The Levinthal paradox of the interactome

Peter Tompa^{1*} and George D. Rose²

¹VIB Department of Structural Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

²Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland MD 21218

Received 6 September 2011; Revised 22 September 2011; Accepted 23 September 2011

DOI: 10.1002/pro.747

Published online 10 October 2011 proteinscience.org

Abstract: The central biological question of the 21st century is: how does a viable cell emerge from the bewildering combinatorial complexity of its molecular components? Here, we estimate the combinatorics of self-assembling the protein constituents of a yeast cell, a number so vast that the functional interactome could only have emerged by iterative hierarchic assembly of its component sub-assemblies. A protein can undergo both reversible denaturation and hierarchic self-assembly spontaneously, but a functioning interactome must expend energy to achieve viability. Consequently, it is implausible that a completely “denatured” cell could be reversibly renatured spontaneously, like a protein. Instead, new cells are generated by the division of pre-existing cells, an unbroken chain of renewal tracking back through contingent conditions and evolving responses to the origin of life on the prebiotic earth. We surmise that this non-deterministic temporal continuum could not be reconstructed *de novo* under present conditions.

Keywords: interactome; protein–protein interaction; Levinthal; protein folding; irreversibility; assembly pathway; steady state; combinatorics

Introduction

Protein folding, the spontaneous acquisition of native conformation under physiological conditions,¹ remains as one of the major unsolved problems in biological chemistry. The underlying search issue was formulated persuasively by Cyrus Levinthal² in a back-of-the-envelope calculation, which demonstrated that a polypeptide chain could not arrive at its native structure in biological real-time by random search because conformational space is far too vast. His formulation has come to be known as the “Levinthal paradox,” although for Levinthal it was no paradox at all but rather a demonstration that folding

proceeds along preferred pathways. Levinthal’s calculation has influenced many current formulations of the search problem in protein folding, see, for example, Dill and Chan.³

Understanding how a protein acquires its native structure, however, is only the initial search problem. Successful cellular function depends upon subsequent interactions with a host of other cellular constituents, resulting in a complex network called the interactome. A comprehensive description of the interactome has become the focus of recent ambitious high-throughput protein–protein interaction studies.^{4,5}

Unlike protein folding, self-assembly of the interactome has not yet prompted such widespread attention, and for understandable reasons. It is a problem of bewildering complexity, far more challenging than the beguiling simplicity of two-state proteins like ribonuclease that can self-assemble *in vitro*.⁶ Where does one begin? Our goal here is to show that assembly of the interactome in biological real-time is analogous to folding in that the functional state is selected from a staggering number of useless or potentially deleterious alternatives. In particular, a simplified calculation is sufficient to show that the number of distinguishable states of

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Korea Research Council of Fundamental Science and Technology (KRCF); Grant sponsor: FP7 Marie Curie Initial Training Network; Grant number: 264257, IDPbyNMR; Grant sponsor: FP7 Infrastructures; Grant number: 261863, BioNMR; Grant sponsors: National Science Foundation and the Mathers Foundation

*Correspondence to: Peter Tompa, VIB Department of Structural Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. E-mail: ptompa@vub.ac.be

the interactome exceeds comprehension. Consequently, the cell cannot self-organize by random assembly of its components. Instead, there must be pathways of hierarchic self-organization that result in functional modules, as proposed by Alberts.⁷ Here, we extend this proposition by incorporating knowledge that the functional interactome requires a continuous influx of energy for its generation and maintenance. This requirement has significant implications in evolution, physiology, pathology, and synthetic biology.

Levinthal Paradox of the Interactome

Levinthal's calculation² assumed nine possible configurations for each ϕ, ψ -pair in the backbone (three staggered configurations for each rotatable bond, like ethane), resulting in $9^{100} \approx 10^{95}$ possible conformations for a chain of 100 residues. Given the time required for single bond rotations (picoseconds), even a small protein that initiated folding by random search at the time of the big bang would still be thrashing about today.⁸ The Levinthal estimate is based on Flory's simplifying assumption⁹ that each ϕ, ψ -pair is sterically independent of the others. That assumption has been challenged,^{10,11} but the search problem persists.

If the protein search problem seems perplexing, the corresponding problem for a cell is bewildering. Taking yeast as a model organism, approximately 4500 different proteins are expressed during log-phase growth, each present in 50 to more than 10^6 copies per cell, with a median value of about 3000 and a median length of about 400 residues $\approx$ 50 kDa molecular weight.¹² Assuming spherical shape and average density 1.1 g/cm^3 , the median protein would have a radius of $26.3 \text{ \AA}$ and a surface area of $8692 \text{ \AA}^2$. Next, assume the surface area of an average protein:protein interface is about $800 \text{ \AA}^2$, the equivalent of 22 interfacial residues, each contributing $36.4 \text{ \AA}^2$.¹³ Also assume that displacement by a residue or rotation by its diameter (where each residue's surface of $36.4 \text{ \AA}^2$ is represented by a circular patch, diameter = $6.8 \text{ \AA}$) would alter the specificity of interaction within each interface. This works out to be $8692 \text{ \AA}^2 / 36.4 \text{ \AA}^2 = 239$ possible interface centers, with rotations producing 14.8 different orientations for each (again, assuming the interface is a circular patch of $800 \text{ \AA}^2$, perimeter = $100 \text{ \AA}$). In all, an average protein would have approximately 3540 distinguishable interfaces.

Assuming the simplest case that each of n proteins is present in a single copy in the proteome and all proteins engage in pairwise interactions (Fig. 1), the total number of possible distinct patterns of interactions is:

$$\frac{n!}{2^{n/2} \times \frac{n}{2}!}$$

Figure 1. The number of possible interactomes increases exponentially with proteome size. The number of possible different states (patterns of pairwise interactions) of the interactome increases exponentially with the number of its constituent proteins. In the simple case of four proteins (A), the number of possible different arrangements is only three. Five proteins (B) may already engage in 15 different pairwise interactions. The first pair (red-blue, red-purple, red-yellow, red-green) is connected by a solid line, followed by any of three possible secondary pairs (with connections indicated by dotted lines), plus three remaining possibilities (not illustrated) in which the first protein (red) is unpaired. The theoretical number for n proteins is $n! / 2^{n/2} \times n/2!$ (cf. text and Supporting Information), which for a realistic interactome of 4500 proteins gives 10^{7200} different possibilities.

(for details of calculations, cf. Supporting Information). For $n = 4500$, this is on the order of 10^{7200} , an unimaginably large number; but a more realistic calculation is yet more complicated. With an average of 3540 distinct interfaces for a single protein, there are $4500 \times 3540 = 1.6 \times 10^7$ entities, resulting in $10^{5.4 \times 10^7}$ possible distinct interaction patterns (cf. Supporting Information). If proteins are present in 3000 copies instead of a single copy, identical pairwise complexes of the same pair should not add to multiplicity of interactions patterns; nevertheless, the number of distinct interactomes increases further because different copies of the same protein can engage in interactions with different partners at the same time. In this case, the estimated number of different interactomes is on the order of $10^{7.9 \times 10^{10}}$ (cf. Supporting Information).

Of course, there are additional complicating factors such as alternative splicing, post-translational modifications, non-pairwise macromolecular interactions, incorrect complex formation that is adventitiously stable, and so forth. However, even neglecting such complications, the numbers preclude formation of a functional interactome by trial and error complex formation within any meaningful span of time. This numerical exercise, a “*Levinthal paradox of the interactome*”, is tantamount to a proof that the cell does not organize by random collisions of its interacting constituents. In analogy to protein folding,^{14,15} an inescapable conclusion from these numbers is that interactome assembly proceeds

along pathways and results in a hierarchy of functional modules.⁷ This conclusion is not altogether surprising when the number of pairwise interactions increases beyond a certain threshold, as shown abstractly for random graphs by Erdős and Rényi¹⁶ and for scale-free real-world networks by Gavin *et al.*⁴

Hierarchic Assembly of the Interactome

At the level of relatively simple multiprotein complexes, such as the bacterial ribosome, effective and spontaneous self-assembly can be observed in reconstitution experiments *in vitro*.^{17,18} In a series of classic papers, Nomura and coworkers have shown that fully active 30S *E. coli* ribosome assembles from its isolated components—16S RNA and 21 purified proteins. This was a remarkable early demonstration that components of the ribosome encode its assembly pathway and final assembled state. Such self-assembling complexes represent fundamental modules in the cellular hierarchy. In a similar vein, *de novo* synthesis of infectious poliovirus in a cell-free system has been demonstrated.¹⁹ This impressive achievement—conducted in an isolated environment, free from extraneous interactions with cellular proteins—is akin to ribosomal self-assembly in both complexity and compartmentalization.

Many subsequent observations of higher-level hierarchic assembly in the interactome recapitulate the early discovery of ribosomal self-assembly, underscoring the notion that the cell can be viewed as an “elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines.”⁷ For example, protein synthesis is spatially and temporally regulated in the cell. About three-quarter of mRNA molecules have non-random cellular localization,²⁰ ensuring that many proteins are made where they are needed, and the sequenced timing of their expression is apparent from the correlation between interaction and expression profiles in yeast.²¹ Also, there is a range of spatial signals that target proteins to functionally relevant cellular sites of interaction, such as the nuclear export signal²² or the endoplasmic reticulum retrieval signal.²³ In essence, a complicated cellular sorting/trafficking and assembly system, made up of membranous organelles, receptors, membrane translocation devices, cytoskeletal tracks, motor proteins, and accessory chaperones guides the proper compartmentalization, localization, and assembly of proteins in the cell.^{24–26} Here, we show that in the absence of energy even this well developed infrastructure would be insufficient to account for the generation of the interactome, which requires a continuous expenditure of energy to maintain steady state.

Limitations of Spontaneous Assembly from Isolated Proteins

Based on these observations that are consistent with hierarchic self-assembly carefully guided by spatial

and temporal signals, it may seem that the interactome can—and would—form spontaneously from its isolated components. In other words, there would be a way to “unboil” the denatured cell, that is, to promote its assembly from a disassembled state, akin to refolding a denatured protein.¹ However, several points suggest that this view is overly simple.

First, even spontaneous (re)folding, typical of small proteins, is often irreversible in larger aggregation-prone proteins. The problem is far more severe in the crowded environment of the cell, where many proteins require chaperones and recombinant proteins tend to aggregate. It is known that chaperone-assisted folding is an energy-requiring process, but the prevailing interpretation is that the chaperone only acts as a catalyst that facilitates formation of the folded state of the protein that could have been attained spontaneously under dilute solution conditions. However, if extrapolated to a macromolecular complex, this view may be too simplistic. The ability of proteins to form prions²⁷ and amyloids²⁸ demonstrates that the physiologically relevant folded state is probably not one of maximum stability, although it may be the most kinetically accessible metastable state. Consequently, Anfinsen’s thermodynamic hypothesis¹ comes with a qualifying corollary, one that may well take precedence in the interactome. Upon initial consideration, misfolding (misassembly) might seem to be an unlikely outcome in the spontaneous assembly of macromolecular complexes, such as the ribosome, but this impression cannot withstand closer scrutiny. Successful self-assembly conditions had to be carefully worked out for the bacterial ribosome,^{17,29} and corresponding conditions are unattainable for the eukaryotic ribosome, which requires as many as 200 accessory proteins *in vivo*, most of them essential.³⁰ Even less-complicated complexes, such as the nucleosome³¹ or the proteasome,³² require assisted assembly in the cell. Such examples illustrate a basic difference between the *in vitro* assembly of 20 isolated components, each introduced in a specific order under controlled conditions, and their *in vivo* assembly amidst a sea of competing components. The underlying problem is well illustrated by calculations showing that physiological interactions are not necessarily the energetically dominant possibilities in the interactome.³³

Over and above combinatorial complexity, there is a fundamental “chicken-and-egg” dilemma: correct interpretation of assembly signals and pathways may require a prior network of interacting proteins, that is, the interactome itself. For example, mRNA localization requires the cytoskeleton, along which transport can proceed.²⁰ In turn, the cytoskeleton requires prior organization, such as the microtubule-organizing centers (MTOCs), for proper assembly,³⁴ and transport along the cytoskeleton requires protein motors, large complexes themselves. Again, the

Figure 2. The interactome cannot assemble from its constituent proteins. Due to the incomprehensible number of possible realizations and the energy needed for all assembly mechanisms, we suggest a discontinuity between a viable interactome and its isolated components, by postulating three conceptually distinct zones of differing complexity. *Zone 1* (order, native state) corresponds to the viable interactome in steady state, where fluctuations are completely reversible. Excursion to *zone 2* (disorder) due to stress, disease, mutations, and large physiological rearrangements can be reversed at the expense of energy. *Zone 3* (chaos) is vast and undifferentiated, representing a lethal level of disorganization brought about by extreme stress: current excursions into this zone are irreversible.

nuclear export signal requires the presence and operation of the nuclear pore complex for proper operation.³⁵ Although cellular function depends upon the “elaborate network of interlocking assembly lines,”⁷ it cannot be established in the absence of its own prior formation, a conundrum at the crux of self-replicating life. In addition, the operation of all these machines requires a continuous input of energy, and therefore it is not feasible that the end result (i.e., the functional interactome) could maintain steady-state conditions in an energy-independent fashion.

Perhaps the most profound conclusion to be drawn from our calculations of combinatorial complexity is that the emergent interactome could not have self-organized spontaneously from its isolated protein components. Rather, it attains its functional state by templating the interactome of a mother cell and maintains that state by a continuous expenditure of energy. In the absence of a prior framework of existing interactions, it is far more likely that combined cellular constituents would end up in a non-functional, aggregated state, one incompatible with life. Even the recent successful creation of an artificial bacterial cell³⁶ only demonstrates that synthetic genetic material can be transplanted into the cytoplasm (i.e., the viable interactome) of a very

closely related bacterium. The spontaneous origination of a *de novo* cell has yet to be observed; all extant cells are generated by the division of pre-existing cells that provide the necessary template for perpetuation of the interactome.

To illustrate the discontinuity between a viable interactome and its isolated components, we postulate a minimum of three conceptually distinct zones of differing complexity (Fig. 2):

- (i) *Zone 1* (order, native state) corresponds to the viable interactome under normal, physiological conditions, defined as a collection of closely related states generated by thermal fluctuations (dissociations/associations) around an equilibrium state. In this zone, spontaneous assembly dominates and fluctuations are completely reversible.
- (ii) *Zone 2* (disorder) is defined by reversible excursions from *zone 1* owing to stress, disease, mutations, large physiological rearrangements such as cell division, and so forth. In this zone, there is somewhat less reversibility, but excursions here can be reversed at the expense of energy by a combination of pathways, compartments, and chaperones.
- (iii) *Zone 3* (chaos) is vast and undifferentiated, representing the lethal level of disorganization brought about by extreme stress, a level that cannot be reversed by self-assembly mechanisms. An excursion into this zone is not reversible. Whereas *zone 1* may represent a steady state in some abstract interaction space, there is no mechanism for reaching it from *zone 3* in a biologically relevant time frame.

An implicit consequence of this conceptual model is that life would have traversed *zone 3* at least once. Presumably, early-earth life forms originated through an accumulation of changes of ever increasing complexity, resulting eventually in photosynthetic prokaryotes. In this sense, extant assembly-pathways almost certainly echo their own evolutionary history, that is, a protein is guided to its cellular destination along a route that was established at an earlier time and subsequently fortified by other, similarly developed, interdependent cellular processes. Supporting evidence for this conclusion is provided by a recent mass-spectroscopy study of the conservation and formation of the quaternary structure of protein homomers.³⁷ This study confirmed that structure alone is sufficient to infer both the evolutionary and physical path of subunit assembly, an example of “ontogeny recapitulates phylogeny” at the cellular level.

Implications

Misfolding errors in proteins can cause assembly errors that propagate across cellular pathways, with

opportunities for malfunction at each successive level. At the level of individual molecules, protein misfolding errors can produce non-native aggregated states, with deleterious consequences to the cell.²⁸ At the level of a pathway, assembly errors can lead to disease-causing mis-localizations and mis-interactions. Typically, such processes are interrelated: misfolding can result in mis-interactions that terminate in an aggregated dead-end.²⁸ Such entanglement is well illustrated by prions, infectious proteins that can propagate in the cell by a self-sustaining autocatalytic conformational change, resulting in the formation of amyloid.²⁷ From the perspective of a protein, the prion catastrophe is a misfolding disease, while from the perspective of the interactome, it is a mis-interaction disease.

It follows that there are many opportunities for disease-associated mutations which can cause mis-localization and mis-interaction of proteins. Whereas most monogenic disease-causing mutations promote destabilization of protein structure,³⁸ such mutations can also affect protein expression, translation, transport, and localization.³⁹ An instructive example is primary hyperoxaluria (abnormally high oxalate excretion). Approximately, one-third of such cases are associated with a protein-sorting defect in hepatic L-alanine:glyoxylate aminotransferase (AGT). The enzyme is peroxisomal under normal circumstances, but in disease it is mistargeted to mitochondria by mutations in its N-terminal region, which generate an aberrant mitochondrial targeting sequence that is misinterpreted by the mitochondrial protein import machinery.⁴⁰

Our view of the interactome may also provide insight into chaperone action, which also functions at both the protein folding and protein assembly level. Indeed, the term “chaperone” was actually coined for a protein-assisted assembly of the nucleosome.³¹ The existence of protein-assisted stabilization prompts the notion of a complementary process of protein-inhibited destabilization, such as the recently proposed “nanny” proteins, which prevent degradation and improper interactions of their partner proteins.⁴¹ The chaperone system, which can stabilize proteins and pathways against stress, is itself subject to stress, and its breakdown under “overload” conditions⁴² may also contribute to disease.

The inability of the interactome to self-assemble *de novo* imposes limits on efforts to create artificial cells and organisms, that is, synthetic biology. In particular, the stunning experiment of “creating” a viable bacterial cell by transplanting a synthetic chromosome into a host stripped of its own genetic material³⁶ has been heralded as the generation of a synthetic cell⁴³ (although not by the paper’s authors). Such an interpretation is a misnomer, rather like stuffing a foreign engine into a Ford and

declaring it to be a novel design. The success of the synthetic biology experiment relies on having a recipient interactome in *zone 1* (or, worst case, *zone 2*) that has high compatibility with donor genetic material. The ability to synthesize an actual artificial cell using designed components that can self-assemble spontaneously still remains a distant challenge.

Acknowledgments

P.T. is indebted to Dr. and Mrs. Kalman Tompa for helpful discussions on the combinatorial aspects of the interactome and Dr. Éva Tüdös (Institute of Enzymology, Hungarian Academy of Sciences, Budapest, Hungary) for help in calculating large factorials.

References

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181:223–230.
2. Levinthal C, How to fold graciously. In: DeBrunner JTP, Munck E, Eds. (1969) *Mossbauer spectroscopy in biological systems*. Allerton House, Monticello, Illinois: University of Illinois Press, pp. 22–24.
3. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. *Nat Struct Biol* 4:10–19.
4. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440:631–636.
5. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Ristone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O’shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440:637–643.
6. Haber E, Anfinsen CB (1961) Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J Biol Chem* 236:422–424.
7. Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92:291–294.
8. Kell DB, Welch GR (1991) No turning back: reductionism and biological complexity. *Times Higher Educational Supplement*, 9th August:p. 15.
9. Flory PJ (1969) *Statistical mechanics of chain molecules*. New York: Wiley.
10. Baldwin RL, Zimm BH (2000) Are denatured proteins ever random coils. *Proc Natl Acad Sci USA* 97:12391–12392.
11. Pappu RV, Srinivasan R, Rose GD (2000) The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc Natl Acad Sci USA* 97:12565–12570.

12. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
13. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein–protein recognition sites. *J Mol Biol* 285:2177–2198.
14. Baldwin RL, Rose GD (1999) Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci* 24:26–33.
15. Baldwin RL, Rose GD (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci* 24:77–83.
16. Erdős P, Rényi A (1960) On the evolution of random graphs. *Publ Math Inst Hungar Acad Sci* 5:17–61.
17. Held WA, Ballou B, Mizushima S, Nomura M (1974) Assembly mapping of 30 S ribosomal proteins from *Escherichia coli*. Further studies. *J Biol Chem* 249:3103–3111.
18. Held WA, Nomura M (1975) *Escherichia coli* 30 S ribosomal proteins uniquely required for assembly. *J Biol Chem* 250:3179–3184.
19. Molla A, Paul AV, Wimmer E (1991) Cell-free, *de novo* synthesis of poliovirus. *Science* 254:1647–1651.
20. Lecuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 131:174–187.
21. Ge H, Liu Z, Church GM, Vidal M (2001) Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat Genet* 29:482–486.
22. Wen W, Meinkoth JL, Tsien RY, Taylor SS (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell* 82:463–473.
23. Nilsson T, Warren G (1994) Retention and retrieval in the endoplasmic reticulum and the Golgi apparatus. *Curr Opin Cell Biol* 6:517–521.
24. Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA (2006) Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem* 75:655–680.
25. Mellman I, Nelson WJ (2008) Coordinated protein sorting, targeting and distribution in polarized cells. *Nat Rev Mol Cell Biol* 9:833–845.
26. Bashor CJ, Horwitz AA, Peisajovich SG, Lim WA (2010) Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys* 39:515–537.
27. Prusiner SB (1998) Prions. *Proc Natl Acad Sci USA* 95:13363–13383.
28. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
29. Held WA, Mizushima S, Nomura M (1973) Reconstitution of *Escherichia coli* 30 S ribosomal subunits from purified molecular components. *J Biol Chem* 248:5720–5730.
30. Strunk BS, Karbstein K (2009) Powering through ribosome assembly. *RNA* 15:2083–2104.
31. Laskey RA, Honda BM, Mills AD, Finch JT (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275:416–420.
32. Bedford L, Paine S, Sheppard PW, Mayer RJ, Roelofs J (2010) Assembly, structure, and function of the 26S proteasome. *Trends Cell Biol* 20:391–401.
33. Wass MN, Fuentes G, Pons C, Pazos F, Valencia A (2011) Towards the prediction of protein interaction partners using physical docking. *Mol Syst Biol* 7:469.
34. Nigg EA, Raff JW (2009) Centrioles, centrosomes, and cilia in health and disease. *Cell* 139:663–678.
35. Patel SS, Belmont BJ, Sante JM, Rexach MF (2007) Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* 129:83–96.
36. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329:52–56.
37. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA (2008) Assembly reflects evolution of protein complexes. *Nature* 453:1262–1265.
38. Yue P, Li Z, Moulton J (2005) Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol* 353:459–473.
39. Shastri BS (2009) SNPs: impact on gene function and phenotype. *Methods Mol Biol* 578:3–22.
40. Purdue PE, Allsop J, Isaya G, Rosenberg LE, Danpure CJ (1991) Mistargeting of peroxisomal L-alanine:glyoxylate aminotransferase to mitochondria in primary hyperoxaluria patients depends upon activation of a cryptic mitochondrial targeting sequence by a point mutation. *Proc Natl Acad Sci USA* 88:10900–10904.
41. Tsvetkov P, Reuven N, Shaul Y (2009) The nanny model for IDPs. *Nat Chem Biol* 5:778–781.
42. Csermely P (2001) Chaperone overload is a possible contributor to 'civilization diseases'. *Trends Genet* 17:701–704.
43. Wade N (2010) Researchers say they created a 'synthetic cell'. *The New York Times*. New York.