Structure
Previews

Hopeful (Protein InDel) Monsters?

Ágnes Tóth-Petróczy[1] and Dan S. Tawfik[1,*]
[1]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel
*Correspondence: tawfik@weizmann.ac.il
http://dx.doi.org/10.1016/j.str.2014.05.013

In this issue of *Structure*, Arpino and colleagues describe in atomic detail how a protein stomachs a deletion within a helix, an event that rarely occurs in nature or in the lab. Can insertions and deletions (InDels) trigger dramatic structural transitions?

It is generally accepted, and is largely the case, that evolutionary transitions occur incrementally via small gradual changes. However, can small gradual changes explain major evolutionary transitions and, foremost, the gaps between micro- and macroevolution? Small gradual changes readily account for changes in substrate specificity within a given enzyme (microevolution) and hence the divergence of different enzyme families within a superfamily. However, how did the major macroevolution transitions in the protein world occur, such as the emergence of different folds and superfamilies? Richard Goldschmidt proposed the concept of *macromutations*: mutations that introduce abrupt, dramatic transitions, also known as "hopeful monsters", that seem better positioned to drive macroevolution.

Protein evolution demands a hierarchy of sequence and structure changes. A single mutation may induce a change of specificity and/or a local structural perturbation. However, "once in a lifetime," a mutation may induce a profound change or even a new fold (Arodź and Płonka, 2012; He et al., 2012). The vast majority of new proteins are born via a "mix and match" of existing domains (duplication, insertion, domain shuffling, circular permutation, etc.). Thus, although very rare, such "hopeful monsters" may be critical to the birth of novel protein architectures. Among such "hopeful monsters" might be insertions and deletions (InDels). For example, a reconstructed deletion within an active-site loop led from nondetectable activity to a k_{cat}/K_M of 10^4 $M^{-1}s^{-1}$; this magnitude of change is not observed with point mutations (Afriat-Jurnou et al., 2012). Similarly, InDels within secondary structure elements—helices and β strands—might comprise *macromutations*. The snag is that, in general, such InDels are extremely deleterious (Tóth-Petróczy and Tawfik, 2013).

How deleterious? The rate by which changes occur in proteins is measured by aligning orthologous proteins with a predefined phylogeny. However, a certain change may happen very slowly because either the corresponding genetic change rarely occurs or it is deleterious at the protein level and thereby purged by selection. Short InDels occur frequently, but their occurrence in proteins is very rare (Tóth-Petróczy and Tawfik, 2013). Considering only in-frame InDels, selection purges InDels within structured proteins at a rate ≥ 9-fold higher relative to point mutations and up to 100-fold more intensely in secondary structure elements.

Given their highly deleterious effects, it is not surprising that protein engineering is mediated almost solely by point mutations. Indeed, in this issue of *Structure*, Arpino et al. (2014) provide a rare example of protein engineering via InDels in which single amino acids are deleted rather than substituted, as is routinely done with GFPs. In fact, InDels are unpopular across the scientific board. In sequence alignments, gaps (as they are called in this context) comprise the most problematic feature. As "gaps," their evolutionary history also remains unassigned in phylogenetic trees, because most current methodologies do not determine whether a given gap is the outcome of an insertion or a deletion. In computational design, calculating new backbone configurations is still a challenge. In rational design, assigning InDels by comparing related proteins is nontrivial, and identifying point mutations that enable these InDels (Tóth-Petróczy and Tawfik, 2013) is even trickier (Afriat-Jurnou et al., 2012). In directed evolution, methods for incorporating point mutations at random are trivial, but methods for random incorporation of InDels are still underdeveloped. Arpino et al. (2014) used a transposon that randomly inserts a cassette. The cassette was designed such that restriction digest resulted in tri-nucleotide deletions. This method is rather effective and has been applied in other proteins with similar results (Simm et al., 2007).

Foremost, this work shows how certain InDels in secondary structural elements may not only be tolerated, but may even be beneficial (Arpino et al., 2014). As observed in natural proteins (Tóth-Petróczy and Tawfik, 2013), the majority of tolerated deletions in GFP occurred within loops or at the edges of helices or strands (Arpino et al., 2014). However, a few tolerated deletions were observed within the cores of secondary-structure elements and especially within helices, including the deletion of Gly4 within the N-terminal helix. The latter even improved stability by virtue of a new set of interactions that successfully replaced the original ones (see Figure 5 in Arpino et al., 2014).

It appears that the fields of protein engineering and evolution may be ready to address the potential role of *macromutations*. However, the challenges are numerous. Although the deletion identified by Arpino et al. (2014) caused a shift in helix registry, it did not alter the GFP's scaffold or its function. Indeed, deletions were not tolerated in the β-barrel scaffold let alone anywhere near the fluorophore. In fact, to our knowledge, no *macromutations* (either InDels or point mutations) that gave birth to novel proteins have yet been identified. Another type of potential "hopeful monster" may stem from a frameshifting InDel that, through a single mutational event, changes the sequence of a long segment. A return to the original frame may initially occur via

transcriptional or translational slippage and, ultimately, by a second frameshift InDel downstream the first one (Rockah-Shmuel et al., 2013). This scenario is supported by the emergence of novel proteins via "overprinting" (Sabath et al., 2012).

What underlies the tolerance of InDels remains also unclear. This and previous work by Jones' lab suggest that deletions are more tolerated in helices than in strands (Arpino et al., 2014; Simm et al., 2007). Insertions in helices might also be relatively tolerated as indicated by the frequent observation of helix bulges (Cooley et al., 2010). However, such a trend is not seen in natural protein phylogenies (upon reanalysis of data in (Tóth-Petróczy and Tawfik, 2013). The accommodation of InDels in natural proteins also seems to be highly dependent on substitutions in the spatial vicinity of the accepted InDel that enable these acute insults (Afriat-Jurnou et al., 2012; Tóth-Petróczy and Tawfik, 2013). This trend was not observed by Arpino et al. (2014); point mutations were incorporated only at the positions flanking the deletions, and these did not mediate deletion tolerance. An interesting experiment might therefore be to identify whether point mutations could rescue deletions that caused loss of function, such as deletions within the scaffold or fluorophore. Other breakthroughs related to InDels, and to protein *macromutations* in general, may stem from a deeper understanding of their role in the evolutionary history of proteins (Arodź and Płonka, 2013), and foremost, from demonstrating their role in mediating abrupt, dramatic transitions of structure and function.

ACKNOWLEDGMENTS

Financial support by the Israel Science foundation is gratefully acknowledged. D.S.T. is the Nella and Leon Benoziyo Professor of Biochemistry.

REFERENCES

Afriat-Jurnou, L., Jackson, C.J., and Tawfik, D.S. (2012). Biochemistry *51*, 6047–6055.

Arodź, T., and Płonka, P.M. (2012). Proteins *80*, 1780–1790.

Arodź, T., and Płonka, P.M. (2013). J. Theor. Biol. *330*, 1–8.

Arpino, J.A.J., Reddington, S.C., Halliwell, L.M., Rizkallah, P.J., and Jones, D.D. (2014). Structure *22*, this issue, 889–898.

Cooley, R.B., Arp, D.J., and Karplus, P.A. (2010). J. Mol. Biol. *404*, 232–246.

He, Y., Chen, Y., Alexander, P.A., Bryan, P.N., and Orban, J. (2012). Structure *20*, 283–291.

Rockah-Shmuel, L., Tóth-Petróczy, Á., Sela, A., Wurtzel, O., Sorek, R., and Tawfik, D.S. (2013). PLoS Genet. *9*, e1003882.

Sabath, N., Wagner, A., and Karlin, D. (2012). Mol. Biol. Evol. *29*, 3767–3780.

Simm, A.M., Baldwin, A.J., Busse, K., and Jones, D.D. (2007). FEBS Lett. *581*, 3904–3908.

Tóth-Petróczy, A., and Tawfik, D.S. (2013). Mol. Biol. Evol. *30*, 761–771.

