

Editorial

Evolution acting on the same target, but at multiple levels: Proteins as the test case

My first Editorial, I thought, would be an excellent platform to share some interesting, less appreciated facets of biology, taking cues from our own research findings, melding them with others in the field and take the issue forward. The editorial is written in way so that less of my own complexity appears and only the complexity that is innate to biology is showcased. Biology, at any length and time-scales of study, is not only exciting but also very intriguing. I will, in my successive editorials, move into various scales of biology, only to wonder about its inner workings.

We all know that protein active sites are shaped into executing specific chemical transformations via the steps of catalysis. Therefore, an active site is really the ‘business end’ of a protein when it is an enzyme, where the directional nature of spatial and charge transfers within the active site seems to mediate the act of catalysis. Active site predictions in a protein must, therefore, invoke methods that combine and compare both spatial congruence and electrostatic potentials of putative active site residues with those in *bona fide* active site motifs. Methods that successfully perform spatial matches of motifs are fairly abundant (Kleywegt 1999; Goyal *et al.* 2007; Debret *et al.* 2009), while those that factor in electrostatic potential comparisons are somewhat rare, even today. We realized this deficit sometime back and tried to develop a method that showed that electrostatic potential difference (PD) between analogous residue pairs in an active site from different proteins of the same enzyme family are similar and therefore can be effectively combined with spatial matching methods. That means we found strong correlation in the electrostatic PD between sets of cognate residues in active sites, where for a given enzymatic activity, pairs of residues in an active site from various proteins of the same family yield highly correlated electrostatic PD with very small standard deviation where all such PD’s lie in a narrow band. Using this robust finding, we pruned out false matches from spatial congruence sets and offered an additional filter for improving accuracy in active site prediction. By demonstrating conservation of electrostatic potential differences in cognate pairs of residues in a wide range of related proteins, we established a computational method, CataLytic Active Site Prediction (CLASP) to detect active sites (Chakraborty *et al.* 2011, 2013a).

We extended the method further and quantified promiscuous active sites in a wide range of proteins (PROMISE) (Chakraborty and Rao 2012; Chakraborty *et al.* 2012a, b). We found that by interrogating any given hypothetical protein using CLASP, we could ‘sniff’ for a potential active site in the bait-protein after comparing with known motifs in the Catalytic Site Atlas (CSA) database and even rank the ‘hit’ matches in the bait. Such ranking comparison allowed CLASP in classifying promiscuous activities associated with the bait-protein. We probed various alkaline phosphatase (AP) structures for well defined motifs from CSA and uncovered a uniquely promiscuous proteolytic active site in shrimp alkaline phosphatase (SAP). We substantiated the same by providing experimental evidence *in vitro*. It turned out that trypsin inhibitors could competitively inhibit AP activity, which was extended to probing other APs as well (Chakraborty *et al.* 2011, 2013b; Chakraborty and Rao 2012).

While it is conceivable that stereochemical equivalence is hardwired for amino acids with similar properties, one finds instances of active sites where residues with different properties occupy the same sequence and spatial location but perform the same function (Lobkovsky *et al.* 1993)! Finally, while choosing a site in a protein as a possible scaffold, it is sometimes useful to relax the constraints to allow identification of moonlighting active sites ‘lurking’ in its vicinity (Jeffery 2009). Indeed, modulating the radius of search defining the ‘active site vicinity’ did help us characterize properties of residues that determine promiscuity (Chakraborty and Rao 2012).

Notwithstanding the general mutational changes in a protein, understanding those that drive novel protein active site functions remains an exciting challenge even today. There the issues under intense debate are: how an evolutionary trajectory shapes the mutation-spectrum of an active site from the pre-existing to the next with altered biochemical functions; whether early mutations set an influential trend

where the mutational paths get additive or get restricted by dynamic adaptive changes or a combination of both and how the newly evolving function trades off with the pre-existing function, leading to conversion from either a ‘generalist’ to ‘specialist’ or vice versa and the routes that shape intermediates leading to high-affinity-high-selectivity sites, etc. (Tawfik *et al.* 2009). It has been proposed that novel activities are acquired by active sites when ‘generalists’ convert to ‘specialists’ via intermediates that exhibit exceptionally wide ranges of promiscuous activities (Matsumura and Ellington 2001). Indeed, examples do exist suggesting that non-overlapping specificities become accessible via ‘generalist’ intermediates emerging under selection (Rockah-Shmuel and Tawfik 2012). Therefore, a ‘hot-soup’ of promiscuity seems to serve as a fertile ground for selections that eventually ‘fix’ specificity in a newly emerging ‘specialist’!

It is well known that entire body of the protein can accumulate mutations over large evolutionary time-scales. Interestingly, sets of correlated-mutations where a particular amino acid change at one location is accompanied by certain other change at another location in the protein (perhaps as a compensatory alteration), supporting a model that 3D structural contacts might influence the mutation landscape within the protein (Jacob *et al.* 2015). There is now a strong hope, backed by powerful statistical tools, that accurate mapping of many such ‘correlated mutations’ during its evolution in a protein could indeed yield sufficient structural constraints that one could start predicting the entire 3D-structure of the protein using such statistically ‘cosher’ constraints!

So the protein evolution involves a balance between ‘whole body’ centric structure driven mutational landscape, within which local active site biochemistry spins out additional trajectories of mutational changes that are tightly coupled to functional specificity of the protein. Moreover, most proteins co-exist within the biochemical niche of its interactome, which in turn imposes additional constraints on mutation landscape of a protein. We know much less on the nature of these constraints as our understanding of interactome evolution is very nascent now. It is conceivable that interactome based constraints adds additional layer of complexity in protein evolution by imposing other adaptive changes in the protein functions. All these arguments put together seems to strongly suggest that a single target, the protein, is subject to multiple mutational pressures, ranging from its local active site centric changes to overall 3D folds-related effects to those emanating from 3D-structural plasticity of the same protein within its dynamic interactome. It is excitingly intriguing to think about all such possibilities.

References

- Chakraborty S and Rao BJ 2012 A measure of the promiscuity of proteins and characteristics of residues in the vicinity of the catalytic site that regulate promiscuity. *PLoS One* **7** e32011
- Chakraborty S, Minda R, Salaye L, Bhattacharjee SK and Rao BJ 2011 Active site detection by spatial conformity and electrostatic analysis—unravelling a proteolytic function in shrimp alkaline phosphatase. *PLoS One*. **6** e28470
- Chakraborty S, Minda R, Ásgeirsson B, Salaye L, Frère JM, Rao BJ 2012a Inhibition of a cold-active alkaline phosphatase revealed by in silico modeling of mono and binuclear metallo- β -lactamase active sites. *FEBS Lett.* **586** 3710-3715
- Chakraborty S, Ásgeirsson B and Rao BJ 2012b A measure of the broad substrate specificity of enzymes based on ‘duplicate’ catalytic residues. *PLoS One* **7** e49313
- Chakraborty S, Minda R, Salaye L, Dandekar A, Bhattacharjee SK and Rao BJ 2013a Promiscuity based enzyme selection for rational directed evolution experiments. *Methods Mol. Biol.* **978** 205–216
- Chakraborty S, Rao BJ, Baker N and Ásgeirsson B 2013b Structural phylogeny by profile extraction and multiple superimposition using electrostatic congruence as a discriminator. *Intrinsically Disord Proteins* **1** e25463-13
- Debret G, Martel A and Cuniasse P 2009 RASMOT-3D PRO: a 3D motif search webserver. *Nucleic Acids Res.* **37** W459–W464
- Goyal K, Mohanty D and Mande SC 2007 PAR-3D: a server to predict protein active site residues. *Nucleic Acids Res.* **35** W503–W505
- Jacob E, Unger R, Horovitz A 2015 Elife Codon-level information improves predictions of inter-residue contacts in proteins by correlated mutation analysis. *eLife* **4**:e08932. doi:10.7554/eLife.08932
- Jeffery CJ 2009 Moonlighting proteins—an update. *Mol. BioSyst.* **5** 345–350
- Kleywegt GJ 1999 Recognition of spatial motifs in protein structures. *J. Mol. Biol.* **285** 1887–1897

- Lobkovsky E, Moews PC, Liu H, *et al.* 1993 Evolution of an enzyme activity: crystallographic structure at 2 Å resolution of cephalosporinase from the ampC gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *Proc. Natl. Acad. Sci. USA* **90** 11257–11261
- Matsumura I and Ellington AD 2001 In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* **305** 331–339
- Rockah-Shmuel L and Tawfik DS 2012 Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. *Nucleic Acids Res.* **40** 11627–11637
- Tawfik DS *et al.* 2009 Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nat. Struct. Mol. Biol.* **16** 1049–1055

Basuthkar J Rao

Editor