Journal of Pharmacogenomics & Pharmacoproteomics

Editorial **Open Access**

Evolutionary Roots of Proteomic Complexity and Lessons for the Drug Designer

Ariel Fernández*

Argentine Institute of Mathematics (I. A. M.), National Research Council (CONICET), Buenos Aires 1083, Argentina,Collegium Basilea-Institute for Advanced Study, Basel CH4053, Switzerland

As it is well known, proteins with common ancestry (homologs) share a common structural topology [1,2]. This similarity introduces major hurdles for structure-based drug design as it precludes an effective and straightforward control of drug specificity. While folding topology bears considerable similarity across homologs, protein wrapping (the extent of structure shielding from water attack) tends to be different, offering an opportunity to focus the impact of a drug on clinically relevant targets [3]. Central to wrapping is the concept of dehydron, a packing defect consisting of an unburied backbone hydrogen bond. We may state that dehydron patterns tend to be different across homologous proteins even if the latter share a common fold. Wrapping variations across homologs have profound consequences for drug design as we aim at engineering target-specific therapeutic agents and attempt to build insightful animal models for disease. In assessing the evolutionary forces that promote differences in the dehydron patterns across orthologous proteins (homologs from different species), we found that random genetic drift plays a central role in causing dehydron enrichment [4]. This type of structural degradation promotes higher protein interactivity and is more pronounced in species with low population, such as humans, where mildly deleterious mutations resulting from random drift have a significant probability of getting fixed in the population. As discussed below, the fitness consequences of nature's evolutionary strategy reveal the high exposure of the human species to the vagaries of random drift that may eventually result in fitness catastrophes. These catastrophes are enshrined in proteins so rich in dehydrons that they become prone to undergo aberrant aggregation.

For the drug designer, the structural similarity across homologs of the protein target is particularly troublesome. Major goals in drug design are inevitably entangled with this evolutionary hurdle. The existence of homologs of the protein target within the species (paralogs) or across species (orthologs) can introduce undesired cross reactivities due to their high level of structural similarity with the target. These cross reactivities, in turn, can result in dangerous side effects or can render inconsequential the animal models for a particular disease.

As said, while the fold is highly conserved across homologs, the wrapping or dehydron pattern is not conserved. We should distinguish orthologs from paralogs when assessing the evolutionary origin of these dissimilarities. In orthologs, the variations arise mostly from differences in the level of efficiency of natural selection across different species [4], whereas in paralogs the dissimilarity is rooted in the imperative to curb dosage imbalances that arise if two copies of the same gene were retained undifferentiated [5].

An emerging picture arises where structural wrapping becomes the molecular dimension exploited by evolution to foster interactome complexity [1-4]. Wrapping differences across paralogs sharing the same fold provide an escape route to mitigate the fitness problems arising from paralog retention [5]. Both evolutionary aspects should be taken into account at the next level of pharmacoinformatics inference [3].

As we explore the evolutionary forces that promote differences across orthologous proteins, we realize that random genetic drift is the main factor causing dehydron enrichment in species with low population [4]. Because dehydrons are markers for protein association, this type of structural degradation promotes higher protein interactivity which becomes more pronounced in species with low population, such as humans. The fitness consequences of this "evolutionary gambit" reveal the high exposure of the human species to fitness catastrophes arising from proteins prone to relinquish their native fold and undergo aberrant aggregation.

Much evolutionary change is detrimental rather than beneficial. It is expected that an evolutionary mutation will likely degrade the soluble protein fold by creating new dehydrons, rather than tighten the structure, shielding it from water attack. Prevailing mildly deleterious mutations typically degrade the wrapping of the protein, enriching its dehydron content, thereby making the soluble protein more reliant on associations with binding partners in order to maintain its integrity [1,2]. Since protein interactivity, regulation and allostery, hallmarks of cooperativity and organismic sophistication, all become more developed with decreasing species population, one may say that nonadaptive factors are likely to play a major role in promoting complexity in higher eukaryotes.

A dosage imbalance implies a rising protein concentration level at specific locations in tissue, metabolic or developmental phases in a way that does not fit the stoichiometry of the complexes formed by the protein with binding partners with fixed concentrations. The complexes may be transient or obligatory vis-à-vis the integrity of protein structure and function and hence the effects of the dosage imbalances may vary accordingly with great latitude. Therefore, dosage sensitivity, that is, the fitness impact of dosage imbalance, is influenced by the level of reliance of the protein on its binding partners to maintain structural integrity and functional competence.

Structural bioinformatics has revealed that protein under-wrapping or packing deficiency is the molecular marker of dosage sensitivity [3,5]. This is because packing deficiency measures the reliance of the protein on associations with binding partnerships to maintain its structural integrity through intermolecular wrapping of dehydrons. A deficiently packed protein is more likely to be involved in an obligatory complex

*Corresponding author: Ariel Fernández, Argentine Institute of Mathematics (I. A. M.), National Research Council (CONICET), Buenos Aires 1083, Argentina,Collegium Basilea-Institute for Advanced Study, Basel CH4053, Switzerland, Tel: 54 11 4954 6781; Fax: 54 11 4954 6782; E-mail: ariel@afinnovation.com

and its dosage imbalances are therefore more likely to impact fitness than those of a well-wrapped protein that can function autonomously.

In humans, insensitivity to dosage imbalances may simply result from selection inefficiency which arises in species with relatively small population size [3,4]. This suggests that the selection pressure exerted on paralogs of the poorly packed human proteins is simply not operative. Alternatively, the higher complexity of expression regulation may offer an escape route from dosage imbalances, introducing a resilience that is not found in unicellular organisms. The molecular mechanisms that humans possess to cope with the deleterious consequences of selection inefficiency have been described elsewhere [3].

Biological functions in humans are far more regulated and resilient than those of other species with larger populations, owing to complex webs of protein-protein interactions. Yet, paraphrasing science writer Philip Ball [6], we may be ultimately doomed because of the way in which our interactome complexity arose: Nature's gambit may have activated a time bomb, and the resulting complexity may be only a short-term fix.

As previously said, it is generally believed that the protein "fold" is conserved as protein orthologs are examined across species. This is because of the tight correspondence between structure and function and attests to the evolutionary robustness of protein function. Consequently, orthologous proteins that retain the same function are expected to keep the same fold even if the protein amino acid sequence varies significantly.

This structural conservation across protein orthologs introduces yet another puzzle in regards to the origin of our complexity, given the relatively modest size of the human genome. If the structure of the proteins is conserved across species and the number of human genes is modest and commensurate with that of lower eukaryotes and prokaryota, where is the source of our presumed higher complexity?

As said, we have discovered subtle structural variations occurring in orthologs from species that diverged from each other billions of years ago. In these structures, a feature subtler than the overall topology can be demonstrated to vary significantly across orthologs [1-4]. The structure in some seems "looser" than in others, less well wrapped, with surface regions that enable surrounding water to penetrate and disrupt the structure by interacting favorably with the protein backbone.

As indicated, structural degradation is an indicator of the species' exposure to random genetic drift: mildly deleterious mutations that would typically degrade the protein structure are more likely to be selected against in bacteria before they can become fixed in the entire population (estimated in trillions of individuals), whereas such a mutation has a far better chance to prevail in humans.

A protein richer in dehydrons than its ortholog is more vulnerable to disruption due to backbone hydration. For this reason, the integrity of its structure becomes more reliant on associations with binding partners. Furthermore, dehydrons are known to promote removal of vicinal water molecules, that is, they are sticky. This fact implies that structurally degraded proteins are more likely to promote protein-protein associations than their better-wrapped orthologous counterparts. This leads us to the striking conclusion that protein-protein interactions, a hallmark of organismic sophistication, are actually promoted by random genetic drift. Interactome complexity does not appear to be naturally selected, but instead emerges as a result of selection inefficiency.

Prions are known to be some of the soluble proteins with the largest accumulation of dehydrons [3]. Their wrapping is so poor that they relinquish their functional fold and form regular aggregates (fibrils) that may result in neuropathies. Thus, prions are "aberrantly needy proteins", serving as primary illustrations of the level of genetic risk to which we are exposed due to our small species population. The prion represents a "fitness catastrophe", becoming a telling story as to where Nature's gambit may lead in regards to the fate of human species. We may prevail as a species providing we are able to mitigate the fitness cost of our complexity. This may require an arduous therapeutic solution.

As indicated, the wrapping differences across orthologous proteins arise due to differences in selection efficiency across species. In the case of paralogous proteins, significant dosage imbalances would arise if paralogs were identical, as they indeed were in the initial stages immediately following gene duplication. To dodge the fitness consequences of such imbalances, dissimilar expression regulation patterns occur across the original duplicate genes and this level of differentiation increases significantly for highly under-wrapped proteins. Thus, in previous work we showed how paralog segregation is evolutionarily achieved and how the extent of segregation correlates with the dehydron richness of the protein (i. e. with its dosage sensitivity) [3].

An alternative route for paralog segregation, required a buffer of the dosage imbalances resulting from gene duplication, arises through dissimilar wrapping of their (otherwise virtually identical) structural topology. This wrapping divergence across paralogs has crucial consequences for drug design. The evolutionary finding enables designers to discriminate paralogs harnessing their differences in dehydron patterns. This next generation of drugs belonging to the so-called "wrapping technology" exploits the design concept that a drug may function as an exogenous wrapper of the packing defects of the target protein.

On the other hand, the wrapping differences found across orthologous proteins may be effectively exploited by the drug designer. Thus, the therapeutic efficacy of drugs is routinely tested in animal models under the unwarranted assumption that the affinity and specificity of the drug for its target in human would be very close to those of the animal ortholog. Given the wrapping differences found across orthologs, there is no assurance that this scenario may be indeed valid. This observation could potentially undermine the very foundation on which animal testing is predicated.

The observed wrapping differences across orthologs may significantly impact the drug-based treatment of infectious diseases such as AIDS when approached within the wrapping technology platform. It would be most helpful for the wrapping-drug designer to detect differences in the dehydron patterns between the target protein in the infecting species and its human ortholog: A drug purposely engineered to wrap dehydrons in the protein of the infecting species that are not conserved in the human orthologous counterpart is likely to be far less toxic than a drug that cannot differentiate between the two orthologs. The issue of species selectivity becomes moot in an AIDS treatment geared at targeting the reverse transcriptase of the HIV-1 retrovirus, simply because this protein is not encoded in the human genome. On the other hand, a molecular design rooted in wrapping differences would be very helpful if the goal is to target a retroviral protein that is also represented in the human proteome. Species selectivity matters in this last context and dehydron patterns may be used as the appropriate selectivity filters.

References

1. Fernández A, Scott R, Berry RS (2004) The nonconserved wrapping of conserved folds reveals a trend towards increasing connectivity in proteomic networks". Proc Natl Acad Sci USA 101: 2823-2827.

2. Fernández A, Berry RS (2004) Molecular dimension explored in evolution to promote proteomic complexity. Proc Natl Acad Sci USA 101: 13460-13465.

3. Fernández Stigliano A (2015) Biomolecular Interfaces: Interactions, Functions and Drug Design. Springer.

4. Fernández A, Lynch M (2011) Nonadaptive origins of interactome complexity. Nature 474: 502-505.

5. Liang H, Rogale-Plazonic K, Chen J, Li WH, Fernández A (2008) Protein under-wrapping causes dosage sensitivity and decreases gene duplicability. PLoS Genetics 4: e11.

6. Ball P (2011) The Achilles' heel of biological complexity. Nature.