FEBS Letters 446 (1999) 6–8 FEBS 21644

Hypothesis

The nuclear protein PH5P of the inter-α-inhibitor superfamily: a missing link between poly(ADP-ribose)polymerase and the inter-α-inhibitor family and a novel actor of DNA repair?

L. Jean[a], J.L. Risler[b], T. Nagase[c], C. Coulouarn[a], N. Nomura[c], J.P. Salier[a,*]

[a] *INSERM Unit-78 and Institut Fédératif de Recherches Multidisciplinaires sur les Peptides, Rouen, France*
[b] *Laboratoire Génome et Informatique, Université de Versailles, Versailles, France*
[c] *Kazusa DNA Research Institute, Kisarazu, Japan*

Received 26 November 1998; received in revised form 27 January 1999

Abstract Poly(ADP-ribose)polymerase is a nuclear NAD-dependent enzyme and an essential nick sensor involved in cellular processes where nicking and rejoining of DNA strands are required. The inter-α-inhibitor family is comprized of several plasma proteins that all harbor one or more so-called heavy chains designated H1–H4. The latter originate from precursor polypeptides H1P–H4P whose upper two thirds are highly homologous. We now describe a novel protein that includes (i) a so-called BRCT domain found in many proteins involved in DNA repair, (ii) an area that is homologous to the NAD-dependent catalytic domain of poly(ADP-ribose)polymerase, (iii) an area that is homologous to the upper two thirds of precursor polypeptides H1P–H4P and (iv) a proline-rich region with a potential nuclear localization signal. This protein now designated PH5P points to as yet unsuspected links between poly(ADP-ribose)polymerase and the inter-α-inhibitor family and is likely to be involved in DNA repair.

Key words: DNA repair; Exon assembly; Homology; Nucleus; Proline-rich

1. Inter-α-inhibitor (IαI) family

A protein family consists of proteins that share at least 40% identical or conservative amino acid residues whereas a superfamily is made up of proteins that share a lower level of conservation along with a still unambiguous evolutionary relationship. The set of plasma proteins comprized of IαI and related molecules is collectively referred to as the IαI family. Our knowledge of this family is currently expanding in three directions, including novel structural features for the genes and proteins, insights of some regulatory steps involved in their mostly hepatic expression and, albeit to a lesser extent, an appraisal of functions for some family members in health and disease such as stabilization of extracellular matrices by virtue of a hyaluronan binding capacity and control of tumor cell growth [1]. Most IαI family members are made up of the assembly of evolutionarily related, so-called heavy (H) chains with one bikunin (B) chain that, as the name says, harbors two protease inhibitory domains of the Kunitz type. Four H chains are currently known and designated H1–H4. One or more of them specifically participate in the structure of a given molecule in the IαI family as examplified by the proteins designated IαIH1,2,B (i.e. IαI itself, made up of H1+H2+B), IαIH2,B, IαIH3,B, etc... [1]. In all mammals studied so far, the four H chains originate from precursor polypeptides H1P–H4P (HxPs) that are 900–930 amino acid residues long and are encoded by four genes, *ITIH1–ITIH4*. The most N-terminal 600 amino acid residues or so of all four HxPs are highly similar (>25% strictly identical residues are shared between them within a species). They all harbor a domain, about 160 residues in length, that is similar to the so-called von Willebrand type A domain and hence suggests a heterophilic binding capacity of the H chains [2,3]. In contrast, the C-terminal third of HxPs is quite variable from one precursor to another. Indeed, it is well conserved between H1P and H3P whereas it is more distantly related in H2P and it is completely different in H4P [1]. Notably, a unique proline-rich domain has been found in the C-terminus third of H4P [4]. Finally, a processing of H1P, H2P and H3P releases their C-terminal third and allows the N-terminal two thirds, i.e. H1, H2 and H3, to assemble with B prior to secretion whereas H4P is not processed nor does it assemble with B. Therefore, the hallmark of all four HxPs, and hence of the IαI family, is made of their strongly conserved N-terminal two thirds only.

2. Poly(ADP-ribose)polymerase (PARP)

PARP is a member of a superfamily of ADP-ribosyl transferases with a shared NAD+ binding fold [5]. PARP is a ubiquitous and chromatin-associated protein, 1014 residues in length, that catalyses the covalent addition of ADP-ribosyl groups from the substrate NAD+ to itself (automodification) and to a limited number of nuclear DNA binding proteins (heteromodification) whereby a resulting ADP-ribose polymer eventually extends up to 20–30 U within the cell nucleus. PARP is an essential nick sensor in cellular processes where nicking and rejoining of DNA strands are required. These cellular processes include differentiation, proliferation and tumor transformation as well as the recovery from DNA damage [6–8]. PARP is well conserved from *Drosophila* to humans and is comprized of three major domains, namely a N-terminal DNA binding domain with two zinc fingers and a bipartite

*Corresponding author. Fax: (33) (235) 14 85 41.
E-mail: jean-philippe.salier@univ-rouen.fr

Abbreviations: B, bikunin; BRCT, BRCA1 C-terminus domain; H, heavy; H1P, H1 precursor; HxP, any H1P–H4P; IαI, inter-α-inhibitor; NLS, nuclear localization signal; PARP, poly(ADP-ribose)polymerase

PII: S0014-5793(99)00173-8

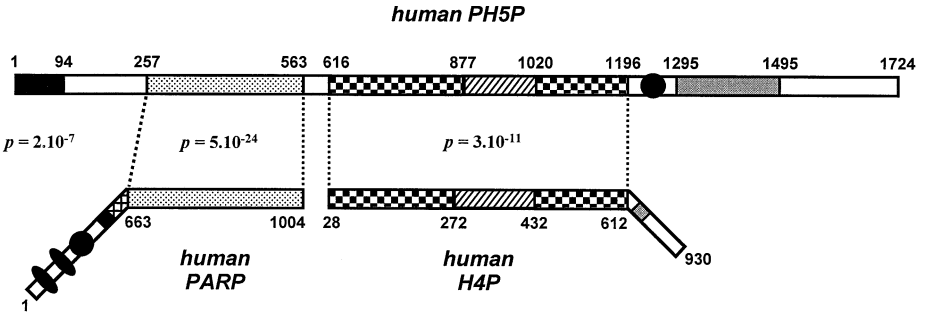

Fig. 1. Overview of PH5P primary structure and its homologies with PARP and the IαI family. The human PH5P (amino acid numbering as in [12]) is shown as a thick straight line. PH5P areas with similarities to human PARP or H4P are detailed whereas other areas that did not reveal any similarity with proteins in data banks are left blank. Human PARP and H4P are depicted below PH5P. In PARP and H4P, an area that presents an amino acid similarity with PH5P is shown as a thick horizontal line whereas all other areas (not drawn to scale) are shown at a 45° angle. Symbols: stippled area, PARP catalytic domain; checkers, hallmark of HxPs in the IαI family; hatched area, von Willebrand type A domain; closed circle, nuclear localization signal; grey area, proline-rich domain; closed box, BRCT domain; crossed hatched area, PARP automodification domain; closed ovals, zinc fingers in PARP DNA binding domain. The *p* values of a significant similarity between an area of PH5P and another polypeptide (PARP, H4P) or a domain (BRCT) as found by gapped BLAST [10] or Pfam [16] searches are indicated. The significant similarities of PH5P with the H1P, H2P and H3P polypeptides of the IαI family are not depicted.

nuclear localization signal (NLS), a central automodification domain and a C-terminal catalytic domain that harbors a highly conserved NAD+ binding site [7,9].

3. A protein with IαI- and PARP-related domains: PH5P

While searching for similarities in data banks with a new version of the BLAST algorithm [10], we have noted that within a novel human cDNA-deduced polypeptide as yet considered a PARP-like protein [11,12], two regions are very similar to the PARP catalytic domain and the hallmark of the IαI family, respectively. Two copies of this novel cDNA have now been cloned by two groups including ours [11,12] and the evidence excludes the possibility that this surprising association of two regions, one of which occurs within a nuclear protein and the other in a plasma protein, could result from cloning artifacts. Also, RT-PCR experiments in our laboratory (not detailed) have further excluded this possibility. As shown in Fig. 1, this novel protein is a large polypeptide with a N-terminal area that is homologous to the PARP catalytic domain, a central area that is homologous to the upper two thirds of HxPs and a C-terminal area that harbors a proline-rich domain (16% Pro residues in this domain versus 4% everywhere else). Accordingly, this novel protein, formerly named KIAA0177 [11], is now designated PH5P (PARP-related/IαI-related H5/proline-rich). The borders of the first two areas in PH5P fairly match the borders previously proposed for the catalytic domain in PARP and the hallmark of HxPs in the IαI family [1,9]. These areas in PARP and HxPs are encoded by a set of 12 exons each [13,14]. Therefore the *PH5P* gene likely resulted, at least partly, from the assembly of two sets of exons derived from the ancestral *PARP* and *ITIH* genes. This process involved more than a simple regional duplication of either ancestor as *PH5P* maps to human chromosome 13 [11,12] whereas *PARP*s map to chromosome 1 [13] and *ITIH1–ITIH4* map to chromosome 3 or 10 [1]. Extensive amino acid sequence alignments between PH5P and the upper two thirds of HxPs (not shown) indicate that the closest paralog of *PH5P* among the *ITIH* genes is *ITIH4*. Also, the location of a Pro-rich domain in the C-terminal area of PH5P is reminiscent of that of another Pro-rich, yet different, domain described in H4P (Fig. 1). Together, the IαI family and the PH5P protein now identify an IαI superfamily whose evolutionary tree is shown in Fig. 2.

4. PH5P as a DNA repair-related protein

Which functions can be ascribed to PH5P? The DPHFII motif that is found in H1P, H2P and H3P and participates in their intracellular assemblies with B [1] is absent at the end of the HxP-related area in PH5P which precludes a (PH5P+B) assembly. This conclusion also fits with the liver-restricted expression of B [1] as opposed to the ubiquitous expression of PH5P [11]. It is unlikely that PH5P binds NAD+ as the GXXXGKG motif required within the human PARP active site (P_{885}VTGYMFGKGIYFAD$_{899}$, where K_{893} is essential [7]) is abolished in PH5P (T_{467}DVGNLGSGIYFSD$_{480}$). Elsewhere, PH5P does not have the bipartite NLS found in the N-terminal DNA binding domain in PARP [9] but it harbors a

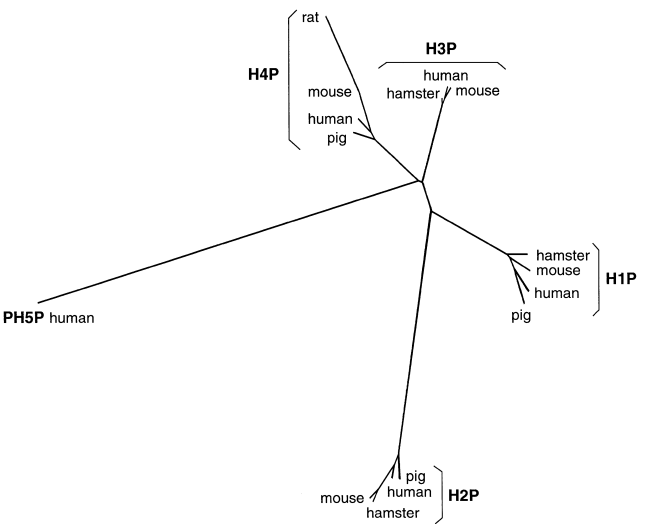

Fig. 2. Unrooted phylogenetic tree of the IαI superfamily. This tree is based on all available cDNA-deduced sequences covering the upper two thirds of HxPs from human, pig and rodents as well as their counterpart within human PH5P. The pairwise distance matrix was calculated by the program Protdist from the PHYLIP package (Dayhoff option) and the neighbor joining tree was obtained by the program Neighbor from the same package [19].

K_{1237}RKHRKIPFSKRK$_{1249}$ sequence (closed circle in Fig. 1) that perfectly matches the consensus for a bipartite NLS [9,15]. Therefore, like PARP, PH5P is likely to have a nuclear location. The von Willebrand type A domain found in PH5P is widespread in adhesive proteins and receptors [2] and hence its presence suggests a heterophilic binding capacity of PH5P for a polypeptide target (yet to be identified). Finally, from a similarity search carried out according to [16] we have found a so-called BRCT domain covering the first 94 residues in PH5P (Fig. 1). Single or multiple copies of such a BRCT domain have been previously observed in many proteins (e.g. the breast cancer-associated protein 1, the retinoblastoma protein, p53 binding proteins and PARP) that all participate in DNA damage-responsive checkpoints via BRCT-mediated, homo or heterophilic protein-protein interactions [17,18]. Therefore, PH5P can now be envisioned as a ubiquitous nuclear protein whose von Willebrand type A and BRCT domains allow for molecular interactions that mediate DNA repair-related functions.

Acknowledgements: We are indebted to Prof. W.J. Larsen (Cincinnati) for a critical reading of the manuscript.

References

[1] Salier, J.P., Rouet, P., Raguenez, G. and Daveau, M. (1996) Biochem. J. 315, 1–9.
[2] Bork, P. and Rohde, K. (1991) Biochem. J. 279, 908–910.
[3] Chan, P., Risler, J.L., Raguenez, G. and Salier, J.P. (1995) Biochem. J. 306, 505–512.
[4] Soury, E., Olivier, E., Daveau, M., Hiron, M., Claeyssens, S., Risler, J.L. and Salier, J.P. (1998) Biochem. Biophys. Res. Commun. 243, 522–530.
[5] Ruf, A., Menissier-De Murcia, J., De Murcia, G.M. and Schulz, G.E. (1996) Proc. Natl. Acad. Sci. USA 93, 7481–7485.
[6] De Murcia, G., Menissier-De Murcia, J. and Schreiber, V. (1991) BioEssays 13, 455–462.
[7] De Murcia, G. and Menissier-De Murcia, J. (1994) Trends Biochem. Sci. 19, 172–176.
[8] Le Rhun, Y., Kirkland, J.B. and Shah, G.M. (1998) Biochem. Biophys. Res. Commun. 245, 1–10.
[9] Schreiber, V., Molinete, M., Bœuf, H., De Murcia, G. and Menissier-De Murcia, J. (1992) EMBO J. 11, 3263–3269.
[10] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[11] Nagase, T., Seki, N., Ishikawa, K., Tanaka, A. and Nomura, N. (1996) DNA Res. 3, 17–24.
[12] Still, I.H., Vince, P. and Cowel J.K. (1998) EMBL accession #AF057160.
[13] Auer, B., Nagl, U., Herzog, H., Schneider, R. and Schweiger, M. (1989) DNA 8, 575–580.
[14] Diarra-Mehrpour, M., Sarafan, N., Bourguignon, J., Bonnet, F., Bost, F. and Martin, J.P. (1998) J. Biol. Chem. 273, 26809–26819.
[15] Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991) Cell 64, 615–623.
[16] Sonnhammer, E.L.L., Eddy, S.R., Birney, E., Bateman, A. and Durbin, R. (1998) Nucleic Acids Res. 26, 320–322.
[17] Bork, P., Hofmann, K., Bucher, P., Neuwald, A.F., Altschul, S.F. and Koonin, E.V. (1997) FASEB J. 11, 68–76.
[18] Zhang, X., Morera, S., Bates, P.A., Whitehead, P.C., Coffer, A.I., Hainbucher, K., Nash, R.A., Sternberg, M.J.E., Lindahl, T. and Freemont, P.S. (1998) EMBO J. 17, 6404–6411.
[19] Felsenstein, J. (1993) http://evolution.genetics.washington.edu/phylip.