Evolutionary Origin of the Tumor Suppressor Hyperplastic Discs Protein

Leonid V. Omelyanchuk\textsuperscript{a,\ast}, Julya A. Pertseva\textsuperscript{a} and John M. Saul\textsuperscript{b}

\textsuperscript{a}Institute of Cytology and Genetics, Russian Academy of Sciences, 10 Lavrent’ev Ave., 630090, Novosibirsk, Russia
\textsuperscript{b}ORYX, 3 rue Bourdaloue, 75009 Paris, France

ABSTRACT: Previous evolutionary study of the tumor suppressor Merlin revealed that this protein family was produced by very early metazoans with the exception of some or all flatworm lineages [Golovnina \textit{et al.}, 2005]. We ask whether other tumor-suppressor proteins had also been in existence in these times and focus our attention on Hyperplastic Discs (Hyd) protein, a classic tumor suppressor in \textit{Drosophila melanogaster} which, when mutated, may cause over-proliferation and malignancy. Phylogenetic analysis of the Hyd protein indicates that it was present among metazoa by the time \textit{Trichoplax adhaerens} had emerged from the common unicellular ancestor of the Animalia.

KEYWORDS: Tumor suppressor, multicellularity, hyperplastic discs

INTRODUCTION

The study of tumor suppressor genes in \textit{D. melanogaster} showed that such genes are active in a variety of tissues but that only one or at most two cell types become malignantly transformed by any particular mutation [Gateff, 1994]. This tissue-specific phenotypic behavior raises questions about the evolutionary origin of such genes. One might expect that these protein families arose relatively late and that their original function had been to control the size and shape or organs. Yet previous evolutionary study of the tumor suppressor protein Merlin revealed that it was produced by early metazoans, with some or all flatworm lineages as the only known exceptions [Golovnina \textit{et al.}, 2005]. It thus seems apparent that tumor suppressor proteins of the Merlin family were being formed as early as the first tissues themselves. In consequence, we decided to investigate the historical biology of another well-known \textit{D. melanogaster} tumor suppressor, Hyperplastic Discs (Hyd).

In \textit{Drosophila}, Hyd is required for the regulation of cell proliferation during development [Mansfield \textit{et al.}, 1994], with mutations in the \textit{hyd} gene resulting in developmental abnormalities that include adult sterility caused by germ cell defects [Callaghan \textit{et al.}, 1998]. In mammals a role of Hyd in cancer is suggested by its commonly observed overexpression in breast and ovarian cancers [Clancy \textit{et al.}, 2003; Fuja \textit{et al.}, 2004]. Further, Hyd has been shown to be an \textit{in vivo} substrate for ERK1 and ERK2 [Eblen \textit{et al.}, 2003].

\textsuperscript{\ast}Corresponding author. E-mail: ome@bionet.nsc.ru.
Structurally, Hyd ligases contain a ubiquitin-associated domain at their N-termini, two nuclear localization signals, a zinc finger-like UBR domain involved in recognition of type 1 N-terminal regions [Tasaki et al., 2005], a domain highly homologous to the PABC (poly(A)-binding protein C-terminal) domain, and a HECT domain at their extreme C-termini. Previous NMR studies showed that the PABC domain is a peptide-binding domain that specifically recognizes a conserved PAM2 (PABP-interacting motif 2) sequence [Kozlov et al., 2001]. This PAM2 site was initially identified in the PABP-interacting proteins Paip1 and Paip2 and in eukaryotic release factor 3 (eRF3). These proteins modulate translational activity, either by stabilizing (Paip1 and eRF3), or by destabilizing the closed loop structure of mRNA (Paip2), formation of which involves the simultaneous interactions of PABP with the poly(A) tail of mRNA and the 5’-cap binding complex [Khaleghpour et al., 2001; Roy et al., 2002; Uchida et al., 2002].

RESULTS AND DISCUSSION

In order to identify putative Hyd sequences in a wide range of eukaryotes, we performed BLAST search of homologous sequences in the database available at the ExPASy Proteomics Server (http://
Fig. 1. A. Phylogenetic tree for Hyd protein family constructed by maximal parsimony method with the Mega-4 program. Both maximal parsimony and nearest neighbor methods give the same tree topology. The branch lengths were evaluated by Mega-4 in NJ option. B. Schematic picture of different Hyd domains. Numbers denote domains position in *Drosophila melanogaster* Hyd amino acid sequence.

ca.expasy.org/) using the *Drosophila melanogaster* Hyd sequence (P51592) as a template. We also performed a search for sequences homologous to Hyd from *D. melanogaster* in species whose genome projects are represented at NCBI (http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=genomeprj) and at the Eukaryotic Genomics web site (http://genome.jgi-psf.org/). Resulting protein sequences that showed similarity dispersing over the entire length of the Hyd protein were chosen for further analysis (Table 1). One particular part of the identified protein sequences (XP_001674329.1 *Caenorhabditis briggsae*, Q7TMY8 *Mus musculus*, Q756G2 *Ashbya gossypii*) was found to be abnormally long and could not therefore belong to the true Hyd protein family. The Decrease Redundancy program from the ExPasy web-site (http://cn.expasy.org/tools/redundancy/) was applied for the search of identical protein sequences in our sequence set. As illustrated in Table 1, all mammalian proteins appear to be identical, so we have simply taken the human protein UBR5 to represent mammals. The alignment and phylogenetic tree for the rest of the protein set showed that the sequences Q76N89, Q8K4P8, Q9P2P5, Q6I6G8 constitute a separate phylogenetic cluster that is only distantly related to *Drosophila* Hyd protein. These were therefore omitted from further consideration. Among *Culex quinquefasciatus* proteins we chose XP_001861835.1 one with the longest homology to reference Hyd. XP_321295.4 protein was identified by computer translation and has only partial similarity with Hyd and so it too was omitted, as were the proteins Q7TMY8, Q756G2 and XP_001674329.1, which were annotated as belonging to a non-Hyd protein family. Protein UFD4 from *Saccharomyces cerevisiae* was taken as an outgroup.

A phylogenetic tree for Hyd homologs (Fig. 1) was constructed using the Mega-4 computer program. The alignment used can be found in Supplement 1. All the tree branches have highly trustworthy positive bootstrap values of over 93, and it is evident that the divergence of both vertebrate and non-vertebrate animals are well described (The Tree of Life Web Project, http://tolweb.org) [Maddison et al., 2007]. Yet the divergence of *Trichoplax adhaerens* and *Caenorhabditis elegans* from the common
unicellular ancestor of multicellular organisms did not fit expectations. Analysis of mitochondrial DNA had indicated that the divergence of Placozoa from the common ancestor is the earliest known divergence among multicellular animals [Dellaporta et al., 2006]. Yet our analysis of the protein alignment showed that the Trichoplax adhaerens protein has a higher degree of similarity with the insect and vertebrate proteins than does the protein of Caenorhabditis elegans (see Supplement 2). From this we conclude that NP_492389.1 from Caenorhabditis elegans is not a true functional ortholog of Hyd. And since we found no other potential orthologs in whole genomic nucleotide and protein searches of this organism’s genome, we conclude the true Hyd ortholog was lost in the Caenorhabditis roundworm lineage or evolved abnormally fast and lost normal Hyd features.

RNA interference silencing of the gene coding for the protein NP_492389.1 of Caenorhabditis elegans did not produce lethality or sterility [Sönrichsen et al., 2005]. In contrast, mutations in Hyd in Drosophila result in evident larval lethality [Gateff, 1994]. There is more evidence that NP_492389.1 from Caenorhabditis elegans does not function in the manner of conventional Hyd.

Thus, the earliest discernable evolutionary appearance of the Hyd protein family is with Trichoplax adhaerens. This probably indicates that proteins with the ability to suppress tumors were created at about the time that animalian multicellularity itself first appeared. Alternatively, it may indicate that archaic genomic characteristics established in the common unicellular ancestor had been readily co-opted for new purposes in late Precambrian times, an exceptional geological interval during which the marine environment was exceptionally enriched in molecular oxygen, phosphorus, and perhaps the degradation products of early members of the collagen family of molecules [Saul, 2008].

As mentioned, Hyd proteins have a PABC domain with evident similarity to the PABC domain of mRNA-binding PABP proteins [Oughtred et al., 2002]. The alignment by Oughtred et al., 2002, of the PABC domains of the PABP protein family and of representatives of Hyd proteins are shown in Supplement 3 from which it can be seen that the PABC domain of Hyd (XP_002114509.1, Trichoplax adhaerens; P51592, Drosophila melanogaster; XP_001923288.1, Danio rerio) have a high degree of similarity to PABC of PABP proteins. By contrast, NP_492389.1 of Caenorhabditis elegans and the outgroup protein P33202 of Saccharomyces cerevisiae do not have any essential similarity. From our inference that the protein NP_492389.1 of Caenorhabditis elegans is not a true Hyd homolog, we are tempted to assume that the PABC domain had been transferred to the Hyd protein later than the divergence of NP_492389.1 from a progenitor ubiquitin ligase, followed by positive selection in the context of cells within tissues. It is also fair to assume this transfer of genetic information was a final or near-final event in the formation of Hyd tumor suppressor protein family.

In sum, the ability of certain protein families to function as proliferation-inhibitors at the tissue level appears to have emerged nearly simultaneously with the first metazoan tissues themselves. This provides indirect support for results obtained by Saul and Schwartz, 2007, from paleontological studies. The main argument for a very early metazoan origin of cancer is based on data in the now defunct “Smithsonian Registry of Tumors in Lower Animals” in which cancer was documented in four phyla, chordates, molluscs, arthropods and flatworms, and neoplasms in three others Cnidaria, Sipuncula and Annelida [Harshbarger and Gibson, 1992; Saul and Schwartz, 2007].

ACKNOWLEDGEMENTS

The work was supported by RFBR project 08-04-00265.
REFERENCES


- Fuja, T. J., Lin, F., Osann, K. E. and Bryant, P. J. (2004). Somatic mutations and altered expression of the candidate tumor suppressors CSNK1E, DLG1, and EDD/hHYD in mammary ductal carcinoma. Cancer Res. 64, 942-951.


