Efficient trafficking of acidic proteins out of the endoplasmic reticulum involves a conserved amino terminal IleProVal (IPV)-like tripeptide motif

Anna S. Nam, **Ying Yin**, **Zofia von Marschall**, and **Larry W. Fisher**[*]
Matrix Biochemistry Section, CSDB, NIDCR, NIH, Bethesda, MD

Keywords

DSPP; dentin sialophosphoprotein; dentinogenesis imperfecta; dentin dysplasia; endoplasmic reticulum; trafficking; cargo receptor

Most proposed extracellular biomineralization processes include the secretion of proteins that interact with mineral ions and/or mineral surfaces. Typically these proteins are acidic or have substantial acidic domains that interact with multivalent cations (calcium, iron, magnesium, silicon, etc.) in the extracellular environment. One important challenge to this process is that secreted proteins are first made in the Ca^{2+}-rich (1-10mM) endoplasmic reticulum (ER) where Ca^{2+}-bridged aggregation could occur unless the proteins are rapidly transported out of this organelle. Recently our studies on the mechanisms by which mutations in the dentin sialophosphoprotein gene, *DSPP*, cause both the milder dentin dysplasia (DD) and the more severe dentinogenesis imperfecta (DGI) nonsyndromic genetic diseases of dentin illustrate what can happen when acidic proteins fail to traffic rapidly out of the ER. DSPP is the most acidic protein in the human body (pI 3.3 before the addition of >100 phosphate groups in the Golgi) with a tremendous potential to bind calcium ions particularly through its ~230 serine-serine-aspartic acid (SSD) tripeptide repeat carboxyterminal domain (pI 2.4) called dentin phosphoprotein (DPP). In mice, a single null allele (i.e., 50% of normal DSPP amount is made and secreted) has no disease phenotype (Dr. Ashok Kulkarni, personal communication) suggesting that mouse dentin can mineralize normally with only half of its normal complement of DSPP protein. The loss of both mouse *Dspp* alleles, however, causes a DGI-like phenotype with incompletely mineralized dentin [1]. Therefore, true *DSPP* null alleles in humans will probably also be shown in the future to be recessive, with the DGI-like double-null patients appearing predominantly in consanguineous populations [2].

Our recent experiments have shown that all of the known *DSPP* mutations have dominant negative effects whereby expression of the mutant DSPP is worse than the loss of a single allele's protein (e.g. null allele) because the retained mutant protein (except Y6D) also

[*]corresponding author: Room 223, Building 30, 9000 Rockville Pike, Bethesda, MD, 20892-4320. Phone: 301-496-5769, Fax: 301-402-0824, lfisher@dir.nidcr.nih.gov.

disrupts secretion of some of the normal allele's DSPP [3]. Mutations more efficient at disrupting the trafficking of the normal allele's DSPP causes the more severe dentinogenesis imperfecta (DGI), while a larger amount of the wildtype protein appears to be successfully secreted by odontoblasts in patients with the milder dentin dysplasia (DD) [3].

Our hypothesis is that the mutant DSPP retained within the ER captures some of the normal DSPP protein via Ca^{2+}-bridges, particularly between their SSD repeat domains. Frameshift mutations (loss of either 1 or 4 bases) within the SSD-encoding repeat domain causing DD and DGI were particularly useful in discovering this process. These -1 frameshifts cause the normal SSDSSDSSD… to become hydrophobic amino acids (e.g. VVTAVIVVT…). The new hydrophobic domain behaves as a repeating transmembrane motif, causing the mutant DSPP to become associated with the ER membrane and fail to traffic out of the organelle. Importantly, mutations that occur early in the repeat domain result in more hydrophobic amino acids but the milder form of the disease (DD). More 3' mutations (i.e., more carboxyterminal) result in fewer mutant hydrophobic amino acids but the more severe DGI disease in all affected family members. This seemingly contradictory effect of more mutant amino acids resulting in the milder disease can be explained by considering the presence of more intact SSD repeats remaining in the DGI patients' mutant proteins. Longer stretches of Ca^{2+}-binding SSD repeats that remain can capture more of the DSPP made by the normal allele, resulting in less total DSPP being secreted out of the odontoblasts and thereby the more severe DGI disease.

The insights obtained from *DSPP* frameshift mutations were then used to understand how the many single base change *DSPP* mutations in the 5' region of the gene cause disease. First, the A15V mutation causes the hydrophobic leader sequence to be predominantly retained resulting in its continued association with the ER membranes [3]. With its full complement of SSD repeats intact, this mutation probably causes DGI by the same mechanism as the more 3' frameshift mutations. Next, we proposed earlier that all other mutations found near the beginning of the sequence (except Y6D) ultimately end up changing the first three amino acids of the mature protein, IleProVal (IPV) (Figure 1). This can happen by direct missense mutations (e.g., ISV, ITV, IPD) or by a variety of splice-site single basepair mutations (most in introns 2 or 3) resulting in exon 3 being skipped (see ref 2). Skipping exon 3 links the first amino acid from exon 4, aspartic acid (D), directly to the last amino acids of exon 2 (IP) to also make IPD. (We noted that in silico predictions of the only reported nonsense mutation, Q45X, changes the last codon of exon 4 and actually results in the skipping of exon 3 and therefore the same mutant IPD starting motif. [2]) Reporter constructs such as GFP trafficked normally out of the cell when fused to either leader+IPV, leader+IPD, or other mutant IPV motifs. The acidic DSPP protein, however, traffics out of the ER only when starting with leader+IPV [3]. This led us to hypothesize that DSPP relies on the amino-terminal IPV motif to interact with a cargo receptor for rapid trafficking out of the ER before it accumulated to sufficient concentrations to form a gel or precipitate in the ER's 1 mM Ca^{2+}. Because all of the IPV-mutant DSPP proteins have their entire calcium-binding SSD repeat domains intact, their retention within the ER explains the observation that nearly all 5' *DSPP* mutations cause the more severe DGI disease. Even the rare 5' *DSPP* mutation (IVS2-6T>G) that causes the milder DD disease fits the proposed

model because it causes exon 3 to be skipped only about 50% of the time [4] thereby raising the production of wildtype DSPP to ~75% with production of problematic IPD-mutant DSPP limited to ~25% of the total. This lesser amount of retained mutant DSPP apparently still captures enough of the more abundant normal DSPP to cause the mild DD phenotype (i.e. <50% of that produced by the normal population with two wildtype alleles) but not enough to cause the DGI like other 5' mutations [3].

We started investigating mutations in *DSPP* to understand the functions of this acidic protein in the formation and mineralization of dentin matrix. However, it has turned out that all published mutations actually result in trafficking errors not in failed functions within the extracellular environment. This has led us to investigate if other secreted acidic proteins may also start with IPV-like sequences in order to use the proposed cargo receptor to traffic out of the ER before they can too can accumulate to levels sufficient to precipitate in this calcium-rich environment. Three other members of the SIBLING family also start with IPV-like sequences and substituting the first ~30 amino acids of DSPP with the starting sequences of either DMP1 or OPN permit the efficient trafficking out of the cell [3]. The IPV-like starting sequence of another highly acidic protein (not in the SIBLING family), the neuroendocrine-associated protein chromogranin A, also supported efficient trafficking of DSPP. In all three hybrid-protein cases, changing the third amino acid from the hydrophobic amino acid valine (V) to the negatively charged aspartic acid (D) caused retention of the protein within the cell [3].

Search of the literature and protein databases resulted in a variety of acidic proteins secreted from mineralizing as well as non-mineralizing tissues that start with IPV-like hydrophobic-proline-hydrophobic (Φ-P-Φ) sequences. Many of the proteins associated with human enamel production, for example, start with Φ-P-Φ motifs including enamelin (MPM), amelogenin (MPL), ameloblastin (VPF) and ODAM (APL). Human amelotin (LPQ) represents a second class of starting motifs whereby one of the hydrophobic amino acids is replaced by a polar but not charged amino acid (H, N, Q, S, T or Y). Indeed, even the highly acidic SIBLING DMP1 in a reptile (caiman) has an uncharged, polar amino acid (HPV) as do two human bone-associated proteins, SPARC (osteopontin, APQ) and SPARCL1 (IPT). Several prohormones [e.x. prolactin (LPI), somatotropin (FPT), leptin (VPI)] and the neuroendocrine secretion-associated proteins chromogranins A and B (LPV and MPV respectively) as well as other blood proteins, fetuin (APH), antithrombin-III (SPV) and complement C3 (SPM), are representative of other vertebrate acidic proteins that have conserved IPV-like motifs (see Table 1).

To see if the proposed IPV-motif cargo receptor is conserved back through evolution, we next looked at secreted acidic proteins, particularly those reported to be associated with more ancient mineralization systems. The amphibian Xenopus, for example, secretes a calcium-binding phosphoprotein that starts with IPI as well as an acidic antibiotic protein, magainin, starting with LPQ. The most abundant protein in the sea urchin tooth is a DMP1-like acidic protein called phosphodontin [5] and starts with an APV motif. Among the mollusks, an acidic protein, MSP-1[6], starts with APL. APV and MPA respectively start two acidic coral mineral-associated proteins, CARP2 and CARP3 [7]. The presence of the IPV-motif cargo receptor in several species of yeast is suggested by the conservation of the

IPV-like motif in the amino termini of two acidic proteins, the alpha mating factor (APV or IPV in various species) and in the magainin-like antibiotic killer protein carried by some yeast strains (LPS). Indeed, much of the industry of making recombinant proteins in yeast involve fusion proteins that use the IPV-like motif of the alpha mating factor's prodomain to ensure efficient trafficking of the proteins.

In conclusion, proteins that can form aggregates must be rapidly trafficked out of the ER before they reach critical concentrations. Because of the high levels of Ca^{2+} ions within the ER, calcium-binding proteins are one broad class of proteins that may come under evolutionary pressure to be trafficked out before the divalent ions form bridges between the accumulating proteins and cause precipitation. Mutations in the IPV-motif of DSPP that either add a negatively charged amino acid or remove the proline cause the proteins to be retained within the ER, similar to mutations that add membrane-associating hydrophobic domains to the otherwise completely hydrophilic, disordered protein. Because many other acidic proteins secreted by mineralizing and non-mineralizing cells also start with IPV-like sequences, we propose that the rapid removal of proteins with the ability to bind multiple calcium ions is facilitated by receptors in the ER that interact with the IPV-motif. Furthermore, the IPV-like motif is conserved over evolution suggesting that this problem was first solved in early eukaryotes. We have found that yeast have conserved IPV-like motifs starting their acidic proteins and likely represent cells first to use a cargo receptor to rapidly traffic acidic proteins out of the ER.

References

1. Sreenath T, Thyagarajan T, Hall B, Longenecker G, D'Souza R, Hong S, Wright JT, MacDougall M, Sauk J, Kulkarni AB. Dentin sialophosphoprotein knockout mouse teeth display widened predentin zone and develop defective dentin mineralization similar to human dentinogenesis imperfecta type III. J Biol Chem. Jul 4; 2003 278(27):24874–80. [PubMed: 12721295]

2. McKnight DA, Suzanne Hart P, Hart TC, Hartsfield JK, Wilson A, Wright JT, Fisher LW. A comprehensive analysis of normal variation and disease-causing mutations in the human DSPP gene. Hum Mutat. Dec; 2008 29(12):1392–404. [PubMed: 18521831]

3. von Marschall Z, Mok S, Phillips MD, McKnight DA, Fisher LW. Rough endoplasmic reticulum trafficking errors by different classes of mutant dentin sialophosphoprotein (DSPP) cause dominant negative effects in both dentinogenesis imperfecta and dentin dysplasia by entrapping normal DSPP. J Bone Miner Res. Jun; 2012 27(6):1309–21. [PubMed: 22392858]

4. Lee SK, Hu JC, Lee KE, Simmer JP, Kim JW. A dentin sialophosphoprotein mutation that partially disrupts a splice acceptor site causes type II dentin dysplasia. J Endod. Dec; 2008 34(12):1470–3. [PubMed: 19026876]

5. Mann K, Poustka AJ, Mann M. Phosphoproteomes of Strongylocentrotus purpuratus shell and tooth matrix: identification of a major acidic sea urchin tooth phosphoprotein, phosphodontin. Proteome Sci. Feb 8.2010 8(1):6. doi: 10.1186/1477-5956-8-6. [PubMed: 20181113]

6. Sarashina I, Endo K. The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop Patinopecten yessoensis. Mar Biotechnol (NY). Jul; 2001 3(4):362–9. [PubMed: 14961352]

7. Mass T, Drake JL, Haramaty L, Kim JD, Zelzion E, Bhattacharya D, Falkowski PG. Cloning and characterization of four novel coral acid-rich proteins that precipitate carbonates in vitro. Curr Biol. Jun 17; 2013 23(12):1126–31. [PubMed: 23746634]

Figure 1.
Legend: Sequence of human *DSPP* illustrating 5' mutations. Normal sequence is on main line with mutation amino acids noted above. Grey highlights indicate dentin dysplasia, black highlights indicate the more severe dentinogenesis imperfecta. Intron base changes that cause skipping of Exon 3 are similarly indicated. * denotes proposed amino acid change that actually results in the skipping of Exon 3. Leader sequence is underlined.

Table 1

Starting three amino acids of selected acidic proteins

```
                              SIBLINGs
IPV  hDSPP (pI 3.3)            IPV for opossum, VPV for elephant
IPV  hOPN (pI 4.3)            LPV for platypus, WPV for chicken
LPV  hDMP1 (pI 4.0)           LPV for platypus, HPV for caiman
APT  hMEPE (pI 8.3)           APT for hedgehog, TPV for chicken
                        Other bone proteins:
APQ  hSPARC (pI 4.7)          APQ for platypus, APQ for chicken
IPT  hSPARCL1 (pI 4.5)        IPT for whale,    IPT for armadillo
                         Enamel Proteins:
MPM  hEnamelin (pI 6.2)       MPM for pig,         VPM for crocodile
MPL  hAmelogenin (pI 6.5)     IPL for opossum,  VPL for frog
VPF  hAmeloblastin (pI 4.7)   VPI for opossum,  FPI for frog
LPQ  hAmelotin (pI 4.7)       LPM for cow,      LPT for manatee
APL  hODAM (pI 4.5)           APL for armadillo, LPL for frog
                          Neuroendocrine:
LPV  hChromogranin A (pI 4.6) LPF for opossum, LPV for chicken
MPV  hChromogranin B (pI 5.0) MPV for cow,     APV for rat
                            Hormones:
LPI  hProlactin (pI 6.4)      LPI for opossum, API for fish
FPT  hSomatotropin (pI 5.2)   FPA for opossum, FPS for frog
VPI  hLeptin (pI 6.0)         VPI for opossum, IPV for fish
                        Blood proteins:
APH  hFetuin (pI 5.4)            TPV for guinea pig, LPT for fish
SPV  hAntithrombin-III (pI 6.0), NPV for guinea pig, APY for bird
SPM  hComplement C3 (pI 6.1)     QPL for platypus,   VPT for bird
                      Misc. Acidic proteins:
VPL  hNucleobindin-1 (pI 5.1) VPL for whale, VPI for frog
VPI  hNucleobindin-2 (pI 5.0) VPI for whale, VPI for opossum
More distantly related: secreted, acidic proteins:
APV  Phosphodontin, sea urchin tooth protein (pI 3.9)
MPA  CARP3, coral mineral-associated protein (pI 3.0)
APV  CARP2, coral mineral-associated protein (pI 4.6)
APL  MSP-1, mollusk shell protein (pI 3.3)
APV  Yeast mating factor alpha (pI 4.5) (IPV for some species)

   Grey = hydrophobic amino acid, Yellow = polar but not charged
```