How gene duplication diversifies the landscape of protein oligomeric state and function

Mallik, Saurav; Tawfik, Dan S.; Levy, Emmanuel

How to cite

MALLIK, Saurav, TAWFIK, Dan S., LEVY, Emmanuel. How gene duplication diversifies the landscape of protein oligomeric state and function. In: Current opinion in genetics & development, 2022, vol. 76, p. 1–9. doi: 10.1016/j.gde.2022.101966

This publication URL: https://archive-ouverte.unige.ch/unige:176878
Publication DOI: 10.1016/j.gde.2022.101966

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

How gene duplication diversifies the landscape of protein oligomeric state and function

Saurav Mallik[1,2,*], Dan S Tawfik[2,#,$] and Emmanuel D Levy[1,†,‡]

Oligomeric proteins are central to cellular life and the duplication and divergence of their genes is a key driver of evolutionary innovations. The duplication of a gene coding for an oligomeric protein has numerous possible outcomes, which motivates questions on the relationship between structural and functional divergence. How do protein oligomeric states diversify after gene duplication? In the simple case of duplication of a homo-oligomeric protein gene, what properties can influence the fate of descendant paralogs toward forming independent homomers or maintaining their interaction as a complex? Furthermore, how are functional innovations associated with the diversification of oligomeric states? Here, we review recent literature and present specific examples in an attempt to illustrate and answer these questions.

Addresses
[1] Department of Chemical and Structural Biology, The Weizmann Institute of Science, Rehovot 7610001, Israel
[2] Department of Biomolecular Sciences, The Weizmann Institute of Science, Rehovot 7610001, Israel

Corresponding authors: Saurav Mallik (saurav.mallik@weizmann.ac.il), Emmanuel D Levy (emmanuel.levy@weizmann.ac.il)
[‡] Twitter account: @ElevyLab
[*] ORCID: 0000–0002-9390–0715
[#] Deceased, May 04, 2021.
[$] ORCID: 0000–0002-5914–8240
[†] ORCID: 0000–0001-8959–7365

Current Opinion in Genetics & Development 2022, **76**:101966

This review comes from a themed issue on **Evolutionary Genetics**

Edited by **Christian Landry** and **Gianni Liti**

For complete overview of the section, please refer to the article collection, "Evolutionary Genetics"

Available online 22th August 2022

https://doi.org/10.1016/j.gde.2022.101966

Introduction

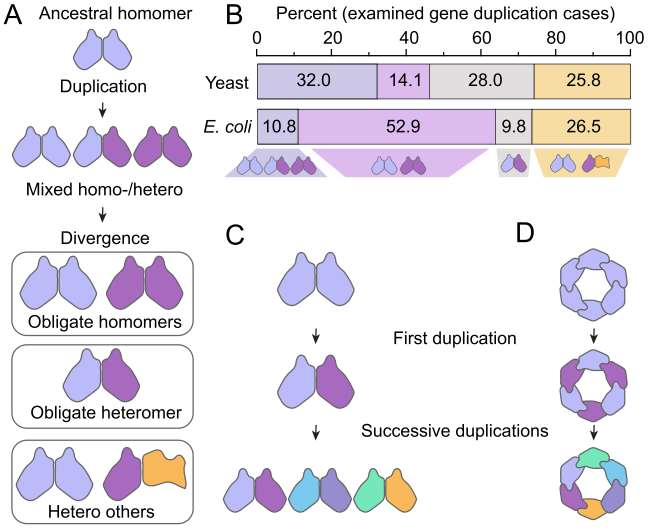
One of the earliest accounts of gene duplication appeared in the 1930s, when doubling of a chromosomal band in *Drosophila melanogaster* was associated with extreme reduction in eye size [1]. Since then, this phenomenon has been recognized as a ubiquitous source of functional novelty across the Tree of Life [2–4]. Indeed, duplicated genes (typically termed as paralogs [5]) initially retain the overall sequence and structural features of the ancestral single-copy gene, but over time, they also diverge, either symmetrically [6] or asymmetrically [7–9], that is, at similar or different rates, and can bring about functional innovations.

Functional innovations are intimately associated with changes in protein structure. In this context, it is important to consider the tertiary and quaternary structure of paralogous proteins when examining the evolution of their function. For example, after duplication, a protein may acquire a novel targeting sequence and perform the ancestral function in a new subcellular compartment [10]. Alternatively, sequence changes in paralogs by point-mutations or by insertion-deletions can lead to functional innovations [11–21] such as subdivision of the ancestral function among the duplicated copies [11], new enzymatic activities [12,13], regulatory modes [12,14], or interaction partners [15,16]. In this review, we survey and classify how changes in quaternary structure relate to the functional diversification of paralogous proteins. First, we present how gene duplication can diversify protein oligomeric states in the context of homomers. Second, we examine molecular and regulatory changes associated with such diversification. Finally, we illustrate how functional innovations can be coupled to changes in oligomeric state.

Gene duplication can drive the emergence of new protein oligomeric states

The duplication of a gene opens numerous possible routes of functional innovation. Once fixed, the functional redundancy conferred by the two copies relaxes the selection pressure on them, allowing the diversification of their sequence, structure, oligomeric state, and in turn, function. A homomeric protein is perhaps the simplest model system to understand these events (Figure 1a). In principle, right after duplication of a gene encoding a homomeric protein, a statistical mixture of homo- and heteromeric complexes would form (*Mixed*). Upon further divergence, various scenarios may occur [22–25]. The statistical mixture may continue to exist [23,26]. Alternatively, the two copies might lose the capacity to cross-interact, resulting in two independent homomers (*Obligate homo*); or, they might lose the capacity to self-interact, resulting in a single heteromeric complex (*Obligate hetero*). The interaction pattern might also diverge asymmetrically, in which one copy retains

Figure 1



Evolution of oligomeric state among homomeric proteins upon gene duplication. **(a)** Duplication of a gene encoding a homomeric protein. Before divergence, interfaces are compatible, so a mixture of homo- and heteromeric complexes can coexist. As the two copies diverge, three outcomes may follow: (i) each duplicate self-interacts, which we refer to as "*Obligate homomers*". (ii) Both duplicates can interact to form a heteromeric complex that we refer to as "*Obligate heteromer*". (iii) The duplicates follow different paths, with, for example, one copy remaining a homomer, while the other gains a new interaction. **(b)** The relative frequency of the four different outcomes as observed in an analysis of *E. coli* and *S. cerevisiae*'s interaction networks [22]. Homomeric interactions are dominant in *E. coli*, whereas cross-reacting paralogs are dominant in *S. cerevisiae*. **(c)** After duplication of a gene coding for a homomeric protein, the two paralogs can fix as an *Obligate heteromeric* complex. Following such an event, further duplications events may expand the family. **(d)** Gene duplication of ring-like homomers often involves the paralogous copies co-assembling into the same ring, making the ring heteromeric.

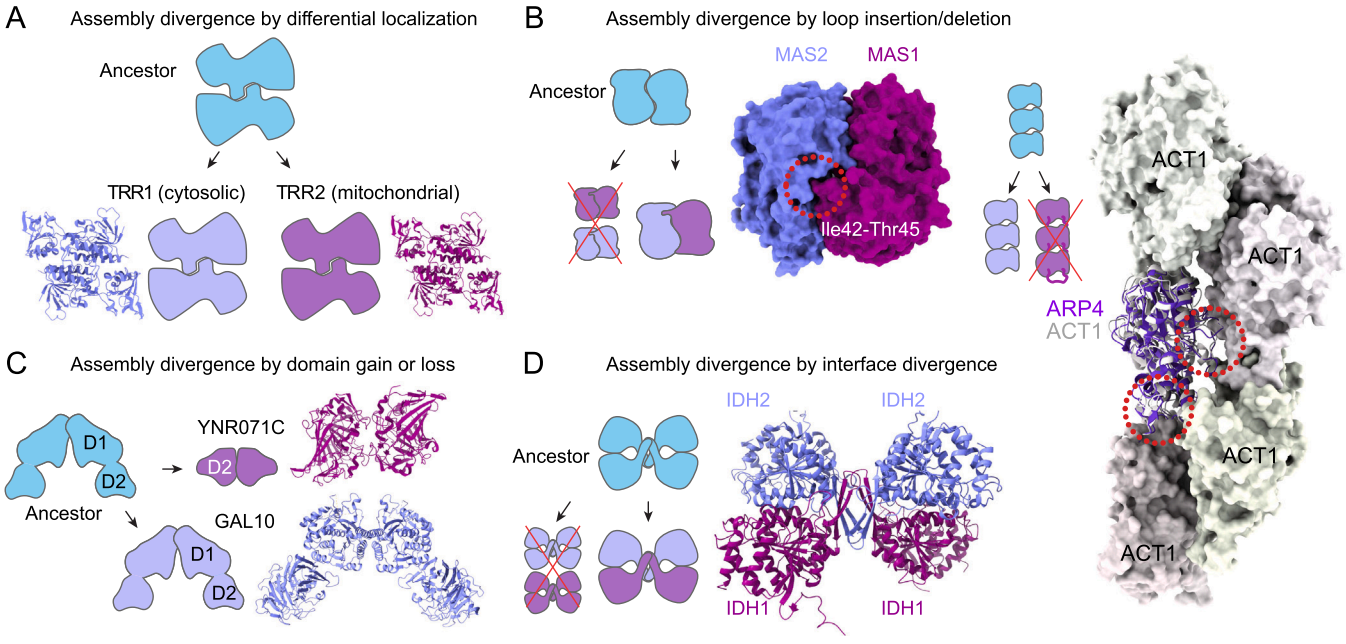
the ancestral homomeric state, whereas the other evolves heteromeric interactions with new partners (*Hetero others*). The relative frequency of these four fates varies, depending on the organism and the symmetry of the homomer [22].

The retention of homomeric interactions is the dominant fate in bacteria, where *Obligate homomers* account for more than half of the examined cases (Figure 1b). In this scenario, paralogous copies tend to be substantially different in both sequence and primary functions [27]. Though the underlying principles await discovery, given that in bacteria horizontal gene transfers are more than ~50 times more frequent than gene duplications [28], many of these *Obligate homomers* may not be *bona fide* paralogs. Instead, they might include a horizontally acquired gene of the same family. In this case, the homologous proteins may significantly differ in sequence and would not cross-interact [29]. In contrast, horizontal transfers are rare in eukaryotes and nearly two-thirds of their homomeric gene duplications yield cross-interacting paralogs (*Obligate hetero+Mixed* dominate, Figure 1b). When an ancestral homomer becomes an *Obligate heteromer*, the primary function before and after the duplication tends not to change. Phosphofructokinase (*PFK*, E.C. 2.7.1.11) [30] forms a homo-octamer in fission yeast. In the lineage of baker's yeast, it underwent a gene duplication event resulting in a hetero-octameric *PFK1/2* complex that retained the ancestral enzymatic activity (E.C. 2.7.1.11) [31]. Though the primary function is conserved, the sequence and structure of each copy can evolve toward the emergence or fine-tuning of secondary function(s) [3,32], or asymmetric binding of partners [23]. This route underlies the emergence of regulatory modes in many eukaryotic enzymes typically absent among their prokaryote orthologs [3].

We saw how gene duplication can yield an *Obligate heteromer*. Interestingly, once two paralogs are fixed as a heteromeric complex, their further duplications might generate an array of heteromeric complexes (Figure 1c). This phenomenon underlies the expansion of many generalist enzyme families (those that work on a broad range of substrates), including transporters, chaperones, or endonucleases. Perhaps, the most intriguing example is the *HSP70* chaperone family. A gene duplication in the *HSP70* family during the emergence of eukaryotes diverged into the *HSP110* cochaperone that lost the canonical chaperone activity, but specialized to become an ATP-exchange factor for *HSP70* [33]. This functional collaboration between *HSP70* and *HSP110* has been retained in their successive duplications, and with the ever-increasing demand for chaperones in eukaryotes, has rendered an array of *HSP70–HSP110* pairs that work in different conditions, subcellular compartments, tissues, and also on different types of substrates [33,34].

For ring-like homomers, paralogous copies typically, but not always, sequester in the same ring, thus making the ring heteromeric (*Obligate hetero* is the dominant fate, Figure 1d). Examples include ATPase rings [35,36], the chaperonin TCP complex [37], the RNA exosome PH ring [38], Lsm/Sm homomers [39], and the V_0 ring of V-ATPase proton pump [40] that have become heteromeric in eukaryotes following one or several rounds of gene duplication. It is interesting to note that for numerous complexes with ring-like (cyclic) symmetry, the patterns of gene duplications appear to be discrete. For example, the proteasome alpha ring (C7 symmetry) is formed by one protein in archaea and seven paralogous proteins in eukaryotes, but intermediate numbers (e.g., 2, 3, 4, 5, or 6 paralogous proteins) have not been observed, presumably due to symmetry constraints [23,41]. Interestingly, after the proteasomal alpha and beta rings were fully heteromeric, whole-genome duplication(s) at the root of jawed vertebrates gave birth to the

Figure 2



Origins of interaction divergence between paralogs. **(a)** Incompatibility may emerge from different subcellular localization of the paralogous copies, such as the case of yeast *Obligate homomers TRR1* (PDB code 3DX8) and *TRR2* (UniProt code P38816). **(b)** Insertions and deletions at the interface can drive incompatibility, as in the case of yeast *MAS1/2* heterodimer where an insertion (Ile42–Thr45) in *MAS1* appears to block its self-interaction. In another example, two insertions, Gln211–Asp231 and Asp299–Thr365, appear to hinder filament formation of *ARP4*, which instead heteromerizes with its paralog *ACT1*. Here, the structure depicts an *ACT1* filament (white, PDB code 6BNO) onto which *ARP4* (purple, PDB code 5NBM) was superposed. **(c)** The gain/loss of the domain mediating the ancestral homomeric interaction is another mechanism for incompatibility. Such a scenario is seen in *Obligate homomers GAL10* (PDB code 1Z45) and *YNR071C* (PDB code 1YGA). **(d)** Accumulation of amino acid substitutions can alter the interface specificity and bring about incompatibility. This scenario occurs in the yeast heteromer *IDH1/2* (PDB code 3BLV). We note that the *IDH* complex shows A4B4 stoichiometry and only half of the complex (A2B2) is shown here for simplicity. The physiological relevance of the quaternary structures highlighted in this figure was inferred based on annotations from the QSBIO.org database [83].

immunoproteasome — differing from the more common "constitutive proteasome" by a few paralogous subunits and specialized to operate during oxidative stress [42].

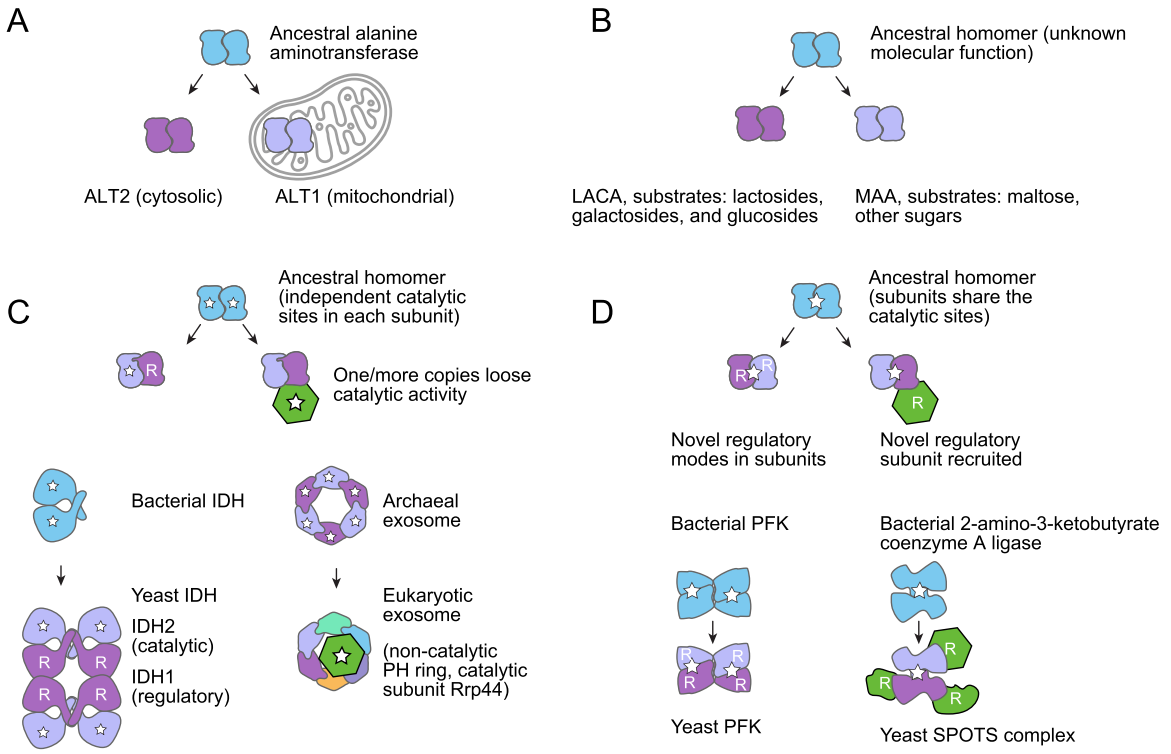
Molecular and regulatory changes associated with interaction divergence

Immediately after a duplication event, the interfaces of both duplicates are identical and therefore compatible with preduplication interaction modes. As the duplicates diverge, at least four distinct mechanisms can drive interaction divergence between the paralogous copies. First, divergence can emerge from different subcellular localizations and/or different modes of expression [43] of the two paralogs (e.g., at different stages of the cell cycle or in different tissues, Figure 2a). Until their interfaces diverge by random genetic drift, such recently duplicated *Obligate homomers* can still cross-interact even though they may not encounter each other *in vivo*. Examples include cytoplasmic and mitochondrial homomeric thioredoxin reductases in yeast (*TRR1* and *TRR2*) that emerged during the yeast whole-genome duplication [44]. These two paralogs are 85% identical in sequence and were observed to cross-interact *in vitro* [45].

Second, insertions/deletions (InDels) at the interface can introduce steric hindrance and create interface incompatibility (Figure 2b) [46,47]. For example, *MAS1* and *MAS2* assemble into the *Obligate heterocomplex* yeast mitochondrial peptidase. The two genes encoding these proteins share a common bacterial ancestor, which forms a homodimer (M16 family peptidase) [22]. The heteromeric interface of the *MAS1/2* complex shares a similar structure to that of the bacterial M16 peptidases and an insertion in *MAS1* (Ile42–Thr45) would hinder its homodimerization. In another example, yeast actin (*ACT1*) forms a homomeric filament [48]. Its paralog, actin-related protein 4 (*ARP4*), does not form homomeric filaments and instead heterodimerizes with *ACT1*. The *ARP4–ACT1* heteromer is a component of chromatin remodeling complexes [49]. As depicted in Figure 2b, two insertions, Gln211–Asp231 and Asp299–Thr365 appear incompatible with self-interaction interfaces found in actin filaments.

Third, divergence can originate in the gain or loss of a domain mediating or preventing an interaction (Figure 2c). An interesting case is that of yeast paralogs *GAL10*

Figure 3



Evolution of the oligomeric state in connection to functional innovations after gene duplication. **(a)** The duplication of a homomer yields two independent *Obligate homomers* sharing the same primary function but different localizations: yeast *ALT1* (mitochondrial) and *ALT2* (cytosolic). **(b)** The duplication of a homomer yields two independent *Obligate homomers* exhibiting different primary functions: *E. coli LACA* and *MAA.* The former is specific for acetylation of galactosyl units [60], while the latter catalyzes acetylation of glucosyl units [61]. **(c)** *Obligate heteromers* with independent active sites can retain the ancestral function (white star), or maintain the function through additional subunits. In the case of yeast *IDH1/2* heteromer, *IDH1* has lost its catalytic activity and became a regulatory subunit [65]. In the case of the exosome, all paralogs forming the PH ring lost the RNAse activity and a new subunit assumed that function. **(d)** Among *Obligate heteromers* where the active site is contributed by both subunits, both subunits are expected to maintain their function. In this case, regulatory activities may emerge in a separate domain, as in the case of yeast *PFK1/2* heteromer [66]. In a different scenario, an *Obligate heteromer* such as the yeast SPOTS complex recruits new regulatory subunits.

and *YNR071C*. As annotated in Pfam [50], *GAL10* comprises an N-terminal GDP-mannose-4,6-dehydratase domain. This domain mediates the homomeric interaction and also contacts the C-terminal Aldose-1-epimerase domain [51]. In contrast, *YNR071C* has lost the N-terminal domain according to Pfam annotations, and comprises only an Aldose-1-epimerase domain, which mediates the homomeric interaction seen in the crystal structure [51]. The cross-interaction between the two paralogs via the Aldose-1-epimerase domain is hindered by the GDP-mannose-4,6-dehydratase domain in *GAL10*.

Last, the accumulation of amino acid substitutions at the interface can also bring about incompatibility and divergence (Figure 2d). Indeed, a few mutations can create or abolish interaction interfaces [52–54]. The mitochondrial NAD-dependent isocitrate dehydrogenase complex in yeast represents such an example. It is an *Obligate heteromer* of *IDH1* and *IDH2* that are 42% identical in sequence but comprise no domain gain, loss, or InDels at the interface. *IDH1* and *IDH2* do not exhibit homomeric interactions *in vitro* [22], so this incompatibility appears to stem from the accumulation of amino acid substitutions at their interface.

Functional innovations are coupled with the diversification of oligomeric states

Functional innovations after gene duplication can be coupled with the divergence of the oligomeric state. For example, *Obligate homomers*, at least in principle, evolve independently without the constraints of paralog interference and can therefore diverge in function more than heteromers of paralogs [55,56]. Functional divergence of *Obligate homomers* can be classified into two categories. In the first category, the primary functions of the duplicated copies do not change but one or both become specialized to operate in different conditions (i.e., subfunctionalization, Figure 3a). For example, the duplicated copy might diverge and perform the same function

in a different subcellular compartment. Yeast *ALT1* and *ALT2* represent such an example: both synthesize pyruvate from L-alanine, but the former localizes in mitochondria and the latter in the cytoplasm [57,58]. Subfunctionalization can also be associated with different expression profiles, as in the case of *Escherichia coli LYSS* and *LYSU*. Both catalyze lysine–tRNA loading, but the former is constitutively expressed, whereas the latter is heat-induced [59]. In the second category, the primary function itself can diverge (neofunctionalization, Figure 3b). For enzymes, the divergence of primary function could mean changes in substrate specificity, whereby both paralogs perform the same chemistry on different substrates [21]. Examples include *E. coli LACA*, which catalyzes the acetylation of galactosyl units [60]; its paralog, *MAA*, catalyzes the acetylation of glucosyl units [61]. In other cases, one copy may lose the enzyme activity altogether. Examples include yeast *Obligate homomers LPD1* and *IRC15* (duplicated during the whole-genome duplication [62]). The former is a lipoamide dehydrogenase enzyme [63], while the latter has lost its enzymatic activity and regulates microtubule dynamics [64].

For *Obligate heteromers*, there is little scope for the divergence of the primary function because before and after the duplication there is only one complex carrying out the ancestral function. However, since two proteins are now mediating the primary function, selection pressure may relax and allow the emergence of secondary functions such as allosteric regulation [3]. Such allosteric regulation can emerge in different ways, depending on the catalytic (in)dependence of the individual subunits.

A homomeric enzyme can exhibit structurally independent catalytic sites in each subunit, or catalytic sites may be shared across multiple subunits. In the former case, once there is an *Obligate heteromeric* complex after gene duplication, regulatory activities can emerge in the following ways. First, one paralog might lose the catalytic activity and instead become a regulatory subunit of the other (Figure 3c), as previously discussed with the *HSP70–HSP110* pair. Another example is the yeast mitochondrial NAD-dependent *IDH1/2* complex (duplicated and diverged from a bacterial homomeric *IDH complex* [22]), in which *IDH1* lost the catalytic activity and became a regulatory subunit of *IDH2* [65]. In a more extreme case, all paralogous subunits may lose the catalytic activity and an external subunit may be recruited for catalysis (Figure 3c). Such a scenario occured with the bacterial homomeric RNase PH ring. After several rounds of gene duplications, it became heteromeric in eukaryotes (RNA exosome) and in that process all paralogous copies lost their catalytic activities. The exonuclease function in contemporary eukaryotic RNA exosome is mediated by a newly recruited, eukaryote-specific Rrp44 subunit [38].

For homomeric enzymes with catalytic sites contributed by multiple subunits, the scope of evolving allosteric regulation is more limited because each subunit is constrained to retain the primary function [85]. In these cases, regulatory activities can emerge within each subunit (Figure 3d). A curious example is yeast ATP-dependent 6-phosphofructokinase, an *Obligate heterocomplex* of *PFK1* and *PFK2*. In this complex, the interacting N-terminal domains of the two subunits share the catalytic sites, whereas the C-terminal domains have evolved allosteric regulation [66]. Alternatively, the newly emerged heteromeric enzyme can act as a catalytic scaffold that recruits additional regulatory subunits (Figure 3d). Examples include the SPOTS complex in yeast, composed of catalytic subunits *LCB1* and *LCB2* (that share the serine palmitoyltransferase catalytic sites) and regulatory subunits *ORM1*, *ORM2*, *SCA1*, and *TSC3* [67,68].

Conclusion

The transition of protein oligomeric states upon gene duplication, once only studied and appreciated for a few prokaryotic homomeric ring-like complexes [35–39], can now be examined as a genome-scale phenomenon. A remarkable trend of this process is that it has transformed homomer-dominant prokaryotic proteomes into heteromer-dominant eukaryotic proteomes [22,69,70]. The spectrum of protein oligomeric-state diversification is wide, and each mode of diversification — combined with the symmetry of the oligomeric state and the catalytic independence of the subunits — constraints functional diversification differently. Conspicuous functional innovations, such as loss of enzyme activity or alteration of substrate specificity, are typically observed when the paralogous pairs form independent complexes and do not cross-react. Comparatively more subtle functional innovations, such as the emergence of allosteric regulation, are typically observed when the paralogous pairs co-assemble into the same complex. Whether or not, two paralogs, descendants of an ancestral homomer, would form a heteromeric complex, depends on post-duplication changes at the interface. These molecular changes in paralogous enzymes have likely played an important role in the emergence of eukaryotic cells with their many subcellular compartments.

Future directions

We saw that specific evolutionary processes govern the interactions and functional divergence of gene duplicates. While this review focuses on specific examples, we still lack a systematic genome-scale understanding of these processes within and between species. In particular, we highlight below four questions. Addressing these will be important to gain fundamental knowledge, but also for understanding the association between gene duplication and various complex diseases [71–73], including Down syndrome [74], Alagille syndrome [75], and cancer [73].

First, on what timescales do evolutionary events accumulate to drive structural and functional divergence? Timescales can vary broadly in different cases, and work will be needed to identify determinants associated with different rates. This point is also tied to a second question: to what extent is divergence driven by adaptive versus neutral processes [76–78]? In the former case, we can anticipate signatures of positive selection and a faster sequence divergence than in the latter. Interestingly, neutral drift is not necessarily expected to disrupt homomeric interactions [79]. In contrast, it can disrupt interactions between paralogs more frequently by altering co-expression and co-localization tendencies. Therefore, we expect that the retention of interactions between paralogs will more often be adaptive than the retention of homomeric interactions [22].

The expected balance of adaptive versus neutral evolution is itself tied to a third question: do functional innovations tend to exist pre duplication, or occur post duplication? Promiscuous activities can emerge in enzymes by neutral drift [76,77] and come to evolve independently in a duplicate. In this context, the new function may predate the gene duplication. Alternatively, the new function could evolve in the newly formed complex post duplication.

Finally, which of the two fates, *Obligate homo* or *Obligate hetero* emerges more frequently and why? *Obligate homomeric* pairs dominate bacterial proteomes, whereas *Obligate heteromers* dominate in eukaryotes [22]. Since a few mutations [52–54,84] or a single insertion–deletion [47,80] can alter the interaction specificity of protein interfaces, the sole effect of mutations on structure is unlikely to explain this dichotomy. Therefore, other factors will need to be investigated.

More generally, addressing questions will require comprehensive information on protein structure and assembly. In that respect, recent structure and interaction modeling methods based on deep learning [81,82] will be instrumental to perform such analyses on a phylogenetic scale.

CRediT authorship contribution statement

Saurav Mallik: Writing – original draft, visualization, **Emmanuel D Levy:** Writing – review and editing, visualization.

Conflict of interest

The authors declare that no conflict of interest exists.

Acknowledgments

S.M. dedicates this article to a friend, colleague, and mentor Dan S. Tawfik, who supervised part of this work. S.M. was supported by a VATAT postdoctoral fellowship, provided by Israel's Council for Higher Education Planning and Budgeting Committee. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 819318), by the Israel Science Foundation (grant no. 1452/18), by a research grant from A.-M. Boucher, by research grants from the Estelle Funk Foundation, the Estate of Fannie Sherr, the Estate of Albert Delighter, the Merle S. Cahn Foundation, Mrs. Mildred S. Gosden, the Estate of Elizabeth Wachsman, the Arnold Bortman Family Foundation.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest.

1. Bridges CB: **The bar "gene" a duplication**. *Science* 1936, **83**:210-211.

2. Ohno S: Evolution by Gene Duplication. Springer Berlin; 1970, https://doi.org/10.1007/978-3-642-86659-3

3. Conant GC, Wolfe KH: **Turning a hobby into a job: how duplicated genes find new functions**. *Nat Rev Genet* 2008, **9**:938-950.

4. Zhang J: **Evolution by gene duplication: an update**. *Trends Ecol Evol* 2003, **18**:292-298.

5. Jensen RA: **Orthologs and paralogs – we need to get it right**. *Genome Biol* 2001, **2**:INTERACTIONS1002.

6. Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV: **Selection in the evolution of gene duplications**. *Genome Biol* 2002, **3**:RESEARCH0008.

7. Holland PWH, Marlétaz F, Maeso I, Dunwell TL, Paps J: **New genes from old: asymmetric divergence of gene duplicates and the evolution of development**. *Philos Trans R Soc Lond B Biol Sci* (1713) 2017, **372**:20150480, https://doi.org/10.1098/rstb.2015.0480

8. Wagner A: **Asymmetric functional divergence of duplicate genes in yeast**. *Mol Biol Evol* 2002, **19**:1760-1768.

9. Scannell DR, Wolfe KH: **A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast**. *Genome Res* 2008, **18**:137-147.

10. Marques AC, Vinckenbosch N, Brawand D, Kaessmann H: **Functional diversification of duplicate genes through subcellular adaptation of encoded proteins**. *Genome Biol* 2008, **9**:R54.

11. Lynch M, Force A: **The probability of duplicate gene preservation by subfunctionalization**. *Genetics* 2000, **154**:459-473.

12. Hittinger CT, Carroll SB: **Gene duplication and the adaptive evolution of a classic genetic switch**. *Nature* 2007, **449**:677-681.

13. • Cople SD: **Evolution of new enzymes by gene duplication and divergence**. *FEBS J* 2020, **287**:1262-1283.
Describes how new enzyme activities can emerge by gene duplication of an ancestral promiscuous enzyme and by successive fixation of mutations in the two copies.

14. •• Jayaraman V, Lee DJ, Elad N, Vimer S, Sharon M, Fraser JS, Tawfik DS: **A counter-enzyme complex regulates glutamate metabolism in Bacillus subtilis**. *Nat Chem Biol* 2022, **18**:161-170.
An intriguing study of a counter-enzyme complex (i.e. an assembly of metabolic enzymes catalyzing opposite rather than sequential reactions) in Bacillus involving a recently duplicated pair of paralogs.

15. Diss G, Gagnon-Arsenault I, Dion-Coté A-M, Vignaud H, Ascencio DI, Berger CM, Landry CR: **Gene duplication can impart fragility, not robustness, in the yeast protein interaction network**. *Science* 2017, **355**:630-634.

16. Veron AS, Kaufmann K, Bornberg-Bauer E: **Evidence of interaction network evolution by whole-genome duplications: a**

case study in MADS-box proteins. *Mol Biol Evol* 2007, **24**:670-678.

17. Latham JA, Chen D, Allen KN, Dunaway-Mariano D: **Divergence of substrate specificity and function in the *Escherichia coli* hotdog-fold thioesterase paralogs YdiI and YbdB**. *Biochemistry* 2014, **53**:4775-4787.

18. Shi D, Yu X, Cabrera-Luque J, Chen TY, Roth L, Morizono H, Allewell NM, Tuchman M: **A single mutation in the active site swaps the substrate specificity of N-acetyl-L-ornithine transcarbamylase and N-succinyl-L-ornithine transcarbamylase**. *Protein Sci* 2007, **16**:1689-1699.

19. VanderSluis B, Bellay J, Musso G, Costanzo M, Papp B, Vizeacoumar FJ, Baryshnikova A, Andrews B, Boone C, Myers CL: **Genetic interactions reveal the evolutionary trajectories of duplicate genes**. *Mol Syst Biol* 2010, **6**:429.

20. Pascarelli S, Laurino P: **Inter-paralog amino acid inversion events in large phylogenies of duplicated proteins**. *PLoS Comput Biol* 2022, **18**:e1010016.

21. • Yang G, Miton CM, Tokuriki N: **A mechanistic view of enzyme evolution**. *Protein Sci* 2020, **29**:1724-1747.
In the light of several examples, the authors discuss that after gene duplication of an ancestral enzyme with a main function and a promiscuous secondary function, one of the duplicated copies might accumulate mutations enhancing the ancestral secondary function, while the other retains the primary function.

22. •• Mallik S, Tawfik DS: **Determining the interaction status and evolutionary fate of duplicated homomeric proteins**. *PLoS Comput Biol* 2020, **16**:e1008145.
A systematic analysis of the fate of gene duplication of homomeric proteins in bacteria and eukaryotes. This study also depicts that a substantial fraction of prokaryotic homomeric protein families has become heteromeric in eukaryotes by gene duplications.

23. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA: **Evolution of protein complexes by duplication of homomeric interactions**. *Genome Biol* 2007, **8**:R51.

24. •• Marchant A, Cisneros AF, Dubé AK, Gagnon-Arsenault I, Ascencio D, Jain H, Aubé S, Eberlein C, Evans-Yamamoto D, Yachie N, *et al.*: **The role of structural pleiotropy and regulatory evolution in the retention of heteromers of paralogs**. *Elife* 2019, **8**:e46754.
Using *in silico* evolution, the authors showed that upon gene duplication of a homomer, heteromerization of the two descendant paralogs could be preserved indirectly due to the selection for the maintenance of the ancestral homomer.

25. Hochberg GKA, Shepherd DA, Marklund EG, Santhanagoplan I, Degiacomi MT, Laganowsky A, Allison TM, Basha E, Marty MT, Galpin MR, *et al.*: **Structural principles that enable oligomeric small heat-shock protein paralogs to evolve distinct functions**. *Science* 2018, **359**:930-935.

26. Maslov S, Sneppen K, Eriksen KA, Yan K-K: **Upstream plasticity and downstream robustness in evolution of molecular networks**. *BMC Evol Biol* 2004, **4**:9.

27. • Dandage R, Landry CR: **Paralog dependency indirectly affects the robustness of human cells**. *Mol Syst Biol* 2019, **15**:e8871.
The authors conducted a systematic analysis of human heteromeric paralogous pairs by investigating the fitness effect of single gene knockouts by CRISPR-Cas9 in several human cell lines; they concluded that the functional co-dependency of heteromeric paralogs accounts for the deleterious nature of the knockout of either copy.

28. • Tria FDK, Martin WF: **Gene duplications are at least 50 times less frequent than gene transfers in prokaryotic genomes**. *Genome Biol Evol* 2021, **13**:evab224.
The authors examined recent horizontal gene transfers among distantly related bacterial species; the results depict that horizontal gene transfers are 50 – 100 times more frequent than gene duplications in bacteria.

29. Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO, Marcotte EM: **Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity**. *Science* 2015, **348**:921-925.

30. Reuter R, Naumann M, Bär J, Haferburg D, Kopperschläger G: **Purification, molecular and kinetic characterization of phosphofructokinase-1 from the yeast Schizosaccharomyces pombe: evidence for an unusual subunit composition**. *Yeast* 2000, **16**:1273-1285.

31. Banaszak K, Mechin I, Obmolova G, Oldham M, Chang SH, Ruiz T, Radermacher M, Kopperschläger G, Rypniewski W: **The crystal structures of eukaryotic phosphofructokinases from baker's yeast and rabbit skeletal muscle**. *J Mol Biol* 2011, **407**:284-297.

32. •• Pillai AS, Chandler SA, Liu Y, Signore AV, Cortez-Romero CR, Benesch JLP, Laganowsky A, Storz JF, Hochberg GKA, Thornton JW: **Origin of complexity in haemoglobin evolution**. *Nature* 2020, **581**:480-485.
The authors reconstructed the evolutionary origin of modern heterotetrameric haemoglobin from an ancestral homodimeric oxygen-binding protein.

33. Rebeaud ME, Mallik S, Goloubinoff P, Tawfik DS: **On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life**. *Proc Natl Acad Sci USA* 2021, **118**:e2020885118.

34. Gong Y, Kakihara Y, Krogan N, Greenblatt J, Emili A, Zhang Z, Houry WA: **An atlas of chaperone-protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways in the cell**. *Mol Syst Biol* 2009, **5**:275.

35. Gille C, Goede A, Schlöetelburg C, Preissner R, Kloetzel PM, Göbel UB, Frömmel C: **A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome**. *J Mol Biol* 2003, **326**:1437-1448.

36. Wollenberg K, Swaffield JC: **Evolution of proteasomal ATPases**. *Mol Biol Evol* 2001, **18**:962-974.

37. Gupta RS: **Evolution of the chaperonin families (HSP60, HSP 10 and TCP-1) of proteins and the origin of eukaryotic cells**. *Mol Microbiol* 1995, **15**:1-11.

38. Lykke-Andersen S, Brodersen DE, Jensen TH: **Origins and activities of the eukaryotic exosome**. *J Cell Sci* 2009, **122**:1487-1494.

39. Veretnik S, Wills C, Youkharibache P, Valas RE, Bourne PE: **Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome**. *PLoS Comput Biol* 2009, **5**:e1000315.

40. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW: **Evolution of increased complexity in a molecular machine**. *Nature* 2012, **481**:360-364.

41. •• Johnston IG, Dingle K, Greenbury SF, Camargo CQ, Doye JPK, Ahnert SE, Louis AA: **Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution**. *Proc Natl Acad Sci USA* 2022, **119**:e2113883119.
The authors propose a non-adaptive mechanism explaining the prevalence of symmetric structures, which require less specific information to encode, and as such are more likely to appear than asymmetric assemblies through random mutations.

42. Kesmir C, van Noort V, de Boer RJ, Hogeweg P: **Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome**. *Immunogenetics* 2003, **55**:437-449.

43. •• Ascencio D, Diss G, Gagnon-Arsenault I, Dubé AK, DeLuna A, Landry CR: **Expression attenuation as a mechanism of robustness against gene duplication**. *Proc Natl Acad Sci USA* 2021, **118**:e2014345118.
The authors measured the fitness effects of the duplication of several essential genes in the budding yeast to show that most duplications have small to undetectable effects on fitness. Further, attenuation of gene expression of the recently duplicated copies might be a mechanism to protect the multiprotein complexes from gene duplications.

44. Marcet-Houben M, Gabaldón T: **Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage**. *PLoS Biol* 2015, **13**:e1002220.

45. Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, *et al.*: **The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions**. *Protein Sci* 2021, **30**:187-200.

46. Akiva E, Itzhaki Z, Margalit H: **Built-in loops allow versatility in domain–domain interactions: lessons from self-interacting domains**. *Proc Natl Acad Sci USA* 2008, **105**:13292-13297.

47. Hashimoto K, Panchenko AR: **Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states**. *Proc Natl Acad Sci USA* 2010, **107**:20352-20357.

48. Gurel PS, Kim LY, Ruijgrok PV, Omabegho T, Bryant Z, Alushin GM: **Cryo-EM structures reveal specialization at the myosin VI-actin interface and a mechanism of force sensitivity**. *Elife* 2017, **6**:e31125, https://doi.org/10.7554/eLife.31125

49. Knoll KR, Eustermann S, Niebauer V, Oberbeckmann E, Stoehr G, Schall K, Tosi A, Schwarz M, Buchfellner A, Korber P, *et al.*: **The nuclear actin-containing Arp8 module is a linker DNA sensor driving INO80 chromatin remodeling**. *Nat Struct Mol Biol* 2018, **25**:823-832.

50. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, *et al.*: **Pfam: the protein families database in 2021**. *Nucleic Acids Res* 2021, **49**:D412-D419.

51. Thoden JB, Holden HM: **The molecular architecture of galactose mutarotase/UDP-galactose 4-epimerase from *Saccharomyces cerevisiae***. *J Biol Chem* 2005, **280**:21900-21907.

52. •• Nocedal I, Laub MT: **Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity**. *Elife* 2022, **11**:e77346.

The authors used the ancestral sequence reconstruction approach to experimentally demonstrate how interaction specificity evolves by fixation of point mutations after gene duplication in bacterial signaling pathways involving a two-component system.

53. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D: **Computational redesign of protein-protein interaction specificity**. *Nat Struct Mol Biol* 2004, **11**:371-379.

54. Netzer R, Listov D, Lipsh R, Dym O, Albeck S, Knop O, Kleanthous C, Fleishman SJ: **Ultrahigh specificity in a network of computationally designed protein-interaction pairs**. *Nat Commun* 2018, **9**:5286.

55. Baker CR, Hanson-Smith V, Johnson AD: **Following gene duplication, paralog interference constrains transcriptional circuit evolution**. *Science* 2013, **342**:104-108.

56. Kaltenegger E, Ober D: **Paralogue interference affects the dynamics after gene duplication**. *Trends Plant Sci* 2015, **20**:814-821.

57. Grandier-Vazeille X, Bathany K, Chaignepain S, Camougrand N, Manon S, Schmitter JM: **Yeast mitochondrial dehydrogenases are associated in a supramolecular complex**. *Biochemistry* 2001, **40**:9758-9769.

58. Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK: **Global analysis of protein localization in budding yeast**. *Nature* 2003, **425**:686-691.

59. Brevet A, Chen J, Lévêque F, Blanquet S, Plateau P: **Comparison of the enzymatic properties of the two *Escherichia coli* lysyl-tRNA synthetase species**. *J Biol Chem* 1995, **270**:14439-14444.

60. Musso RE, Zabin I: **Substrate specificity and kinetic studies on thiogalactoside transacetylase**. *Biochemistry* 1973, **12**:553-557.

61. Lo Leggio L, Dal Degan F, Poulsen P, Andersen SM, Larsen S: **The structure and specificity of *Escherichia coli* maltose acetyltransferase give new insight into the LacA family of acyltransferases**. *Biochemistry* 2003, **42**:5225-5235.

62. Byrne KP, Wolfe KH: **The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species**. *Genome Res* 2005, **15**:1456-1461.

63. Ross J, Reid GA, Dawes IW: **The nucleotide sequence of the LPD1 gene encoding lipoamide dehydrogenase in *Saccharomyces cerevisiae*: comparison between eukaryotic and prokaryotic sequences for related enzymes and identification of potential upstream control sites**. *J Gen Microbiol* 1988, **134**:1131-1139.

64. Keyes BE, Burke DJ: **Irc15 Is a microtubule-associated protein that regulates microtubule dynamics in *Saccharomyces cerevisiae***. *Curr Biol* 2009, **19**:472-478.

65. Lin A-P, Demeler B, Minard KI, Anderson SL, Schirf V, Galaleldeen A, McAlister-Henn L: **Construction and analyses of tetrameric forms of yeast NAD+-specific isocitrate dehydrogenase**. *Biochemistry* 2011, **50**:230-239.

66. Laurent M, Seydoux FJ, Dessen P: **Allosteric regulation of yeast phosphofructokinase. Correlation between equilibrium binding, spectroscopic and kinetic data**. *J Biol Chem* 1979, **254**:7515-7520.

67. Breslow DK, Collins SR, Bodenmiller B, Aebersold R, Simons K, Shevchenko A, Ejsing CS, Weissman JS: **Orm family proteins mediate sphingolipid homeostasis**. *Nature* 2010, **463**:1048-1053.

68. Han G, Gupta SD, Gable K, Bacikova D, Sengupta N, Somashekarappa N, Proia RL, Harmon JM, Dunn TM: **The ORMs interact with transmembrane domain 1 of Lcb1 and regulate serine palmitoyltransferase oligomerization, activity and localization**. *Biochim Biophys Acta Mol Cell Biol Lipids* 2019, **1864**:245-259.

69. Marsh JA, Teichmann SA: **Structure, dynamics, assembly, and evolution of protein complexes**. *Annu Rev Biochem* 2015, **84**:551-575.

70. Lynch M: **The evolution of multimeric protein assemblages**. *Mol Biol Evol* 2012, **29**:1353-1366.

71. Conrad B, Antonarakis SE: **Gene duplication: a drive for phenotypic diversity and cause of human disease**. *Annu Rev Genom Hum Genet* 2007, **8**:17-35.

72. Bailey JA, Eichler EE: **Primate segmental duplications: crucibles of evolution, diversity and disease**. *Nat Rev Genet* 2006, **7**:552-564.

73. Verhaak RGW, Bafna V, Mischel PS: **Extrachromosomal oncogene amplification in tumour pathogenesis and evolution**. *Nat Rev Cancer* 2019, **19**:283-288.

74. Ronan A, Fagan K, Christie L, Conroy J, Nowak NJ, Turner G: **Familial 4.3 Mb duplication of 21q22 sheds new light on the Down syndrome critical region**. *J Med Genet* 2007, **44**:448-451.

75. Moog U, Engelen J, Albrechts J, Hoorntje T, Hendrikse F, Schrander-Stumpel C: **Alagille syndrome in a family with duplication 20p11**. *Clin Dysmorphol* 1996, **5**:279-288.

76. Khersonsky O, Tawfik DS: **Enzyme promiscuity: a mechanistic and evolutionary perspective**. *Annu Rev Biochem* 2010, **79**:471-505.

77. Nam H, Lewis NE, Lerman JA, Lee D-H, Chang RL, Kim D, Palsson BO: **Network context and selection in the evolution to enzyme specificity**. *Science* 2012, **337**:1101-1104.

78. Lynch M: **The evolution of genetic networks by non-adaptive processes**. *Nat Rev Genet* 2007, **8**:803-813.

79. •• Hochberg GKA, Liu Y, Marklund EG, Metzger BPH, Laganowsky A, Thornton JW: **A hydrophobic ratchet entrenches molecular complexes**. *Nature* 2020, **588**:503-508.

The authors show that evolutionary conservation of the oligomeric state can come from fixation of point-mutations that destabilize the monomeric state but not the multimeric state.

80. Plach MG, Semmelmann F, Busch F, Busch M, Heizinger L, Wysocki VH, Merkl R, Sterner R: **Evolutionary diversification of protein-protein interactions by interface add-ons**. *Proc Natl Acad Sci USA* 2017, **114**:E8333-E8342.

81. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583-589.

82. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, *et al.*: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science* 2021, **373**:871-876.

83. Dey S, Ritchie DW, Levy ED: **PDB-wide identification of biological assemblies from conserved quaternary structure geometry**. *Nat Methods* 2018, **15**:67-72.

84. Garcia Seisdedos H, Levin T, Shapira G, Freud S, Levy ED: **Mutant libraries reveal negative design shielding proteins from supramolecular self-assembly and relocalization in cells**. *Proc Natl Acad Sci USA* 2022, **119**:e2101117119.

85. Abrusán G, Marsh JA: **Ligand-binding-site structure shapes allosteric signal transduction and the evolution of allostery in protein complexes**. *Mol Biol Evol* 2019, **36**:1711-1727.