBG-GAD and GBKC-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>TrmD</td>
<td>TrmD</td>
<td>Y2H: GBG-GAD and GBKC-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>AsnB</td>
<td>AsnB</td>
<td>Y2H: GBKC-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>DnaE</td>
<td>DnaX</td>
<td>Y2H: GBKC-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>HolC</td>
<td>DnaX</td>
<td>Y2H: GBKC-GAD</td>
<td>This study</td>
</tr>
<tr>
<td>HolC</td>
<td>HolC</td>
<td>Y2H: GBKC-GAD</td>
<td>This study</td>
</tr>
</tbody>
</table>

3. Results and discussion

3.1. Bacterial complexes

3.1.1. The *E. coli* DNA polymerase III complex

In order to validate how well multiple variants of the two-hybrid system work for mapping the topology of protein complexes, we have selected the *E. coli* DNA polymerase III complex as an example of a well-characterized protein complex (protein complex 42 in [7]). Hu et al. affinity-purified this complex and identified seven subunits, namely, AsnB, DnaE, DnaX, HolC, HolD, HolE, and TrmD. We did not include other proteins known to be more loosely associated with DNA polymerase III such as the epsilon subunit (DnaQ) which catalyzes the 3′ to 5′ proofreading exonuclease activity of the holoenzyme.

*E. coli* DNA polymerase contains two sub-complexes: the catalytic polymerase/exonuclease sub-complex (with alpha, beta, delta and epsilon subunits), plus the DNAx complex, a heptamer that includes the tau and gamma products of the dnaX gene and confers structural asymmetry that allows the polymerase to replicate both leading and lagging strands. We have curated all the known protein–protein interactions between co-purified complex members from several studies (Fig. 1C and Table 2).

In order to validate whether we can capture all the known interactions of the *E. coli* DNA polymerase III proteins, we subjected its proteins to an Y2H matrix screening. The yeast two-hybrid assay was able to detect 70% known interactions (seven out of 10 interactions, including homo-dimers) and was able to connect four protein (out of seven proteins) in the complex compared to five proteins connected by literature data (Fig. 1C). Some of the known interactions are not reproduced here, for example the theta (HolE) subunit of DNA polymerase III binds to the epsilon subunit (DnaQ) but not to the alpha subunit (DnaE) [22]. This binding appears to enhance the interaction between alpha (DnaE) and epsilon (DnaQ) as well as slightly stimulating epsilon activity [23]. Since DnaQ is not a member of the protein complex described here, the interaction between HolE → DnaQ → DnaE has not been tested. These results demonstrate that Y2H screening can detect majority of the direct interactions within a complex and aid the mapping of its topology. The MntR complex is a largely uncharacterized protein complex that has been identified as "complex 34" by AP/MS [7]. This complex consists of eight proteins of which five proteins have been characterized and three uncharacterized proteins (Table 3, Fig. 1D). Even though the complex contains several well-characterized proteins the direct interaction and functional associations among the complex members are unknown.

The Y2H screening of all the members of this complex identified seven interactions (plus four homo-dimers) and based on these interactions we were able to precisely map the topology of the complex (Fig. 1B and D), confirming that these proteins actually form a complex. These protein–protein interactions should help to characterize the function of unchartered proteins in the complex. For example, the interaction between the transcriptional regulators MntR and PerR supports the predicted DNA-binding and
transcriptional activity of the latter. While PerR has been studied in *Bacillus subtilis*, its ortholog in *E. coli* is poorly understood. PerR and related members of the LysR family have been shown to interact with other members of the family to form heterodimers, but the physiological significance of this is unknown [24]. In order to study the function of these complexes in more detail, the topology of interactions suggests several strategies to crystallize this complex, e.g. either as dimers (YjhC–YjhC), trimers (MntR–HycG–MetN or MntR–MetN–YjhC), tetramer (MntR–HycG–MetN–YjhC), or as hetero-hexamer. Such crystallization experiments are now under way.

3.2. Previously published complexes

In addition to the new interactions in the previous section, we have also re-visited the literature on several protein complexes that have been studied by Y2H methods. They cover sizes ranging from ribonucleotide reductase, a small protein complex of four subunits, to the spliceosome, one of the largest complexes in living systems with hundreds of proteins. These complexes are discussed in the order of increasing complexity.
3.3. Ribonucleotide reductase

Ribonucleotide reductases (RNRs) convert ribonucleotides to deoxyribonucleotides and thus provide the raw materials for DNA synthesis. All living systems that are based on DNA as their genetic material, either encode RNRs, or must obtain their dNTPs from an outside source, as several parasites and endosymbionts do [25]. Most eukaryotes and some of their viruses encode so-called type I RNRs that are characterized by a $\alpha_2\beta_2$ quaternary structure [26].

RNRs are interesting complexes for several reasons (reviewed in [25–27]). First, they all appear to be homologous, yet show a variety of quaternary structures and enzymatic mechanisms. For instance, class I RNRs form $\alpha_2\beta_2$ quaternary structures while class II enzymes are either $\alpha$ monomers or $\alpha_2$ dimers. Second, different species use different cofactors (e.g. thioredoxin vs. formate), are aerobic or anaerobic, and have different modes of interaction. Third, they are highly dynamic complexes that undergo conformational changes during a reaction cycle, which also affects their interactions. Fourth and last, RNRs and their subunit interactions are affected by allosteric regulators such as dGTP.

In the course of our studies on human herpesviral interactomes [28,29], we have analyzed RNRs of several viruses. However, only the RNR of Varicella Zoster Virus (VZV) was analyzed with multiple Y2H vectors (Fig. 2, [16]) which allowed us to draw some more general conclusions about interactions that can be detected within protein complexes. First, under favorable conditions, majority of interactions in a complex can be detected by Y2H assays. In the case of VZV RNR, all interactions among the four RNR subunits were found: R1–R1, R1–R2, and R2–R2. Second, full-length proteins not necessarily work in such assays. Here, the full-length R2 protein neither interacted with R1 nor with itself. However, fragments clearly worked in multiple orientations. As is shown in Fig. 2B, N- and C-terminal fragments of R2 interacted in a homo-dimeric fashion (ORF18C–ORF18C), as pseudo-homo-dimers (ORF18N–ORF18C), as well as true hetero-dimers (ORF18C–ORF19). Third, and finally, the N- and C-terminal fusions of the Gal4 DNA-binding and activation domains had very distinct

![Diagram of VZV RNR complex](image)

**Fig. 2.** The ribonucleotide reductase (RNR) complex of Varicella Zoster virus (VZV). (A) Model of the VZV RNR complex based on crystal structures of *Salmonella typhimurium* RNR (2BQ1 [32]). VZV ORF18 = R2 (Uniprot P09247), ORF19 = R1 (P09248) with chains in each homo-dimer numbered as I and II. R1 chain II and R2 chain II are colored in rainbow from N-terminal (blue) to C-terminal (red), with terminals labeled as NT and CT, respectively. The C-terminal of the R2 (II) chain was replaced with the corresponding peptide taken from 2BQ1 (thick red tube on the picture) where it is in contact with the groove in the R1 (I). The remaining part of R2 (II) CT is hidden because it is disordered in 2BQ1. The C-terminal of R1 (I) and the N-terminal of R2 (II) are shown as thick tubes in regions disordered in 2BQ1. The remaining terminals are behind the structures of the symmetric homo-dimers in this view and thus not labeled. (B) Interactions among RNR subunits detected in Y2H assays. The Gal4 DNA-binding (DBD) and activation domains (AD) are shown as green boxes that are either fused to the N or C-terminus of the RNR proteins. ORF18 was tested as full-length protein and as N- or C-terminal fragments (and each fragment was tested as N- and C-terminal fusion to DBD and AD). Y2H data in (B) from [16].
patterns of interaction. For instance, while an N-terminal fusion of ORF18C interacted with almost all other proteins, both N- and C-terminal fusions of ORF18N interacted very specifically only with ORF18C.

These observations are largely supported by the crystal structure of the RNR holocomplex (Fig. 2A) [30,31], even though the biological complex appears to be a dynamic assembly of proteins in which the R2 subunits moves and rotates [32]. This fact has impaired crystallization of the biological complex and it is still not entirely clear how the subunits move during a catalytic cycle in vivo.

Our homology model superimposed on the available asymmetric template (Fig. 2A) shows C- and N-terminals either located completely away from the interface areas or having likely disordered/flexible ends, thus leaving enough room for accommodating the Y2H fusion constructs in all cases.

3.4. The proteasome

The proteasome is a complex assembly of proteins found within all eukaryotic cells and some bacterial species. In its fully-assembled state, this protein complex recycles peptides within the cell by degrading misfolded or otherwise unnecessary proteins. Each full 26S proteolytic complex is formed by a 20S core particle (CP) and at least one regulatory particle (RP), including six AAA-ATPase subunits and thirteen non-ATPase subunits. In eukaryotes, each of these 26S proteasome particles may be found in the cytoplasm and the nucleoplasm [33]. Several sets of yeast two-hybrid analyses have been used alone and in concert with other methods to clarify the protein interactions required for assembly of a functional proteasome. Work by Cagney et al. [34] employed a yeast two-hybrid approach to search for interactions between 31 known proteasome proteins and an array of nearly 6000 different Saccharomyces cerevisiae proteins. This genome-wide screen revealed 55 potential protein–protein interactions, more than a third of which involved only proteasome components rather than non-proteasomal proteins. The specific pairs of interactions found in these screens demonstrated how many proteins of the 19S and 20S subcomplexes specifically interact with other proteins within those subcomplexes (Fig. 3A and B). One interaction between a 20S α ring protein, Pre8, and a 20S β ring protein, Pup1, was observed, confirming structural arrangements seen in the 20S crystal structure [35]. Some interactions did occur between subunits, as comparisons with the crystal structure of the 20S core showed. Out of 14 interacting protein pairs in the 20S subcomplex, just 3 involved proteins...
not predicted to be neighbors in the crystal structure. As Cagney et al. studied yeast proteasomal proteins in an Y2H system; it is quite likely that endogenous proteasomal proteins bridged the two hybrid proteins. This background may have contributed to artificial interactions not seen in the 20S crystal structure. Use of the Y2H system may therefore be better suited to the study of heterologous protein pairs from sources other than yeast.

Work by Davy et al. [36] used the Y2H system to generate interaction data for 30 proteasome subunits from Caenorhabditis elegans. Each of the proteins assayed in this study are orthogonal to those of S. cerevisiae. Though this study found many interactions correlating with those seen in with yeast proteasome proteins, a number of new interactions were also observed (Fig. 3). Among these are four interactions between 19S subunit proteins and three 20S α subunits.

An independent study of the human proteasome [37] also revealed protein–protein interactions by Y2H, finding 114 potential interacting pairs. Many of these interactions were found to be structurally relevant. Interactions between the α and β rings of the 20S core were observed, confirming observations by Cagney et al. (Fig. 3B, Supplementary Table 1). Subsequent studies have confirmed many of the interactions revealed by Y2H assays using other methods. A mass spectrometry-based approach [38] generated 64 potential proteasome-associated protein–protein interactions beyond those seen between proteasome subunit proteins.

Unfortunately, the fragile nature of the 26S proteasome has rendered it difficult to crystalize for extensive X-ray crystallographic analysis. The 20S core particle crystal structure prepared by Groll et al. [35] provides an example of many potential protein interactions on the basis of the barrel-like structure of this subunit, but it provides no evidence for protein interactions beyond those of the core particle. Combined data from crystallography, cryoelectron microscopy, and protein co-purifications have been used to clarify proteasomal structural organization predicted by Y2H results [39–41]. One such model prepared by Lasker et al. displays the overall architecture of the 26S proteasome holoenzyme. While the overall structure is based on cryo-EM mapping and molecular modeling (including a model prepared by Förster et al. [42]), the specific locations of each protein are based on crystal structures, residue-specific lysine crosslinks, and known protein–protein interactions, including those found though Y2H screens.

In summary, the three Y2H studies found a total of 183 PPIs, while structural studies revealed a total of 38 PPIs. No Y2H study found more than 40% of the interactions assumed to take place in the proteasome, but all three studies together found 79% of all interactions because each study found a different subset (Supplementary Table 1). This is certainly due to the fact that proteins from different species as well as different Y2H systems were used.

Clearly, Y2H screens produced a substantial number of false positives in these studies, (as shown by interactions not seen in the proteasome structure), but some of them may be truly physiological interactions not seen in the 20S crystal structure. Use of the Y2H system may therefore be better suited to the study of heterologous protein pairs from sources other than yeast.

Work by Davy et al. [36] used the Y2H system to generate interaction data for 30 proteasome subunits from Caenorhabditis elegans. Each of the proteins assayed in this study are orthogonal to those of S. cerevisiae. Though this study found many interactions correlating with those seen in with yeast proteasome proteins, a number of new interactions were also observed (Fig. 3). Among these are four interactions between 19S subunit proteins and three 20S α subunits.

An independent study of the human proteasome [37] also revealed protein–protein interactions by Y2H, finding 114 potential interacting pairs. Many of these interactions were found to be structurally relevant. Interactions between the α and β rings of the 20S core were observed, confirming observations by Cagney et al. (Fig. 3B, Supplementary Table 1). Subsequent studies have confirmed many of the interactions revealed by Y2H assays using other methods. A mass spectrometry-based approach [38] generated 64 potential proteasome-associated protein–protein interactions beyond those seen between proteasome subunit proteins.

Unfortunately, the fragile nature of the 26S proteasome has rendered it difficult to crystalize for extensive X-ray crystallographic analysis. The 20S core particle crystal structure prepared by Groll et al. [35] provides an example of many potential protein interactions on the basis of the barrel-like structure of this subunit, but it provides no evidence for protein interactions beyond those of the core particle. Combined data from crystallography, cryoelectron microscopy, and protein co-purifications have been used to clarify proteasomal structural organization predicted by Y2H results [39–41]. One such model prepared by Lasker et al. displays the overall architecture of the 26S proteasome holoenzyme. While the overall structure is based on cryo-EM mapping and molecular modeling (including a model prepared by Förster et al. [42]), the specific locations of each protein are based on crystal structures, residue-specific lysine crosslinks, and known protein–protein interactions, including those found though Y2H screens.

In summary, the three Y2H studies found a total of 183 PPIs, while structural studies revealed a total of 38 PPIs. No Y2H study found more than 40% of the interactions assumed to take place in the proteasome, but all three studies together found 79% of all interactions because each study found a different subset (Supplementary Table 1). This is certainly due to the fact that proteins from different species as well as different Y2H systems were used.

Clearly, Y2H screens produced a substantial number of false positives in these studies, (as shown by interactions not seen in the proteasome structure), but some of them may be truly physiological interactions (i.e. taking place in vivo), given that the α and β subunits of the proteasome are closely related proteins that probably interact in more than the canonical combinations even within the assembled proteasome (Fig. 3).

Notably, cross-linking studies found a total of 32 interactions, of which nine (32%) were seen in the structure. However, these studies may have not been as comprehensive as the Y2H screens, but it is remarkable that they also produced a number of false positives that is similar to those found in Y2H screens.

3.5. Phage Lambda

Bacteriophage lambda virion consists of ~14 different proteins and a total of ~1075 subunits. After its discovery in 1960 hundreds of studies have revealed its structure in great detail and we know now that its subunits are connected by at least a dozen different protein–protein interactions (Fig. 4A and B). Interestingly, not even in this extremely well-studied phage all interactions are known for sure. Rajagopala et al. [43] therefore curated the literature and compiled a list of 33 published PPIs among lambda proteins (note that this list includes PPIs among proteins that are not in the particle). This set of 33 known interactions is considered as a “gold-standard” set here. These authors then tested all possible protein pairs encoded by the lambda genome for interactions using a matrix-based Y2H screen that employed 6 different Y2H vectors (Fig. 4). Several lessons can be learned from this analysis. First, the screen detected more than half of all previously known interactions (including 4 interactions among regulatory proteins not present in the virion). Rajagopala et al. speculate that the remaining interactions were not detected because of the lack of chaperones, assembly factors, post-translational modifications, or other effects. Second, each vector pair detected a certain fraction of the known “gold-standard” set of PPIs; which interaction was detected depended primarily on the vector pair used (Fig. 4C). For instance, pGBK7/pGADC produced the largest absolute number of “gold-standard” interactions although the pDEST vectors produced the largest fraction of gold-standard interactions. 11 out of 16 interactions were detected only by one vector pair, namely pGBK7/gPADCg (5 PPIs), pDEST (4), and pGBK7/pGADCg (2). The other 5 PPIs were detected with multiple vectors each. Interestingly, only 4 of the gold-standard PPIs were detected exclusively with N-terminal fusions, the system with which the vast majority of all Y2H screens are carried out. Two gold-standard PPIs were detected exclusively with C-terminal fusions. The majority was detected with either NC fusions or multiple vectors.

3.6. The spliceosome

The spliceosome is probably the largest protein complex known, at least in terms of complexity, containing more than 200 different proteins as well as multiple RNAs [44]. It is thus much bigger than the ribosome. Although spliceosomal proteins assemble into smaller complexes, such as the U1, U2, U4/6 and U5 snRNPs, they transiently associate with each other during the process of splicing, which makes them a “functional” complex. Given its complexity, the spliceosome has not been crystalized and there are only certain subcomplexes whose structure is reasonably well-known, such as the U1 snRNP [45]. Hegele et al. [46] recently presented a systematic analysis of protein–protein interactions among 244 spliceosomal proteins, of which 141 proteins are classified as “core proteins”, based on their high abundance. Initially, Hegele et al. cloned 244 proteins known to be associated with the human spliceosome into yeast two-hybrid bait and prey vectors and tested them in a pairwise fashion for interactions. This study found a total of 632 interactions among 196 of the 244 proteins. Notably, 390 interactions were among non-core and 242 interactions were found among core proteins. Importantly, this study also curated 311 binary interactions previously published in 201 papers. Of these 311 PPIs, 72 were reproduced by Hegele’s Y2H analysis. While this number corresponds to only 23% of the published PPIs, the fraction of reproduced PPIs among the core proteins was 41% (43 PPIs). Reproducing only a quarter of all published interactions certainly have the same reasons as discussed above: first, only one vector pair was used in the Y2H analysis. Second, spliceosomal interactions may depend on both RNA support as well as ternary interactions. Third, many interactions are dynamic and thus rather weak, and may require additional factors such as assembly proteins or other catalysts [47]. The Hegele paper unfortunately did not attempt to integrate their data into structural models of the spliceosomal subunits, so it remains unclear to what extent these interactions facilitate a structural understanding of
the spliceosome. However, many interactions suggested local subcomplexes that can be subjected to crystallography and other analyses.

A good example how Y2H screens can dissect the interactions with a large complex is an analysis of the U1 snRNP. This subcomplex of the spliceosome consists of 10 U1 proteins and a ring of seven Sm proteins in yeast. Several groups have solved the overall structure of the complex by cryo-EM analysis and crystallography [48,49]. However, their model of the human U1 snRNP contains only three U1 proteins and the Sm ring (7 proteins). It remains unclear where the remaining proteins are that are clearly parts of the yeast U1 snRNP. Ester & Uetz [50] showed that a subcomplex of 3 proteins (Snu71, Prp40, and Luc7) can be detected within the yeast U1 snRNP and these authors mapped their interaction domains. It is hoped that such subcomplexes are stable enough for crystallization, so that their atomic structure can be resolved eventually.

4. Conclusions

From the data discussed here, it becomes clear that the Y2H system can contribute significantly to the understanding of protein complexes. However, there is also much room for improvements. First, multiple vectors need to be used routinely to achieve maximum coverage. In addition, the existing vectors can be certainly further improved. Notably, vector variations as mentioned here have not even been used for alternative Y2H systems such as the split-ubiquitin system or protein fragment complementation assays. Second, and similar to crystallography, it is likely that comparative studies will shed much light on protein complexes. For instance, while crystallization of *E. coli* and eukaryotic ribosomes was unsuccessful for many decades, the ribosomes (or ribosomal subunits) of other species could be crystallized more easily. Similarly, it is quite likely that many interactions may be easier to detect with proteins from non-model organisms although this issue has not been systematically studied on a larger (experimental) scale [51]. That said, it will be interesting to see how complexes and their interactions look in different, distantly related species. There is certainly much to be discovered by comparative analysis using improved methodology.

Acknowledgements

SVR and PU were supported by NIH Grant R01GM079710; AT was supported by NSF Grant 1048199.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ymeth.2012.07.015.

References
