FEBS Letters 581 (2007) 2870–2876

Minireview

The hard cell: From proteomics to a whole cell model

Matthew J. Betts, Robert B. Russell*

Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received 14 May 2007; accepted 22 May 2007

Available online 30 May 2007

Edited by Thomas C. Marlovits

Abstract **Proteomics has provided a wealth of data related to the nature of the proteome, including subcellular location, copy number, interaction partners and protein complexes. This raises the question of whether it is feasible to combine these data, together with other data related to overall cellular structure, to construct a static picture of the cell. In this minireview, we discuss these data, and the issues of turning them into whole cell models.**

Keywords: Cell structure; Protein; Interaction; Model; Proteomics

1. Introduction

Following the sequencing of the first complete genomes, many studies have been performed to study different aspects of the associated proteomes. For many species, we now know a great deal about protein–protein interactions, determined in large-scale two-hybrid screens [1–7] or captured individually [8,9]. Affinity capture techniques, such as the TAP-MS system, have also discovered hundreds of complexes in several species [10–15]. Within the budding yeast *Saccharomyces cerevisiae*, many additional details are also available beyond interactions and complexes, including the approximate copy number for most proteins [16] and cellular locations [17,18] – both of which have been determined in high-throughput assays.

At the same time, the knowledge of the overall structure of cells is pushed to an ever higher resolution. Sophisticated imaging techniques can now pinpoint specific proteins inside cells [19] and the maturation of electron tomography (e.g. [20]) is providing images of whole cells (or large parts thereof) at resolutions approaching 3 nm. There has been much discussion on the use of predicted structures for complexes or interacting components as a means to bridge the gap between the interactome and whole cell models (e.g. [21,22]). However, to date this has focussed mostly on complements to electron microscopy or tomography, usually to provide suitable modelling templates that might be used as search objects in order to locate particular complexes inside images of cellular entities.

In this minireview we discuss a different idea – specifically whether or not it might be feasible to use information that is available from proteomics data such as cellular locations and copy number, together with details of cell-structure and the models or structures for individual proteins or complexes, to construct a static model of a cell consistent with what is already known. In this, we were very much inspired by the cellular paintings of Goodsell [23,24], which are wonderfully illustrative, despite very often being based on incomplete or potentially inaccurate data. The concept is distinct from, though complementary to, methods to add a spatial dimension to whole cell simulations [25,26]. It is also distinct from studies of interaction networks that consider only one copy of each molecule, and which tend to treat processes in the cell as blueprints or wiring diagrams (e.g. [27–29]).

Given an idea of the overall dimensions and structure of the cell, the set of proteins in it, the number of copies of each that typically exist, how they interact and where in the cell scaffold they are located, it would, in principle, be possible to get a static picture of the cell by constructing a model that satisfies the constraints imposed by these observations. However, the nature of the high-throughput techniques which determine the above details are such that one is immediately faced with both missing details, contradictions, false information, and a problem of working over many orders of magnitude in resolution.

2. The current scaffold of the cell

Cells themselves vary over a large size range, with the smallest (e.g. mycoplasma) being roughly 0.15 microns in diameter, and the largest eukaryotic cells being as large as 1 mm (or up to 3 m for giraffe nerve cells). For the purposes of this paper we will consider a typical budding Yeast cell, at about 5 μm in diameter, to represent a typical size. Compartments in the cell consist of the nucleus (~1.5 μm diameter), the mitochondria (0.1–1 μm), vacuoles (0.1–1 μm), and the golgi (0.1–0.5 μm). Much smaller than these are the macromolecules doing most of the functional business inside the cell: for instance proteins, which are about 3–5 nm across, and the molecular assemblies in which multiple macromolecules act together, at 10–100 nm.

Cellular entities of interest can thus differ in size by roughly four orders of magnitude. No single experimental technique is able to view the cell across the entire scale. When investigating details at one resolution, one often loses the context at the others. For instance, X-ray and NMR structures give atomic level details of proteins and often the details of their interactions

*Corresponding author. Fax: +49 6221 387 8517.
E-mail address: russell@embl.de (R.B. Russell).

0014-5793/$32.00
doi:10.1016/j.febslet.2007.05.062

with each other, but say little about their place in the overall picture. Light microscopy can show living cells and organelles, in contrast to the other methods which usually work with samples that are frozen (literally or otherwise). Particular proteins can be seen through the use of fluorescence tags, potentially in a high-throughput manner [19], but the resolution is much too low to see the individual molecules involved. Tomography lies somewhere between the two, being able to view segments of small cells that can be grown flat enough (e.g. *Dictyostelium discoideum* [30]) and sometimes to reconstruct entire volumes of cells that have been cryo-immobilised, embedded in plastic and sectioned (e.g. *Schizosaccharomyces pombe* [31]), and to identify large protein complexes such as the actin cytoskeleton, microtubule bundles and the 26S proteosome which can then be modelled in to the electron density [32].

3. How many molecules and where?

Complete genomes provide a list of the types of nucleic acids and proteins that a cell contains, but to place these molecules in the cell one must also know how many of each type is typically present, and in which cellular compartment one can expect to find them. Fortunately, for Yeast, genome-scale studies using immuno-detection of tandem affinity purification (TAP) tagged proteins, tagging with green fluorescent protein (GFP), and immuno-detection of epitope tagged proteins have provided estimates of the number of copies of each protein [16] and their subcellular locations [17,18].

These data are a great boon to attempts to model the cell, though it is important to remember their limitations. First, like many proteome studies, the techniques simply do not work for a certain fraction of molecules, owing most often to a failure of proteins to express with the labels used to identify them, and results can vary either in repetitions of the same technique, or between different techniques. For example, copy numbers of 206 proteins when measured three times each by the same method, gave standard deviations between 1% and 50% of the mean value, with the largest variations being observed for the largest copy numbers [16]. It is also telling that of the proteins common to two large-scale studies of subcellular localisation in yeast [17,18], 65% (457 of 701) had at least one location that was detected in only one of the studies.

Second, the experiments are fundamentally limited to the set of conditions under which the proteome has been studied. Currently, nearly all studies are performed in exponentially growing yeast. This means that molecules that are not expressed under these conditions will not be seen. Ideally, these genome-scale experiments would be repeated under a set of diverse conditions, though currently the cost of these studies precludes such a set of repetitions. Third, even under a single condition, what one sees is an average over various times in the life of a cell: all stages of the cell cycle, mating, etc. Fourth, the experiments themselves have resolution limits, or require human interpretation of variable experimental data. What one sees, then, is a somewhat restricted set of proteins, with both precise copy numbers and locations subject to some error.

Despite these limitations, the data are still very informative. For instance, one can readily assess the protein component of a Yeast cell. Taken together, the information on copy number [16] suggests a total of 46.5 million protein molecules in a cell. Modelling proteins as spheres with diameters proportional to their sizes (Fig. 1) gives a total protein volume of 3.1 μm^3 for the proteins observed in copy number studies, and a value of 4.2 μm^3 when extrapolating to compensate for the 27% of the proteome that were not seen in the dataset. An average yeast cell is about 5 μm in diameter, giving a volume of 65.44 μm^3, and taking away 15% to account for the volume occupied by cellular membranes, gives 57 μm^3. The studies suggest, then, that about 10% of the accessible cell volume is occupied by soluble proteins, which is in good agreement with, though to the low end of, estimates of 7–40% macromolecular occupied volume determined previously [33].

Data for cellular locations and copy numbers are independent, which makes it somewhat difficult to decide what fraction of the number of proteins resides in each location. Moreover, different experimental limitations and errors mean that one has copy numbers for some proteins but not subcellular locations, and vice versa. A naive approach to assigning copy numbers to subcellular locations is to split the whole-cell copy number for each protein equally (or proportionally depending on location size) between the locations in which that protein was detected. A more sophisticated approach would be to consider the copy numbers of the things that each protein interacts with and the stoichiometry of these interactions, along with the subcellular locations. This could also help to locate the more than 2 million copies of proteins whose subcellular locations were not detected (nor found in other annotated sources). It should be remembered, however, that the stoichiometry of interactions is rarely available, although for some complexes it can be seen in a known or modelled three-dimensional structure (e.g. the exosome [34]).

4. Who interacts with whom?

Recent years have seen a great increase in the number of detected interactions between macromolecules. The two-hybrid system has been applied genome-scale to several species, and has identified tens of thousands of interactions between proteins [1–7]. Complementary to this are affinity purification techniques coupled to mass-spectrometry, which have identified hundreds of protein complexes [10–15]. At the same time, efforts are underway to capture individually detected interactions by curation of hundreds of thousands of articles in the scientific literature [8,9,35,36], and importantly to assign confidence values to each technique and thus each identified interaction (e.g. [37–39]). Though they are lacking in details such as the structure or strength of the interaction, or (in the case of affinity purification data) precise details of binary interactions, these data provide important information about the proximity of proteins that could be incorporated into any whole cell model.

Again, limitations must be remembered if one is to use this information to place molecules into the cell. The two-hybrid and affinity purification systems are well known to miss real interactions, i.e. they have false negatives, as well as to propose incorrect associations, i.e. false positives (e.g. [37,40]). Some measure of the rate of both of these problems, and a means to correct for them are necessary if interactions or complexes are to be used as constraints for whole cell models.

More importantly, though there are many thousands of interactions known, biophysical details such as kinetic or

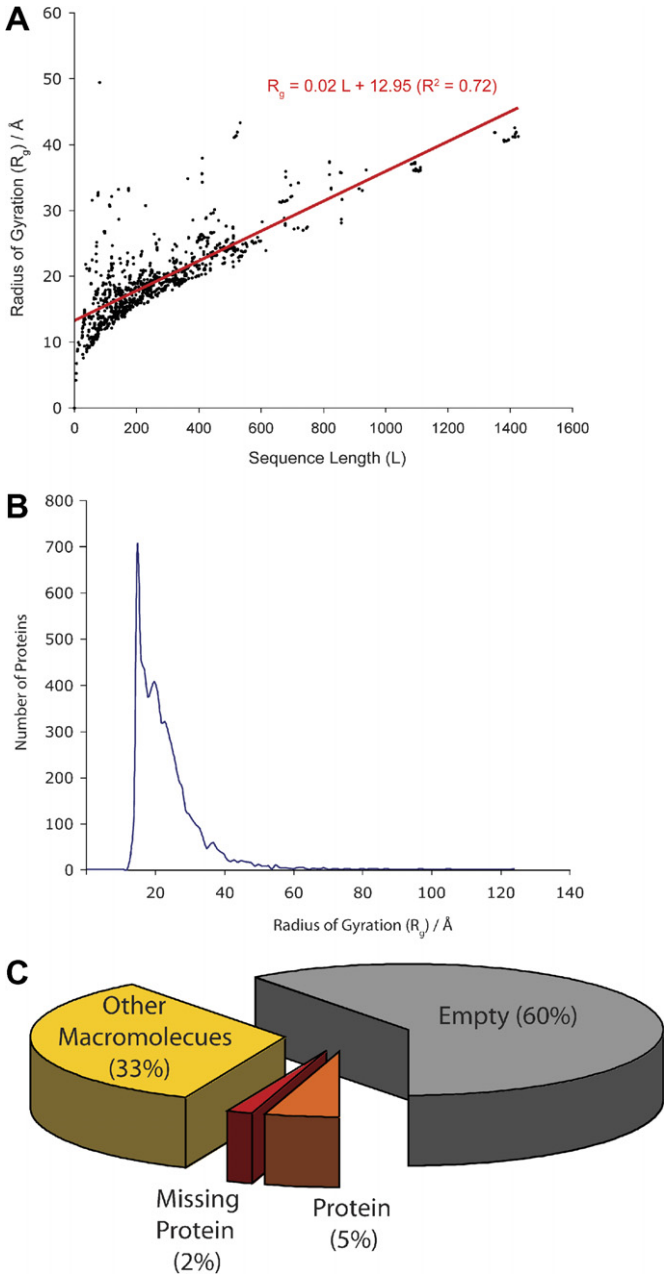

Fig. 1. Estimating macromolecular crowding from proteomics data. (A) Radius of gyration (R_g) vs. sequence length (L) for roughly 1000 yeast proteins in the PDB [65] that were named by the Saccharomyces Genome Database as homologues of their ORFs [66]. The red line shows that a least-squares fit of these data to a straight line is reasonable ($R^2 = 0.72$). (B) The parameters from the straight line fit were used to calculate an approximate R_g for every yeast ORF (i.e. including those without known structure). These are plotted here as a distribution, showing that most proteins have a diameter of 3–8 nm. (C) Yeast cells are between 5 and 10 μm in diameter, giving a volume of >65 μm^3. The total volume of each ORF= $N \times \frac{4}{3}\pi R_g^3$, where N = copy number [16]. This gives a total protein volume of 3.09 μm^3, which is 5% of the volume of the smallest cell. Another 1.14 μm^3 (2%) can be calculated from the 27% of annotated ORFs whose copy numbers were not observable [16], although some of these could be false ORFs or not expressed in log-phase growth. Previous studies estimate that 7–40% of the cell is occupied by macromolecules [33], which suggests that 0–33% is occupied by non-protein macromolecules such as nucleic acids, polysaccharides and membrane lipids.

thermodynamic constants are known for only a very small fraction of these: a previous estimate is that they are available for just 54 of probably more than a million physically interacting protein pairs [10]. In an ideal world, a full set of these details could help guide the construction of a whole cell model, though in the short term one must cope with the fact that details such as interaction affinity are simply not going to become generally available. Efforts to deduce numbers approximating affinities directly from proteomics data [10,41] might provide an intermediate solution.

5. When one knows the answer?

When using the datasets mentioned above, one must consider all of the possible limitations. It is also informative to consider

how these data look when projected onto a situation where we know reasonably well what the "true" picture should be. For this one can consider molecular machines that have been the subject of intense individual study. For a small fraction of complexes inside the cell, we know a great deal. For instance, the proteasome, RNA polymerase II, Arp2/3 complexes all have high-resolution crystal structures, and a lot of other supporting information such as stoichiometry is known for some complexes often before structures have been determined.

For example, the exosome is involved in 3′ to 5′ degradation of mRNA in the nucleus, and combines with the ski complex to perform cytoplasmic 3′ to 5′ mRNA degradation [42]. The core of the yeast exosome has 12 components [10]. Early experiments suggested that the principle players in this complex are present as single copies [43], which was confirmed for the nine core components in the X-ray structure [34]. The structure and previous in depth two-hybrid studies [44] also revealed the way the components were assembled: ie. who interacts with whom (Fig. 2A). The Ski complex is also thought to have a 1:1:1 stoichiometry [45].

Genome scale two-hybrid data [1,2] identifies only two of the interactions known to occur inside the exosome, and says

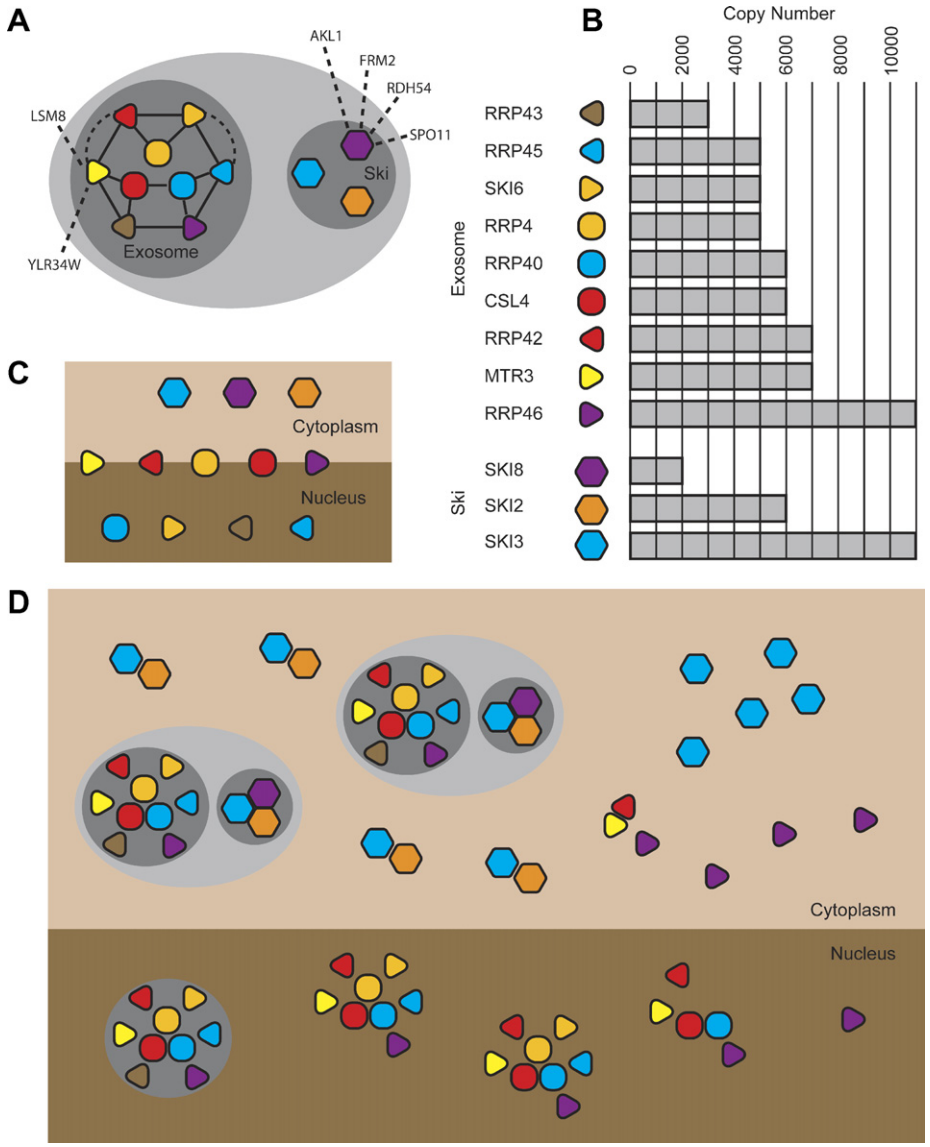

Fig. 2. Merged structural and proteomics data of the exosome and ski core complexes. (A) Interactions between proteins in the exosome and ski complexes. Solid lines = direct contacts in the human exosome structure. Dashed lines = yeast two-hybrid interactions [1,2], including to those proteins that are not part of the exosome or ski (shown by their gene names). All pairs of proteins in dark grey areas have a socio-affinity score $\geqslant 10$ [10]. Light grey areas connect proteins through pairs with socio-affinity score $\geqslant 5$ and <10. (B) Copy numbers [16], to the nearest 1000 copies. (C) Sub-cellular locations [17,18]. Proteins that were detected in both the nucleus and the cytoplasm are shown straddling the regions. (D) One way of putting it all together. Each triangle, square and hexagon represents 1000 molecules. As many molecules as possible are placed close to their interaction partners. For proteins that were detected in both the nucleus and cytoplasm, the number of copies is divided equally between the two regions. Exosome components are shown in the cytoplasm even if they were not detected there, since they are required for the interaction of Exosome with Ski. The figure includes the yeast homologues of the nine components of the human exosome structure [34] plus the components of the ski complex. Rounded triangles = components of the exosome ring, rounded squares = components of the Exosome cap, hexagons = components of the Ski complex. Colours and shapes are constant throughout. Gene names are given in (B).

nothing about the ski complex (Fig. 2A). Given the tendency for false negatives, it is perhaps not surprising that so many interactions were missed. Possibilities to correct for these could come from methods to infer interactions by homology, either *interologs* (interactions between equivalent proteins in different species [46] or by modelling protein interfaces [47,48]).

Inspection of the copy numbers for these proteins (Fig. 2B) gives a curious picture. For the exosome, proteins that should be there in equal copies range from 3138 (RRP43) to 10757 (RRP46), and one sees an even greater range for Ski components (1707–11931). This gives some hints as to both the experimental accuracy and other variability that is impossible to capture in such experiments, such as different rates of degradation.

For both complexes, there have been investigations into their cellular locations. The exosome core is thought to exist both in the cytoplasm and the nucleus with different accessory proteins modulating its function accordingly [49]. The Ski complex, in contrast, is thought to act only in the cytoplasm. Within the high throughput data [17,18], the ski complex components were found only in the cytoplasm as expected, but some of the core exosome components are seen only in the nucleus (Fig. 2C). Again, this highlights limitations of the approach – something prevented the entire exosome from being seen in the cytoplasm, despite the fact that there is evidence for all core components to be there.

Combining these data together naively (Fig. 2D) would give a picture that was out of keeping with what is known. Specifically, it would appear to suggest partial structures, subunits in isolation, or perhaps complexes with stoichiometries differing from that known. This suggests that it would be necessary somehow to correct for what are probably limitations in the techniques to measure copy number and location. The combination of high-throughput complex purification datasets might be the key: the core exosome components are consistently together in many purifications [10,11], suggesting that these data could be used to correct for such anomalies.

6. Putting things together

How do we integrate all the data in to a structural model of a cell? More importantly, do we need to reconcile all inconsistencies, and fill in all gaps before obtaining anything useful? And moreover what can we hope to use such a model for? Several models of cells and large subcellular structures currently exist. Perhaps most striking are the beautiful pictures of Goodsell that integrate structures of different sizes and information from a variety of sources [23,24]. Elsewhere, Takamori et al. recently published a structural model of a synaptic vesicle [50] that combined experimental information on vesicle size and structure, protein and lipid types, copy numbers, and protein and membrane structures, and used comparative modelling, molecular dynamics, and hand-editing to fit this data in to a model of a single vesicle. Many methods for simulating reaction dynamics model structure either implicitly or explicitly [25,51–56], although they usually do not explicitly model proteins and structures not directly involved in the reaction of interest.

Cells are, of course, not static entities. Even ignoring drastic variations, such as division, or mating, they are in a constant state of change. Some cellular structures form spontaneously (e.g. [57]), while others require the input of energy, a constant turnover of components and/or the participation of other factors (e.g. [58–63]). Most of the data discussed above are an average over time and multiple cells, and will clearly not capture this dynamism. This begs the question as to whether it is at all possible to construct a static model without a more detailed consideration of the dynamics.

The true picture of the cell does not fit well onto the blueprint or wiring diagram analogies that are so common in textbook biology. When building the whole cell it is probably worth considering other analogies, such as ant colonies, bee swarms, the flocking of birds, traffic patterns and the behaviours of crowds (e.g. [64]), which may help to explain some of the data.

A static whole cell model would enable many studies, such as virtual fluorescence labelling, investigations of diffusion and other cell-scale processes, the identification of contradictions and artefacts in the data, the production of approximate physical constants, and would focus attention on blurry parts of the picture where data are lacking. As such it would be a boon to efforts to model the whole cell in a fashion that considers as much of what is currently known as possible.

Acknowledgment: This work was supported by the EU grant 3D-Repertoire, contract number LSHG-CT-2005-5120828.

References

[1] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98 (8), 4569–4574.

[2] Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S. and Rothberg, J.M. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 403 (6770), 623–627.

[3] Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H. and Wanker, E.E. (2005) A human protein–protein interaction network: a resource for annotating the proteome. Cell 122 (6), 957–968.

[4] Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P. and Vidal, M. (2005) Towards a proteome-scale map of the human protein–protein interaction network. Nature 437 (7062), 1173–1178.

[5] Rain, J.C., Selig, L., De Reuse, 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 (6817), 211–215.

[6] Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.F., Lamesch, P., Xu, L., Tewari, M., Wong, S.L., Zhang, L.V., Berriz, G.F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H.W., Elewa, A., Baumgartner, B., Rose, D.J., Yu, H., Bosak, S.,

Sequerra, S., Fraser, A., Mango, S.E., Saxton, W.M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K.C., Harper, J.W., Cusick, M.E., Roth, F.P., Hill, D.E. and Vidal, M. (2004) A map of the interactome network of the metazoan *C. elegans*. Science 303 (5657), 540–543.

[7] Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C.A., Finley Jr., R.L., White, K.P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R.A., McKenna, M.P., Chant, J. and Rothberg, J.M. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302 (5651), 1727–1736.

[8] Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U. and Eisenberg, D. (2004) The Database of Interacting Proteins: 2004 update. Nucleic Acids Res 32 (Database issue), D449–D451.

[9] Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R. and Hermjakob, H. (2007) IntAct – open source resource for molecular interaction data. Nucleic Acids Res 35 (Database issue), D561–D565.

[10] Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B. and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440 (7084), 631–636.

[11] Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrin-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A. and Greenblatt, J.F. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440 (7084), 637–643.

[12] Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415 (6868), 141–147.

[13] Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W., Figeys, D. and Tyers, M. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 415 (6868), 180–183.

[14] 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 (7025), 531–537.

[15] Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., Hopf, C., Huhse, B., Mangano, R., Michon, A.M., Schirle, M., Schlegl, J., Schwab, M., Stein, M.A., Bauer, A., Casari, G., Drewes, G., Gavin, A.C., Jackson, D.B., Joberty, G., Neubauer, G., Rick, J., Kuster, B. and Superti-Furga, G. (2004) A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat. Cell Biol. 6 (2), 97–105.

[16] Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S. (2003) Global analysis of protein expression in yeast. Nature 425 (6959), 737–741.

[17] Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S. and O'Shea, E.K. (2003) Global analysis of protein localization in budding yeast. Nature 425 (6959), 686–691.

[18] Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., Cheung, K.H., Miller, P., Gerstein, M., Roeder, G.S. and Snyder, M. (2002) Subcellular localization of the yeast proteome. Genes Dev. 16 (6), 707–719.

[19] Pepperkok, R. and Ellenberg, J. (2006) High-throughput fluorescence microscopy for systems biology. Nat. Rev. Mol. Cell Biol. 7 (9), 690–696.

[20] Nickell, S., Kofler, C., Leis, A.P. and Baumeister, W. (2006) A visual approach to proteomics. Nat. Rev. Mol. Cell Biol. 7 (3), 225–230.

[21] Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A.C., Bork, P., Superti-Furga, G., Serrano, L. and Russell, R.B. (2004) Structure-based assembly of protein complexes in yeast. Science 303 (5666), 2026–2029.

[22] Sali, A., Glaeser, R., Earnest, T. and Baumeister, W. (2003) From words to literature in structural proteomics. Nature 422 (6928), 216–225.

[23] Goodsell, D.S. (2005) Visual methods from atoms to cells. Structure 13 (3), 347–354.

[24] Goodsell, D.S., http://mgl.scripps.edu/people/goodsell/. 2007.

[25] Ander, M., Beltrao, P., Di Ventura, B., Ferkinghoff-Borg, J., Foglierini, M., Kaplan, A., Lemerle, C., Tomas-Oliveira, I. and Serrano, L. (2004) SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. Syst. Biol. (Stevenage) 1 (1), 129–138.

[26] Takahashi, K., Arjunan, S.N. and Tomita, M. (2005) Space in systems biology of signaling pathways – towards intracellular molecular crowding in silico. FEBS Lett. 579 (8), 1783–1788.

[27] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34 (Database issue), D354–D357.

[28] BioCarta, "BioCarta", http://www.biocarta.com/. 2007.

[29] Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature 411 (6833), 41–42.

[30] Medalia, O., Weber, I., Frangakis, A.S., Nicastro, D., Gerisch, G. and Baumeister, W. (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. Science 298 (5596), 1209–1213.

[31] Hoog, J.L., Schwartz, C., Noon, A.T., O'Toole, E.T., Mastronarde, D.N., McIntosh, J.R. and Antony, C. (2007) Organization of interphase microtubules in fission yeast analyzed by electron tomography. Dev. Cell 12 (3), 349–361.

[32] Frangakis, A.S. and Forster, F. (2004) Computational exploration of structural information from cryo-electron tomograms. Curr. Opin. Struct. Biol. 14 (3), 325–331.

[33] Ellis, R.J. and Minton, A.P. (2006) Protein aggregation in crowded environments. Biol. Chem. 387 (5), 485–497.

[34] Liu, Q., Greimann, J.C. and Lima, C.D. (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell 127 (6), 1223–1237.
[35] Chatr-aryamontri, A., Ceol, A., Palazzi, L.M., Nardelli, G., Schneider, M.V., Castagnoli, L. and Cesareni, G. (2007) MINT: the Molecular INTeraction database. Nucleic Acids Res. 35 (Database issue), D572–D574.
[36] Bader, G.D., Betel, D. and Hogue, C.W. (2003) BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 31 (1), 248–250.
[37] von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. Nature 417 (6887), 399–403.
[38] von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B. and Bork, P. (2007) STRING 7 – recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. 35 (Database issue), D358–D362.
[39] Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F. and Gerstein, M. (2003) A Bayesian networks approach for predicting protein–protein interactions from genomic data. Science 302 (5644), 449–453.
[40] Aloy, P. and Russell, R.B. (2002) Potential artefacts in protein-interaction networks. FEBS Lett. 530 (1–3), 253–254.
[41] Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S. and Krogan, N.J. (2007) Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. Mol. Cell Proteomics 6 (3), 439–450.
[42] Orban, T.I. and Izaurralde, E. (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA 11 (4), 459–469.
[43] van Hoof, A. and Parker, R. (1999) The exosome: a proteasome for RNA? Cell 99 (4), 347–350.
[44] Raijmakers, R., Egberts, W.V., van Venrooij, W.J. and Pruijn, G.J. (2002) Protein-protein interactions between human exosome components support the assembly of RNase PH-type subunits into a six-membered PNPase-like ring. J. Mol. Biol. 323 (4), 653–663.
[45] Brown, J.T., Bai, X. and Johnson, A.W. (2000) The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. RNA 6 (3), 449–457.
[46] Matthews, L.R., Vaglio, P., Reboul, J., Ge, H., Davis, B.P., Garrels, J., Vincent, S. and Vidal, M. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or "interologs". Genome Res. 11 (12), 2120–2126.
[47] Aloy, P. and Russell, R.B. (2002) Interrogating protein interaction networks through structural biology. Proc. Natl. Acad. Sci. USA 99 (9), 5896–5901.
[48] Lu, L., Lu, H. and Skolnick, J. (2002) MULTIPROSPECTOR: an algorithm for the prediction of protein–protein interactions by multimeric threading. Proteins 49 (3), 350–364.
[49] Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D. and Mitchell, P. (1999) The yeast exosome and human PM-Scl are related complexes of $3' \rightarrow 5'$ exonucleases. Genes Dev. 13 (16), 2148–2158.
[50] Takamori, S., Holt, M., Stenius, K., Lemke, E.A., Gronborg, M., Riedel, D., Urlaub, H., Schenck, S., Brugger, B., Ringler, P., Muller, S.A., Rammner, B., Grater, F., Hub, J.S., De Groot, B.L., Mieskes, G., Moriyama, Y., Klingauf, J., Grubmuller, H., Heuser, J., Wieland, F. and Jahn, R. (2006) Molecular anatomy of a trafficking organelle. Cell 127 (4), 831–846.
[51] Andrews, S.S. and Arkin, A.P. (2006) Simulating cell biology. Curr. Biol. 16 (14), R523–R527.
[52] Coggan, J.S., Bartol, T.M., Esquenazi, E., Stiles, J.R., Lamont, S., Martone, M.E., Berg, D.K., Ellisman, M.H. and Sejnowski, T.J. (2005) Evidence for ectopic neurotransmission at a neuronal synapse. Science 309 (5733), 446–451.
[53] Ridgway, D., Broderick, G. and Ellison, M.J. (2006) Accommodating space, time and randomness in network simulation. Curr. Opin. Biotechnol. 17 (5), 493–498.
[54] Slepchenko, B.M., Schaff, J.C., Carson, J.H. and Loew, L.M. (2002) Computational cell biology: spatiotemporal simulation of cellular events. Annu. Rev. Biophys. Biomol. Struct. 31, 423–441.
[55] Slepchenko, B.M., Schaff, J.C., Macara, I. and Loew, L.M. (2003) Quantitative cell biology with the virtual cell. Trends Cell Biol. 13 (11), 570–576.
[56] Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C. and Hutchison Jr., C.A. (1999) E-CELL: software environment for whole-cell simulation. Bioinformatics 15 (1), 72–84.
[57] Anfinsen, C.B., Haber, E., Sela, M. and White Jr., F.H. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl. Acad. Sci. USA 47, 1309–1314.
[58] Howard, M. and Kruse, K. (2005) Cellular organization by self-organization: mechanisms and models for Min protein dynamics. J. Cell Biol. 168 (4), 533–536.
[59] Kuthan, H. (2001) Self-organisation and orderly processes by individual protein complexes in the bacterial cell. Prog. Biophys. Mol. Biol. 75 (1–2), 1–17.
[60] Nedelec, F., Surrey, T. and Karsenti, E. (2003) Self-organisation and forces in the microtubule cytoskeleton. Curr. Opin. Cell Biol. 15 (1), 118–124.
[61] Wittmann, T., Hyman, A. and Desai, A. (2001) The spindle: a dynamic assembly of microtubules and motors. Nat. Cell. Biol. 3 (1), E28–E34.
[62] Misteli, T. (2005) Concepts in nuclear architecture. Bioessays 27 (5), 477–487.
[63] Olson, M.O. and Dundr, M. (2005) The moving parts of the nucleolus. Histochem. Cell Biol. 123 (3), 203–216.
[64] Helbing, D., Farkas, I. and Vicsek, T. (2000) Simulating dynamical features of escape panic. Nature 407 (6803), 487–490.
[65] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. Nucleic Acids Res. 28 (1), 235–242.
[66] Hong, E., Balakrishnan, R., Christie, K., Costanzo, M., Dwight, S., Engel, S., Fisk, D., Hirschman, J., Livstone, M., Nash, R., Oughtred, R., Park, J., Skryzpek, M., Starr, B., Andrada, R., Binkley, G., Dong, Q., Hitz, B., Miyasato, S., Schroeder, M., Weng, S., Wong, E., Zhu, K., Dolinkski, K., Botstein, D. and J. Cherry. Saccharomyces Genome Database, ftp://ftp.yeastgenome.org/yeast/, 2007.