A Computationally Guided Protein-Interaction Screen Uncovers Coiled-Coil Interactions Involved in Vesicular Trafficking

Hong Zhang1,2†, Jinan Chen1,2†, Yongqiang Wang1†, Lin Peng1, Xiaoxi Dong1,2, Yi Lu1,2, Amy E. Keating3* and Taijiao Jiang1*

1National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
2Graduate School of the Chinese Academy of Sciences, Beijing 100080, China
3Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Received 3 February 2009; received in revised form 15 June 2009; accepted 2 July 2009
Available online 8 July 2009

Mapping protein–protein interactions at a domain or motif level can provide structural annotation of the interactome. The α-helical coiled coil is among the most common protein-interaction motifs, and proteins predicted to contain coiled coils participate in diverse biological processes. Here, we introduce a combined computational/experimental screening strategy that we used to uncover coiled-coil interactions among proteins involved in vesicular trafficking in Saccharomyces cerevisiae. A number of coiled-coil complexes have already been identified and reported to play important roles in this important biological process. We identify additional examples of coiled coils that can form physical associations. The computational strategy used to prioritize coiled-coil candidates for testing dramatically improved the efficiency of discovery in a large experimental screen. As assessed by comprehensive yeast two-hybrid assays, computational prefiltering retained 90% of positive interacting pairs and eliminated >60% of negatives from a set of interaction candidates. The coiled-coil-mediated interaction network elucidated using the combined computational/experimental approach comprises 80 coiled-coil associations between 58 protein pairs, among which 21 protein interactions have not been previously reported in interaction databases and 26 interactions were previously known at the protein level but have now been localized to the coiled-coil motif. The coiled-coil-mediated interactions were specific rather than promiscuous, and many interactions could be recapitulated in a green fluorescent protein complementation assay. Our method provides an efficient route to discovering new coiled-coil interactions and uncovers a number of associations that may have functional significance for vesicular trafficking.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: vesicular trafficking; coiled coils; yeast two-hybrid; computational prediction; protein interactions

Introduction

Mapping which proteins interact in an organism, and how those interactions occur, is essential for understanding the systems-level organization of cellular structure, biological information processing and molecular mechanism. Numerous initiatives to map protein–protein interactions at the level of entire proteomes, particularly for the yeast Saccharomyces cerevisiae, have begun to provide a view of the organization of the interactome.1–8 Discovery of new interactions is inefficient using all-versus-all screens, however, because most proteins do not interact. Compounded with high false-negative rates, this means that only ~50% of protein interactions are
estimated to have been detected in yeast, following 6 years of intensive effort using many approaches. Furthermore, most large-scale studies do not provide information about how proteins interact because most proteome-wide assays are done using full-length gene products. Only a few studies have been carried out using protein fragments that localize interactions to a domain, subdomain or motif.10,11 Computational techniques have been developed to identify the regions of proteins that are responsible for interactions.12 These are highly promising, but both the quantity and the quality of predictions remain low for now.

An alternative approach to unbiased high-throughput screening using full-length proteins is a strategy that is computationally and structurally focused from the outset. For example, several laboratories have targeted high-throughput studies at protein interactions involving conserved domains and/or interactions that occur among proteins of related function.11–16 The hit rates in these types of tests are higher than those for proteome-wide systematic studies. Assaying protein interactions at the domain level also provides a variety of technical advantages. Isolated, well-folded domains are less prone to aggregation and non-specific interaction, and interaction sites that may be masked in full-length proteins can be exposed to enable association with protein, peptide or small-molecular partners. Finally, analysis at the domain level immediately localizes an interaction to a subregion of the full protein, which can be useful for mechanistic analyses or the development of protein-interaction inhibitors.

The coiled-coil interaction motif is particularly suitable for this type of study. Coiled coils have a simple structure consisting of two or more α-helices that associate in a supercoiled bundle. The motif is ubiquitous in the proteome. Its occurrence can be predicted with reasonable reliability using a variety of algorithms, and roughly 10% of proteins are estimated to contain coiled-coil regions.17 An advantage when seeking to identify coiled-coil-mediated interactions is that the partners of coiled-coil helices are, by definition, other coiled-coil helices. An advantage for experimental studies is that coiled coils often fold and interact autonomously, out of the context of a full-length protein.18 In this work, we used a combined computational/experimental strategy to examine coiled-coil interactions in the biological process of vesicular trafficking.

Vesicular trafficking provides a means of transporting material between the interior and the exterior of cells via exocytosis or endocytosis, as well as between intracellular membrane-enclosed compartments, including the endoplasmic reticulum, the Golgi apparatus and the lysosome/vacuole.19–21 A typical trafficking process involves vesicle formation, vesicle delivery, vesicle tethering and fusion between donor and acceptor membranes.22 In vesicle formation, cargo is selected and packaged into a membrane-coated vesicle. The vesicle is then delivered to its target membrane, usually along a track formed by microtubules or the actin filament system. Tethering brings the vesicle and the target membrane into close proximity, and the final step is membrane fusion, which releases vesicle contents. Although defects in these steps have been implicated in numerous human diseases, many details of the molecular mechanisms underlying these processes are not known.

Coiled coils are known to play critical roles in vesicular trafficking. The SNARE (Soluble N-Ethylmaleimide-Sensitive Factor Attachment Protein Receptor) proteins assemble into coiled-coil tetramers to promote membrane fusion,23 and long coiled-coil proteins, such as Uso1p (yeast homologue of mammalian p115) and Coy1p (yeast homologue of mammalian CASP), have been proposed to form coiled-coil-based elongated complexes that may bridge vesicles to their target membranes.24 Given that coiled coils clearly play important roles in many steps of vesicular trafficking, a more systematic understanding of such interactions may further elucidate mechanisms underlying these processes.

In this work, we implemented a strategy for discovering coiled-coil-mediated interactions in vesicular trafficking that involves computational filtering followed by experimental testing. We began by compiling lists of proteins associated with vesicular trafficking and predicting the number and location of coiled-coil regions in these proteins. A computational prefiltering strategy was then used to prioritize pairs of coiled coils for experimental testing. Coiled-coil-encoding fragments were cloned, and pairs of coiled coils that passed the computational filter were tested using the yeast two-hybrid (Y2H) assay. The result was 80 experimentally verified coiled-coil interactions among 58 pairs of proteins involved in vesicular trafficking. Because many of the coiled-coil proteins involved in trafficking are conserved, our work in yeast will help unravel related mechanisms in other organisms, including human.

Results

Proteins involved in vesicular trafficking are enriched in predicted coiled coils and potential coiled-coil-mediated interactions

Proteins annotated to have roles in trafficking were compiled from GO and MIPS FunCat25,26 and augmented using a manual survey of the literature to give a total of 415 proteins (Fig. 1a and Table S1). Using Paircoil2, a program that predicts the propensity of sequences to form coiled coils, we predicted 177 coiled-coil motifs in 119 proteins27 (see Materials and Methods). A protein was designated as a “coiled-coil protein” if it was predicted to contain at least one coiled-coil region of length greater than 28 residues. Figure 1b shows that proteins associated with vesicular trafficking are significantly enriched in predicted coiled coils ($P<e^{-20}$).
We found that predicted coiled-coil proteins are highly enriched in all four steps of vesicle formation, delivery, tethering and fusion, especially in the latter two, with frequencies up to 53% and 46%, respectively (Table S2). The coiled-coil proteins are widely distributed among various complexes/families, including the well-known SNAREs and long coiled-coil tethers (Table S1).

To explore whether coiled-coil-mediated protein interactions might be enriched among trafficking proteins, we compared the number of reported physical interactions among coiled-coil proteins involved in vesicular trafficking with that among the same number of randomly chosen non-coiled-coil proteins associated with trafficking (see Materials and Methods). The number of connections among the coiled-coil proteins is 287, 1.74 times of that (165±20) of the non-coiled-coil proteome ($P\leq e^{-5}$) (Fig. 1c). This suggests that coiled-coil proteins tend to associate with one another, possibly via interaction of their coiled-coil motifs.

A computational filtering strategy identifies candidate coiled-coil-mediated interactions between proteins involved in vesicular trafficking

Given both the high frequency of coiled coils in trafficking proteins and the relatively high frequency of reported interactions between these coiled-coil proteins, we postulated that the predicted coiled coils are likely to mediate many interactions. To uncover such interactions efficiently, we developed a computational filtering strategy. The procedure involved two steps: First, we compiled a protein association network. This network linked pairs of proteins annotated to function in trafficking if they were (1) each predicted to contain a coiled coil and (2) judged to have high potential to interact (see...
Then, we analyzed pairs of proteins in the association network at the level of their predicted coiled-coil motifs, using several metrics to prioritize pairs of motifs for experimental testing. The computational prefiltering step was followed by extensive Y2H testing for interactions between predicted coiled coils; this was done using protein fragments corresponding to just the predicted coiled-coil regions (Fig. 2a).

The protein-level association network was designed to comprehensively link coiled-coil-containing protein pairs judged likely to interact. Edges corresponding to reported physical associations were compiled from a variety of sources (see Materials and Methods). To increase the coverage of the network, we also included edges based on two other criteria: First, edges based on annotated or predicted similarity of function (as defined using high-resolution GO terms) were added because similarity of function is known to correlate with physical interaction. Second, interolog pairs for which homologous counterparts interact in other species were included. This resulted in a high-coverage association map of 826 protein pairs judged to have high potential to interact, among 113 predicted coiled-coil proteins involved in vesicular trafficking (Fig. 2a and b and Fig. S1).

To infer whether specific coiled-coil regions from these proteins were likely to interact, we considered five factors: stand-alone coiled-coil interaction score by the program bZIP_SCORE (bZIP_SCORE), the proportion of residues in the interacting proteins predicted to form a coiled coil (CC_PROP), confidence in the underlying protein-level interaction (EDGE_CONF), domain-level analysis of coiled-coil interaction potential (CC_DOMAIN) and coiled-coil interactions for which there is structural evidence from homologous proteins (CC_STRUCT) (Fig. 2c). To identify putative coiled-coil interactions, we first combined measures bZIP_SCORE, CC_PROP and EDGE_CONF and then integrated the predictions by CC_DOMAIN and CC_STRUCT (see Materials and Methods; for the contributions of these factors to the prediction, see Fig. 3a and Fig. S2). In total, we predicted 778 coiled-coil pairs between 150 coiled proteins.
At the protein level, this corresponded to 261 putative coiled-coil-mediated protein–protein interactions (CCMPIs) involving 94 proteins (Fig. S2). Experimental testing using the Y2H assay confirms many predicted interactions. The Y2H assay has been widely used to analyze interactions between proteins as well as between protein domains and motifs. The technique measures physical interactions (some may be indirect) between proteins or between domains by testing their ability to reconstitute a transcription factor in the nucleus of a yeast cell.32 The main deficiency of the assay is the large number of false negatives (i.e., protein interactions that cannot be detected in this way). False positives were once a significant concern, but modern two-hybrid methods have significantly reduced false-positive rates.8 To assess the performance of the Y2H assay under our experimental conditions for detecting coiled-coil interactions, we measured pairwise interactions among six human bZIP transcription factor coiled coils previously characterized using protein microarrays.33 Eleven of 12 pairwise interactions for which both assays gave unambiguous results indicated the same outcome in both experiments (Table S3). This indicates that the Y2H assay under our conditions is suitable for detecting coiled-coil interactions.

**Fig. 3.** Performance of computational prefiltering. (a) The number of predicted coiled-coil interactions that satisfied various filtering criteria. “High bZIP SCORE” refers to interactions with high bZIP SCORE and either high or low EDGE_CONF. “Lower bZIP_SCORE & CC_PROP” refers to lower bZIP_SCORE, high CC_PROP and either high or low EDGE_CONF (see Materials and Methods for details). (b) Evaluation of the computational filtering strategy. The plot compares the experimental interaction rate for coiled-coil pairs that passed the computational filter, designated “passed filter,” with the experimental interaction rate for those that did not pass the filter, designated “filtered out.” (c) The number of experimentally observed coiled-coil interactions that satisfied various criteria. (d) Supporting information for 80 coiled-coil interactions identified by Y2H assays. Numbers indicate how many coiled-coil interactions met different criteria; numbers in parentheses indicate interaction counts at the protein level. The “prior coiled-coil interaction” sector describes interactions reported by others to occur via coiled coils, whereas the “prior protein interaction” sector describes interactions previously reported at the protein level. The 30 coiled-coil interactions detected between previously unreported protein interactions were further tested by BiFC assay (see Materials and Methods).

**Fig. 4.** The 85 experimentally detected coiled-coil interactions. Each protein is represented as a gray horizontal bar under the protein name, with the location of the predicted coiled coil(s) indicated using thicker and darker gray bars. The complex and/or family that each protein belongs to was annotated manually. Complexes and families are bordered with dashed lines. Lines indicate experimentally observed interactions in the Y2H assay. Thick black lines are interactions previously unreported in compiled databases at the protein level, gray lines are interactions previously unreported at the coiled-coil level and thin black lines are interactions previously reported as mediated by coiled coils. Asterisks indicate interactions that did not pass the computational filter but were uncovered in further experimental testing.
Coiled-Coil Networking in Vesicular Trafficking
To test the predicted coiled-coil interactions, we cloned almost all coiled-coil regions involved in the predicted CCMPIs (only 1 of the 150 coiled coils failed, so 2 pairs were removed; Fig. 2d and Table S4). We then used the Y2H assay to test the 776 predicted interactions between these 149 stand-alone coiled coils. After eliminating 18 self-activators, we analyzed 611 coiled-coil pairs in a mating-type two-hybrid assay. The complete results are given in Fig. 4 and Table S5.

A total of 80 coiled-coil pairs were identified as positives by the yeast two-hybrid, which resulted in an experimental interaction network comprising 58 interaction pairs at the protein level. Among the 80 pairs of coiled-coil interactions, 50 were associated with 37 previously reported physical protein-level interactions, and the remaining 30 pairs were associated with 21 protein pairs not previously reported to physically interact (Fig. S1, Fig. 3d). We regard these 21 pairs as new protein-protein interactions. (In an extensive search of the literature, we found one of the 21 new interactions recently reported at the protein level, after our experiments were completed.) Among the 50 pairs of interactions previously known at the protein level, 15 were annotated as coiled-coil interactions (Table S6). This supports the reliability of our data set. For the 30 coiled-coil interactions representing new interactions, we used bimolecular fluorescence complementation (BiFC)34 to further verify interaction reliability (see Materials and Methods). Yu et al. have discussed how this assay is appropriate for validating Y2H results.8 Of the 30 pairs identified in the Y2H assay, excluding 9 pairs with ambiguous results due to the failure of cloning or self-activation, ~71% (15) of the remaining 21 pairs were confirmed as positives by BiFC (Fig. 3d and Table S7). This is a very high recovery rate, given the different requirements of the two assays in terms of protein folding geometry and cellular expression and localization. The much higher rate of reproducibility for these coiled-coil pairs, compared with other similar studies in the literature,7 probably arises because we are assaying simple structures, an advantage of our approach.

Computational filtering was effective

To evaluate the computational filtering, we subsequently tested all 1025 remaining coiled-coil interactions that could potentially occur between 565 protein pairs linked in the protein association network involved in vesicular trafficking. These interactions were derived from association network edges that were not predicted to interact via coiled coils. After eliminating self-activators, we analyzed 923 coiled-coil pairs. Only 5 of 923 interactions were found to be positive (asterisk-tagged interactions in Fig. 4).

Our experimental data suggest that the computational prefiltering, although fairly crude, was effective. Remarkably, the positive rate among coiled-coil interactions that passed the filter was greater than 13.1% (80/611), while the positive rate among coiled-coil pairs not predicted to mediate interactions was 0.5% (5/923) (Fig. 3b). This indicates that the computational prefiltering removed 63% (918/1449) of non-interacting coiled coils and retained 94% (80/85) of all positives.

We evaluated which aspects of the prefiltering strategy were most effective, as summarized in Fig. 3c and Figs. S1 and S2. At the protein level, 37 of 58 of the positives came from proteins that were linked via a reported physical interaction in the protein association network (and thus received a high EDGE_CONF value); the other 21 came from indirect indicators of interaction and so may represent the first detection of a physical association (Fig. S1c). At the coiled-coil interaction level, the bZIP_SCORE method was very useful. Nearly all the positives (77 of 80) contained coiled-coil pairs that passed one or more of the bZIP_SCORE criteria. The CC_DOMAIN and CC_STRUCT predictors added very few additional positives (2 and 1, respectively; see Fig. 3c).

Coiled-coil interactions among vesicular trafficking proteins are specific

The Y2H data showed that coiled-coil interactions among trafficking proteins do not occur indiscriminately and instead are quite specific. Even after employing a computationally directed strategy to target those protein pairs most likely to interact, only 80 of 611 coiled-coil pairs were positive. Including the larger set of results that we used to assess the computational framework, 1534 pairs were tested and only 85 were positive. The average number of interactions per coiled-coil segment was 2.2, of an average of 19.1 interactions tested (1534 pairs). The computational prefiltering strategy were most effective, as summarized in Fig. 3c and Figs. S1 and S2. At the protein level, 37 of 58 of the positives came from proteins that were linked via a reported physical interaction in the protein association network (and thus received a high EDGE_CONF value); the other 21 came from indirect indicators of interaction and so may represent the first detection of a physical association (Fig. S1c). At the coiled-coil interaction level, the bZIP_SCORE method was very useful. Nearly all the positives (77 of 80) contained coiled-coil pairs that passed one or more of the bZIP_SCORE criteria. The CC_DOMAIN and CC_STRUCT predictors added very few additional positives (2 and 1, respectively; see Fig. 3c).

Coiled-coil assays provide structural localization of the interaction motif

Our domain-level interaction assays provide a greater level of structural resolution than those employing full-length proteins. Figure 4 shows how many known (gray continuous lines) and previously unknown (thick black lines) interactions were localized to specific coiled coils by our tests. For example, the SNX4/SNX41/ATG20 complex participates in endosomal sorting and autophagy.35,36 A previous study showed that Smx4p interacts with Snx41p and Atg20p.35 In our work, we further observed that coiled-coil regions of Smx4p interact with C-terminal coiled coils from both Snx41p and Atg20p, suggesting that coiled-coil associations may constitute the structural basis for organization of this complex.

Similarly, the kinesin-related protein Smy1p and the class V myosin motor Myo2p have previously been shown to interact via their globular tails. Our additional observation of an interaction between predicted coiled coils in the head region is consistent with the observation that overexpression of Smy1p can overcome defects in both head and tail domains of Myo2p.37
The coiled-coil interactions in ESCRT (endosomal sorting complex required for transport)-III are also interesting (Fig. 4). ESCRT-III and other ESCRT complexes constitute a major pathway for lysosomal degradation of transmembrane proteins and are critical for many cellular processes and some pathological pathways (e.g., budding of HIV particles). Six protein components of ESCRT-III in yeast have been identified—Did4p, Vps20p, Vps24p, Snf7p, Vps60p and Did2p, all of which contain predicted coiled coils. Vps20p and Snf7p have been reported to interact with each other. We observed that there are multiple coiled-coil interactions between Vps20p and Snf7p, indicating that coiled-coil associations are likely to underlie the complex assembly. Furthermore, we identified two other novel interactions mediated by coiled coils in ESCRT-III proteins: Snf7p with Vps60p and Vps20p with Did4p. All ESCRT-III subunits have a conserved primary structure including a highly basic N-terminal region and an acidic C-terminal region. Both regions contain coiled coils, and a simple model proposed for association of these subunits is that the basic and acidic regions form an antiparallel coiled coil. Among the five coiled-coil associations we identified between the ESCRT-III subunits, three followed this N-to-C rule (Vps20pCC3–Snf7pCC1, Vps20pCC3–Snf7pCC2 and Vps20pCC3–Did4pCC). But there are two exceptions: the basic N-terminal CC1 of Snf7p can interact with the basic N-terminal CC1 of Vps20p and the basic N-terminal CC1 of Vps60p. These unusual basic–basic coiled-coil interactions may also have roles in the formation of the ESCRT-III complex and/or ESCRT-III lattices on the endosomal membrane.

Coiled-coil associations are featured in exocytosis and endocytosis

To probe the roles of coiled-coil interactions in different trafficking processes, we manually assigned each protein to a pathway according to the literature (Table S8). We observed that coiled-coil associations exist within and between many trafficking pathways, including the following: cytoplasm to vacuole, endoplasmic reticulum to Golgi, intra-Golgi, Golgi to endosome, endosome to Golgi and endosome to vacuole and degradative multivesicular body exocytosis and endocytosis (Fig. S3). In exocytosis, Sec2p homodimerization, Gyp5p–Gyl1p heterodimerization and interactions between SNARE proteins Sso1p–Snc1p and Sso1p–Snc2p are known to be mediated by coiled coils. However, we identified six more CCMPIs within the pathway. These are involved in each step of exocytosis. For example, Myo2p–Gyp5p/Gyl1p functions in delivery, Exo84p homodimer functions in tethering and Sso1p–Snc1/2p pairs in fusion. Many CCMPIs were also observed in endocytosis, another well-studied vesicular trafficking pathway in yeast. In endocytosis, coiled-coil protein Ede1p initiates vesicle formation, which is followed by recruitment of coiled-coil proteins Pan1p, End3p, Sed5p, Yap1802p, En12p and Sla2p. CMT1p indirectly regulates actin assembly and internalization, and Rvs161p and Rvs167p are involved in membrane scission. The CCMPI network further reveals that some other proteins may also function in endocytosis pathways through coiled-coil motifs, such as Sgm1p (with Ede1p) and Rud3p (with Rvs161p). These observations suggest that coiled coils have the potential to play critical roles in coupling events in dynamic trafficking pathways. Notably, six coiled-coil interactions occur between endocytosis and exocytosis, showing extensive connections through coiled coils between the two related processes.41

SNARE-protein coiled coils exhibit specific interactions

SNARE proteins mediate fusion via coiled-coil heterotetramers that involve the association of proteins in apposing membranes. Heterotetramers composed of one each of four categories of SNAREs—Qa, Qb, Qc and R—are required for fusion in vitro. However, it is clear that other SNARE complexes...
can form in vitro; several have been crystallized.42–45 In this study, we linked 14 pairs of SNARE proteins via coiled-coil-mediated interactions identified by Y2H, including 4 proteins of type Qa, 1 protein of type Qc, 2 proteins of type Q, and 5 proteins of type R (Fig. 5a; Table S9). Because the pairwise assay is highly unlikely to detect Qa(Qc)(Q)(R) heterotetramers (integral membrane endogenous proteins would have to be recruited into the nucleus by the two-hybrid bait or prey), the observed interactions likely involve just two SNARE coiled coils.

The SNARE coiled coils formed significantly more interactions with other SNARE proteins than with non-SNARE proteins (Fisher’s exact test, \( P = 9.5 \times 10^{-10} \)). Furthermore, within the SNAREs, we observed 14 pairs among SNARE proteins, and 7 occurred among components of reported complexes42,46 (thin lines in Fig. 5a). The other 7 occurred between proteins in distinct reported complexes (thick lines in Fig. 5a). The BiFC assay verified that 4 of these 7 previously unobserved pairs interact at the level of their coiled coils (Ykt6p–Sso1p, Sso2p–Vam3p, Sec22p–Sso2p and Tgl1p–Sec9p). These coiled-coil interactions were much more common among Q SNAREs or between Q and R SNAREs than among R SNAREs (Fig. 5b). This is consistent with genetic and biochemical tests that have indicated that complexes containing at most one R-type SNARE are functional.47,48 In contrast, coiled-coil complexes lacking an R SNARE but containing multiple Q SNAREs can be functional, and several have been crystallized.43–45 Two-component Q and R SNARE interactions have also been reported.49,50

Discussion

Identifying new protein–protein interactions and localizing interactions to specific protein domains or motifs are important steps in comprehensively mapping the interactome. Given the appealing simplicity of the coiled-coil motif, it may be possible in the future to computationally predict coiled-coil-mediated interactions with useful accuracy. This has been demonstrated for some coiled-coil-containing transcription factors.50–51 Currently, the large number of structurally annotated coiled-coil interactions that are needed to develop and improve computational prediction methods is not available. In this work, we therefore combined a computational prefiltering approach with extensive experiments to identify new coiled-coil-mediated interactions. Our computational method involved combining multiple weak indicators to identify those pairs of proteins judged most likely to interact via coiled coils. It worked surprisingly well, retaining 94% of positive coiled-coil interactions from the association network and eliminating 63% of non-interactions. Limitations of the approach, discussed below, suggest areas where further work and/or data may lead to improved performance.

We began by identifying trafficking proteins and predicting the locations of coiled-coil regions. At the Paircoil2 cutoff used, we likely missed \( \sim 23\% \) of positives and included \( \sim 1\% \) of all non-coiled-coil residues as false positives.27 Furthermore, computational coiled-coil detection likely led to the improper truncation of coiled coils and/or the fragmentation of longer coiled coils into shorter pieces. Such artifacts are a consequence of using a computational coiled-coil predictor to recognize poorly defined motif boundaries. Manual annotation of coiled-coil boundaries, informed by multiple-sequence alignment of homologues, could be helpful in this regard.

Our strategy involved searching for likely coiled-coil-mediated interactions only among proteins reported or inferred to have some association (physical, functional or evolutionary). Current estimates of the coverage of the yeast interactome suggest that \( \sim 50\% \) of all protein–protein interactions have been reported.53 For trafficking processes, where membrane proteins are common, this is likely to be an overestimate. We included all 287 reported physical associations involving coiled-coil proteins annotated with vesicular trafficking, and, using genetic interactions as well as functional and homology relationships, we also included 539 more probable associations. These edges in the association network are likely to include a significant number of the remaining unreported physical protein-level interactions. In fact, more recent large-scale studies of yeast interactions have identified only 8 additional protein pairs among proteins annotated to be involved in trafficking.7,8

The second step in the filtering strategy evaluated coiled-coil motif pairs. For this purpose, postevaluation showed that bZIP_SCORE was very useful (compare Fig. 3a with c). This was somewhat surprising. The bZIP_SCORE program is trained on parallel dimeric leucine-zipper transcription factors, where it performs well.50 However, the interaction topologies of the coiled-coil complexes assessed in the present study are not known. Parallel dimers are likely to be present, but antiparallel dimers, as well as trimers and/or tetramers may also be included. The effectiveness of filtering using this method suggests either that many of the detected pairs do form parallel dimers or that bZIP_SCORE can provide useful information even when the underlying interaction model is incorrect. The latter may be reasonable, as some amino acid interactions (e.g., electrostatic interactions between residues at the helix–helix interfaces) are similar in parallel coiled coils of different stoichiometry. In addition to bZIP_SCORE, the coiled-coil proportion (CC_PROP) was useful in predicting interaction partners when used in conjunction with lower bZIP scores.

Computational prefiltering led to a higher hit rate than is found in most large-scale Y2H studies. Criteria for positive interactions vary widely among different reports and are not readily comparable. In all-versus-all screens, rates for observing reproducible positives are around 0.002%–1%, whereas screens using protein fragments have led to higher reported positive frequencies, reaching \( \sim 0.1\%–0.2\% \).10,11 With a focus
Coiled-Coil Networking in Vesicular Trafficking

on motifs or domains that clearly have the structural potential to interact, discovery efficiency is expected to increase further. For example, in a study of coiled-coil motifs in proteins annotated to participate in chromosome segregation, Newman et al. observed a positive rate of ~ 1.6%.15 In our work, with the application of a computational filter, the Y2H assay achieved a positive rate of 13.1% (i.e., 80 positives of 611 predicted positive pairs tested). The increase came from appropriate identification of coiled-coil pairs with high interaction propensity. In addition, assaying short coiled-coil peptides may give lower false-negative rates than experiments employing full-length proteins or less well-defined protein fragments.

In conclusion, high-throughput screening of protein–protein interactions has proven valuable for obtaining insights into protein function and suggesting models for how the interactome is organized. Focused screens, such as the one we present here, are not as comprehensive as all-versus-all studies, but they can rapidly provide a larger number of positive interactions. Our computational focusing strategy led to the discovery of new coiled-coil-mediated interactions among proteins involved in vesicular trafficking in yeast. We found that most coiled-coil segments of trafficking proteins formed just a few complexes and did not associate indiscriminately with many other partners, consistent with findings of interaction specificity among other coiled coils in yeast and human proteins.15,33,52 This high degree of specificity is quite interesting, given the low complexity of coiled-coil sequences and structures, and it is an ongoing subject of investigations.53 The focused tests of this study also served to localize known protein interactions to specific regions of sequence. Localization information can be used in the future to engineer mutations or reagents that disrupt some interactions while minimally affecting the rest of the protein. As our structural understanding of interaction specificity improves, computational prefiltering strategies, such as the one presented here, should become much more accurate. Ultimately, this will reduce the number of experiments required to understand how protein interactions are wired at the systems level.

Materials and Methods

Identification of proteins involved in vesicular trafficking

The proteins involved in trafficking pathways were selected from the functional annotation databases GO and MIPS FunCat.26 We used yeast GO-Slim terms provided by the SGD database to identify trafficking functions; these included general terms and broad categories from the GO ontology for S. cerevisiae.54 We selected 234 proteins annotated as GO-Slim term “vesicle-mediated transport.” Additionally, for some terms related to vesicular trafficking, such as “sporulation” and “pseudohyphal growth,” we checked each protein by reading the summaries in SGD. For the MIPS FunCat database, we selected 265 proteins annotated as vesicular trafficking, vesicular cellular export, vesicular cellular import or intracellular transport vesicles. For some better-documented protein families functioning in vesicular trafficking, such as coats, tethers and SNAREs, we collected examples from reviews.

Coiled-coil prediction

Coiled coils were predicted from the yeast genome at the MIPS Comprehensive Yeast Genome Database by using the computer program Paircoil2 with a P-value cutoff of 0.99 and a window of 28 residues.27 In the case that there was a discontinuity in the predicted heptad register, multiple shorter coiled coils each with continuous register were defined. The enrichment of coiled-coil proteins in trafficking pathways and that in each step of vesicular trafficking were calculated using the hypergeometric test as follows: Suppose there are N proteins coded in the yeast genome of which m proteins contain coiled coils. The hypergeometric distribution describes the probability of finding k coiled-coil proteins when sampling n proteins randomly without replacement [denoted as P(X=k)]. Then, the P-value can be calculated as the cumulative probability P(X≥k).

Construction of protein networks

The physical interaction data used for the analysis in Fig. 1c were obtained from DIP_CORE55 and MIPS56 as of November 2006. The protein-level association network consisted of reported interaction data, AVID (Annotation Via Integration of Data) predictions and computationally predicted interologs. Interaction data were collected from six databases in November 2006: BIND,57 BIOGRID,58 DIP,59 MINT,55 MIPS56 and INTACT.60 AVID is a method for predicting functional relationships among proteins.25 The functional categories that lead to linkages in AVID are more specific than “vesicular trafficking.” Interologs were predicted by the generalized interolog mapping procedure proposed by Yu et al.70 Forty-three inferred interactions were transferred from human, fruit fly and worm to yeast. Human protein-interaction data were downloaded from the Human Protein Reference Database, whereas fruit fly and worm interaction data were downloaded from BIOGRID.58

Scoring coiled-coil interactions using bZIP_SCORE

To calibrate the performance of bZIP_SCORE and to choose appropriate scoring thresholds, we used data from the work of Newman et al.15 Their study identified coiled-coil interactions in chromosome segregation by using the Y2H assay. By supposing that protein pairs whose coiled coils interact also interact, we mapped these coiled-coil interactions to the protein level and obtained 160 CCMPIs of 7381 possible protein pairs involving 121 proteins. To estimate the performance of bZIP_SCORE on these interactions, we ran Paircoil2 and predicted coiled-coil sequences and registers for 119 proteins (coiled coils were not predicted in 2 proteins). Then, we ran the bZIP_SCORE method of Fong et al. using base-optimized weights to evaluate all pairs of predicted coiled coils.30 The Paircoil2-assigned heptad register was used in these calculations. The score assigned was the maximum over all possible alignments that had matching heptad assignments and a length ≥21 residues. If more than 1 coiled-coil

‡ http://db.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl
pair could form between 2 proteins (i.e., there were 2 or more coiled coils predicted in one or both of the 2 proteins), the highest score over all coiled-coil pairs was used. Of 7140 protein pairs tested, 6888 included coiled-coil alignments long enough to be assigned scores by bZIP_SCORE, covering 153 observed CCMPIs. By adjusting the bZIP_SCORE cutoff, we identified a value that retained 117 (76.5%) of CCMPIs while filtering out 3546 (52.6%) pairs of proteins whose coiled coils did not interact in the study of Newman et al. (Fig. S4). CC_PROP indicates the geometric average of the proportion of total residues in a pair of proteins that are predicted to form a coiled coil by Paircoil2 (Fig. S5). Marginal improvements of both the coverage and the efficiency of filtering true negatives in the Newman et al. data set were observed when combining the contribution of CC_PROP. To accomplish this, we designated positive pairs as those that satisfied either a high threshold of 30 for bZIP_SCORE or a lower threshold of 15 for bZIP_SCORE and an additional threshold of 20% for the geometric average of coiled-coil proportions. Applying these criteria, we retained 123 (80.4%) CCMPIs while filtering out 3952 (58.7%) of protein pairs with coiled coils that did not interact. Thus, we combined bZIP_SCORE and CC_PROP in this way when applying the computational filter to coiled-coil interactions involved in vesicular trafficking.

For filtering candidate coiled-coil interactions, we ran the bZIP_SCORE program on all possible coiled-coil pairings for proteins that interact in the association network, as described above. Also as described above for the Newman et al. data, a pair of proteins was identified as containing a candidate coiled-coil-mediated interaction if (1) the maximum score was above 30 or (2) the geometric average of the proportion of coiled-coil residues in each partner was high (CC_PROP, ≥20%) and the bZIP score was ≥15. For physically interacting protein pairs from the literature (i.e., those with a higher EDGE_CONF), less strict thresholds were used (i.e., the score cutoff was 20, or the score was above 10 and CC_PROP was higher than 10%).

Structural evidence for CCMPIs (CC_STRUCT)

We used default settings to run the computer program SOCKET62 to identify coiled coils from protein structures downloaded from the Protein Data Bank and PQS databases.63,64 These were used to infer the interaction of homologous proteins. Because of the low complexity of coiled-coil sequences, we cannot search for homologous examples using BLAST at the coiled-coil domain level. Instead, two criteria were introduced: First, we searched for homologous proteins by BLASTing full-length proteins containing SOCKET hits against full-length yeast coiled-coil proteins with an E-value cutoff of e−10.65 Second, we checked whether the coiled-coil regions in the protein structure and the coiled-coil region predicted by Paircoil2 overlapped. An overlap of at least 28 residues was required.

Inferring CCMPIs from domain–domain interactions (CC_DOMAIN)

We used RPS-BLAST against the conserved domain database (CDD)66 to identify domains in coiled-coil proteins associated with vesicular trafficking.67 Interactions with the potential to occur via these domains were downloaded from DOMINE.68 If a pair of potentially interacting domains each contained a significant proportion of predicted coiled-coil residues (≥20%), we assumed that their interactions were highly likely to be mediated by a coiled coil.

Generating AD and BK coiled-coil fusion constructs

The coiled-coil sequences used are listed in Table S4. Primers for predicted coiled-coil sequences were designed by using Web Primer3. The forward primer for each coiled coil was flanked by an NdeI restriction site; the reverse primer, by a BamHI restriction site. PCR products were amplified from genomic DNA of yeast strain W303BLa (ade2-1 ura3-1 his3-11, 15 trp1-1 leu2-3, 112 can1-100 lys2::hisG bar1::hisG MATa) and cloned into the Y2H vector pGBK7 (BK, Clontech). When the BK constructs were confirmed to contain the desired inserts by sequencing, the coiled coils were subcloned into pGADT7-Rec (AD, Clontech) and confirmed by restriction digest. In total, we cloned 169 coiled coils into both BK and AD vectors (Table S4). A predicted coiled coil in Vps1p failed to clone.

Y2H assay

To identify coiled-coil interactions by Y2H, we transformed pGBK7 constructs into the yeast haploid strain Y187 (MATα, ura3-52, his3-200, ade2-101, leu2-3, 112, gal4, met–, gal80D, URA3::GAL1UAS–GAL1TATA-lacZ) and pGADT7-Rec into AH109 (MATα, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4, gal80D, LYS2::GAL1UAS–GAL1TATA–HIS3, GAL2UAS–GAL2TATA-ADE2, URA3::MEL1UAS–MEL1 TATA-lacZ) as described by the manufacturer. The two plasmids were cotransformed into the same cell by mating haploid transformants on YPD plates at 30 °C for 24 h. Cells were replicated onto SD/-Leu/-Trp plates for 48 h to screen for the diploids containing both AD and BK constructs. Then, the diploid cells were replicated onto SD/-Leu/-Trp/-His/-Ade plates. Cells were grown at 30 °C for 14 days before recording positive results. All AD and BK strains were mated with the opposite mating type "empty" strain, which contained the plasmid pGBK7 or pGADT7-Rec to eliminate autoactivators. The autoactivation assay was carried out under the same conditions as above. All mating reactions were repeated three times, and all three were required to be positive to assign an interaction. The autoactivating coiled coils from both plasmids (BK and AD) were removed from the final results (Table S5).

Bimolecular fluorescence complementation assay

The BiFC assay used in this study is based on split green fluorescent protein (GFP). GFP was split into two non-fluorescent fragments: N-terminal (residues 1–157) and C-terminal (residues 158–238) fragments,34 designated as GN158 and GC158, respectively. Coiled coils involved in novel interactions identified by Y2H were subcloned both upstream and downstream of GN158 or GC158. The GN158 constructs were transformed into the yeast haploid strain ySC7 (MATα, ade2-1 ura3-1 his3-11, 15 trp1-1 leu2-3, 112 can1-100 lys2::hisG bar1::hisG pep4::kanMX), and the GC158 constructs were transformed into ySC8 (MATα, ade2-1 ura3-1 his3-11, 15 trp1-1 leu2-3, 112 can1-100 lys2::hisG bar1::hisG pep4::kanMX). After mating, diploid cells were incubated in SCR/-Leu/-Trp medium (synthetic

§http://www.yeastgenome.org/cgi-bin/web-primer
Among with 2% raffinose) for 40 h at 30 °C. Haploid transformants were also mated with the opposite mating type strains expressing only GN158 or GC158 fragments to serve as negative controls. The cells were analyzed by a type strains expressing only GN158 or GC158 fragments to serve as negative controls. The cells were analyzed by a fluorescent microscopy (excitation = 485 nm; emission = 535 nm). For a BiFC test, three independent repeats for the testing group (e.g., GN-CC-A/GC-CC-B) and its two controls (GN-CC-A/GC and GN/GC-CC-B) were made respectively under the same conditions. If the fluorescence intensity of a fusion construct was significantly higher than that of its parental GFP fragment alone, this construct was regarded as autoactive. If the GFP signal of a testing group was higher than both controls for all three repeats (P < 0.05), the pair (e.g., GN-CC-A/GC-CC-B) was regarded as a positive interaction. If none of the three tests was higher than its controls, the pair was regarded as a negative interaction. If three tests were not consistent, the pair was regarded as an uncertain result. For a certain pair of coiled-coil interaction (CC-A/CC-B), amino-terminal and carboxyl-terminal fusions were used to test eight distinct combinations (i.e., GN-CC-A/GC-CC-B, GN-CC-B/GC-CC-A, GC-CC-B/GC-CC-A, CC-A-GN/GC-CC-B, CC-B-GN/GC-CC-A, CC-A-GN/CC-B-GC and CC-B-GN/GC-CC-A). If one of them is positive as assessed above, the pair (CC-A/CC-B) is regarded as positive (P). If all of them are negative (N), the pair is negative (N). If there is an uncertain result in the eight combinations, this pair of coiled-coil interaction is uncertain (UN). The pair containing an autoactive construct is regarded as uncertain due to autoactivation [UN(A)] (Table S7). GN-Fos/GC-Jun (the same coiled-coil regions used in Table S3) was used as a positive control in each plate to make sure that the testing procedure was correct.

Acknowledgements

This work was supported by the Bai Ren Project of the Chinese Academy of Sciences, the National Natural Science Foundation of China (grant no. 30570428 to T.J.), Project “973” (grant nos. 2009CB18503 and 2006CB911002 to T.J. and grant no. 2009CB82506 to Y.W.) and by a National Science Foundation CAREER award (MCB-0347203 to A.E.K.). We are very grateful to W. Li, H. Hang and members of the Jiang and Keating labs for their comments on the manuscript.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.07.006

References


Coiled-Coil Networking in Vesicular Trafficking 241


