

Analyzing multiple functionalities in proteins

15.3.09

Motivation and Background. Many cases of proteins with multiple functions are now known. The extra function has almost always come as a surprise and was discovered serendipitously. The true extent of multiple functionalities is unknown and would be of great interest to evaluate. Proteins consist of many atoms and have complex movements of time spans of a μ -second to a second, so theoretically it would be possible pack many actions into one protein. To exemplify, describing the action of 1 protein could entail the positions of 10^4 atoms and to describe its motion in 1 μ -second would need 10^9 time steps, in total 10^{13} coordinates. Its function might be summarized by one action (such as hydrolyzing 1 bond) in this time period. However, designed machines are different from evolved machines and these numbers also illustrate that there could be room for several functions, possibly many.

Moonlighting proteins are clearly understudied and could be extremely important in the biology of an organism.

Project. A series of questions could be addressed by standard methods of molecular evolution. However, hard questions would remain and could not be solved by computational methods alone – hard experimental evidence on many more moonlighting proteins would be needed.

It is of interest to

1. Tabulate known cases of moonlighting proteins
2. Evolutionary modeling the phenomena of moonlighting
3. Measuring selection strength of known moonlighting proteins
4. If possible to separate the selection into component stemming from the two functionalities?
5. Is it possible to assign functional constraint to individual positions from the two functionalities with homologous proteins that have combinations of the functionalities
6. Dating the acquisition of a newest function
7. Detecting multiple functions from structure, molecular evolution and molecular dynamics

1) has to some extent already been done in a series of papers (Jeffery99,03,04). However, given the unsystematic nature of detection of extra functions, one should expect the phenomena to have been seriously underestimated. The total number of investigated cases is also low. The natural investigation of this would be to see how the number and fraction has evolved over the last decades. Especially cases with homologous proteins where one has 2 functions, the other only 1 would be useful to study.

2) evolutionary modelling of moonlightening as competition between the evolutionary advantage of a new function and the choice of creating a new protein with the new function or adding the new function to an existing function. This is clearly laudable, but would be faced with the same difficult questions: how difficult is it to fit an extra function into an existing protein and how strong is the advantage of a new function? Evolutionary moonlighting must have some similarity to overlapping reading frames in that they both impose overlapping constraints. There are also serious differences: Overlapping reading frames are easy to find, they impose the double constraint on the genetic material and there is a well defined limit to how many overlaps can be postulated (6 is maximum, only triple overlap has been seen). Moonlighting imposes additional constraints on the same protein, it might more frequent as the protein is produced anyway and there is no simple limit to how many functionalities could be imposed on the same protein.

3) could be done by standard methods of measuring selection without much difficulty (Yang, 2006). One would clearly expect more constraints and thus slower evolution of proteins with multiple functions. However, there is a very large variation in the amount of selection on single function proteins and their evolutionary rates. Thus tabulation of evolutionary rates might not give a clear cut picture.

4) could have much potential if proteins can be found where a new function was added recently and there was many closely related sequence of the protein with and without the function. This could provide much information on the strength and spatial distribution of selection along the protein. It would also be interesting to know if there was a systematic bias towards the original function dominating the functional constraints. Again, this ideal situation could be complicated by the functional constraint of one function can fluctuate over time.

5) would be clearly be of much value, but probably difficult as there is little knowledge about how functional constraints is distributed along the protein. If it was possible to get many sequences from homologous variants with and without the additional function, progress on this point would be possible.

6) is different from the question 4) above and should be applied to all known moonlighting proteins. There will be two kinds of datings possible: dating origin of new function from deceleration of evolutionary rate and from the distributions of functionalities on the leaves of the phylogeny relating the proteins. Experimental knowledge of function is superior, but harder to come by. Dating of loss/gains of extra functions could provide extremely valuable information on how frequent the phenomena is: If most moonlighting is very recent then it is probably a frequent phenomena.



To the left is shown a protein (thin black line) with the functional constraints imposed from two different functions (blue and red) as bars above it. Since little is known (or there might be few general rules) about the distribution along the sequence of functional constraints, it could be difficult to disentangle into two components.

To the right a phylogeny is relating a series of homologous moonlighting proteins, where the functionalities are experimentally known in a series of species (shown as colored balls). Such phylogenetic information could be highly valuable in detect when the set of functionalities have changed. A star is here indicated the event of a gain of function whose date it is the goal to infer.

7) This clearly is *the* challenge and harder than any of the above. Techniques have been developed to predict functions from sequence de novo (without resorting to homology) (Kannan et al., 2008 and Hawkins et al., 2007). Functional prediction when function is experimentally known in closely related sequences is much easier.

Research. We will focus on 3 and 6 above as these can be addressed with standard methods. 3. Will need set of homologous moonlightening proteins. Their rate of evolution, strength of selection and validity of molecular clock could be measured by a program like PAML (Yang, 2006). 6. Will be the same as 3 except some extension are needed as now leaves are labelled by 1 or 2 functionalities and the rate of evolution and strength of selection will be depend on that labelling valid at any point in time. Possibly programs exists to do this, possibly this will have to be coded from scratch. It will in this context be of interest to determined the altered selection pressure position by position.

Comments. Our group (Hein and colleagues) has worked on several problems involving overlapping functionalities, such as finding overlapping protein genes (McCauley and Hein), aligning using information from both DNA and protein level at the same time (Hein and Støvlbæk, 1994a,b 1996) and protein genes also subject to RNA secondary structure constraints (Pedersen et al., 2004a,b). This probably motivates our interest in moonlighting proteins, however it is unlikely that techniques can be transferred. All the problems mentioned have very well defined known structure such a protein genes and RNA secondary structure.

References.

- Daniel (2008) BAD: undertaker by night, candyman by day *Oncogene* 1:553-70.
- Hawkins T, Kihara D. Function prediction of uncharacterized proteins. *J Bioinform Comput Biol.* 2007 Feb;5(1):1-30.
- Hein, J. & Støvlbæk, J. (1994) A maximum likelihood approach to the evolution of overlapping reading frames. *J. Mol. Evol.* 40, 181-190
- Hein, J. (1994) An algorithm combining DNA and protein alignment. *J. Theor. Biol.* 167, 169-174
- Hein, J & Støvlbæk, J. (1996) Combining DNA and Protein Alignment. *Methods of Enzymology* 266, 402-415
- Constance J. Jeffery Moonlighting proteins *TIBS* 24.8- – *JANUARY 1999*
- Constance J. Jeffery Moonlighting proteins: old proteins learning new tricks *TRENDS in Genetics* Vol.19 No.8 August 2003 415-
- Carlos Gancedo* and Carmen-Lisset Flores Moonlighting Proteins in Yeasts *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, Mar. 2008, p. 197–210
- Hult K, Berglund P. Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.* 2007 May;25(5):231-8.
- Kannan S, Hauth AM, Burger G. Function prediction of hypothetical proteins without sequence similarity to proteins of known function *Protein Pept Lett.* 2008;15(10):1107-16.
- Khersonsky O, Roodveldt C, Tawfik DS. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol.* 2006 Oct;10(5):498-508.
- McCauley, S. & Hein, J. (2006) Using hidden Markov models (HMM) and observed evolution to annotate ssRNA viral genomes. *Bioinformatics*, 22(11), 1308–1316
- Pedersen, et al. (2004) A comparative method for finding and folding RNA secondary structures in protein-coding regions. *Nucleic Acids Research*, 32, 4925–4936
- Pedersen et al. (2004) An evolutionary model for protein-coding regions with conserved RNA structure. *Mol. Biol. Evol.*, 21, 1913–1922.
- Taglieber A, Höbenreich H, Carballeira JD, Mondière RJ, Reetz MT. 2007 “Alternate-site enzyme promiscuity” *Angew Chem Int Ed Engl.*;46(45):8597-600.
- Thorne, Painter and Kishino (1998) **Estimating the rate of evolution of the rate of molecular evolution** *Mol Evol* 15(12):1647-57
- Whisstock JC, Lesk AM. 2003 Prediction of protein function from protein sequence and structure. *Q Rev Biophys.* 36(3):307-40.
- Yang, Z (2006) *Computational Molecular Evolution* OUP