Archives of Oral Biology (2005) **50**, 599–609

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Archives of Oral Biology

www.intl.elsevierhealth.com/journals/arob

REVIEW

A review of protein structure and gene organisation for proteins associated with mineralised tissue and calcium phosphate stabilisation encoded on human chromosome 4

N. Laila Huq, Keith J. Cross, Men Ung, Eric C. Reynolds *

Cooperative Research Centre for Oral Health Science, School of Dental Science, The University of Melbourne, 711 Elizabeth Street, Melbourne, Vic. 3010, Australia

Accepted 23 December 2004

KEYWORDS
Multiple phosphoseryl-containing proteins;
Conserved chromosomal synteny;
Codon usage bias;
Gene duplication

Summary Several proteins associated with mineralised tissue (teeth and bone) or involved in calcium phosphate stabilisation in the body fluids, milk and saliva have been mapped to the q arm of human chromosome 4. These include the dentine/bone proteins dentine sialophosphoprotein (DSPP), dentine matrix protein 1 (DMP1), bone sialoprotein (BSP), matrix extracellular phosphoglycoprotein, osteopontin (OPN), enamelin, ameloblastin, milk caseins, salivary statherin, and proline-rich proteins. The proposed function of those that are multiphosphorylated is: (i) the stabilisation of calcium phosphate in solution (e.g. casein, statherin) preventing spontaneous precipitation and seeded-crystal growth or (ii) promoting biomineralisation (e.g. the phosphophoryn domain of DSPP), where the protein described as a template macromolecule, is proposed to act as a nucleator/promoter of crystal growth. The genes of these proteins have been subjected to conserved chromosomal synteny during mammalian evolution. The multiphosphorylated proteins statherin, caseins, phosphophoryn, BSP and OPN have been characterised as intrinsically disordered. The codon usage patterns for the amino acid serine reveal a bias for AGC and AGT codons within the human genes *dspp*, *dmp1* and *bsp*, mouse *dspp* and *dmp1* but not significantly for statherin or caseins. This pattern was also observed in the gene encoding hen phosvitin that also contains stretches of multiphosphorylated serines and in the *dmp1* gene sequences of mammalian, reptilian and avian classes. In conclusion, these intrinsically disordered multiphosphorylated proteins are the translation products of genes displaying examples of codon usage bias, internal repeats and conserved chromosomal synteny within the mammalian class.

* Corresponding author. Tel.: +61 3 9341 0270; fax: +61 3 9341 0236.
E-mail address: e.reynolds@unimelb.edu.au (E.C. Reynolds).

0003-9969/$ – see front matter
doi:10.1016/j.archoralbio.2004.12.009

Contents

Introduction

Several proteins involved in the stabilisation of calcium phosphate in body fluids and/or associated with mineralised tissue (teeth and bone) are encoded on chromosome 4 in the human genome (Table 1 and Fig. 1). These include the phosphophoryn domain and dentine sialoprotein domains of the dentine sialophosphoprotein (DSPP),[1] dentine matrix protein 1 (DMP1),[2] bone sialoprotein (BSP),[3] matrix extracellular phosphoglycoprotein (MEPE)[4] and osteopontin (OPN)[5] generally found in hard tissues; enamelin (ENAM)[6] and ameloblastin (AMBN)[7] found in enamel; statherin (STATH),[8] histatin (HTN)[9] and proline-rich proteins (PROL)[10] found in saliva[11]; and the caseins (CSN)[12] found in milk. Most of the proteins are post-translationally modified being phosphorylated and/or glycosylated. Many of those phosphorylated proteins contain multiple phosphoseryl residues in clusters. The proposed functions of the multiphosporylated proteins are: (i) the stabilisation of calcium phosphate in solution (e.g. casein, statherin) preventing spontaneous precipitation and seeded-crystal growth and/or (ii) biomineralisation (e.g. phosphophoryn), where the protein, described as a template macromolecule, is proposed to act as a nucleator/promoter of crystal growth. The tissue distribution, post-translational modifications and proposed functions of these proteins are summarised in Table 1. Some of the proteins are proposed to have multifunctions and these are discussed in more detail later.

Dentine and bone extracellular proteins (the SIBLING family)

Within a 375 kb region on human chromosome 4q21 (Fig. 1), there is a cluster of genes coding for the proteins DSPP, DMP1, BSP, MEPE and OPN that have similar exon structures. Exon 1 is non-coding. Exon 2 encodes for the leader sequence plus the first two residues of the mature protein. Exons 3 and 5 often contain the consensus sequence for casein kinase II phosphorylation (SSEE). Exon 4 is usually proline-rich. The last one or two exons encode the vast majority of the protein and usually contain the integrin-binding tri-peptide Arg—Gly—Asp (RGD). Since BSP and OPN (both ~300 aa) and DMP1 (513—657 aa) are small molecules, these proteins are known as the Small Integrin-Binding LIgand N-linked Glycoprotein (SIBLING) family.[13,14] This terminology has limitations as ''OPN typically may not be *N*-glycosylated''[15] and the sizes of the DSPP protein can vary with the species as shown in Fig. 2. It has been suggested that the genes coding for the proteins DSPP, DMP1, OPN, BSP and MEPE have arisen from a common ancestor.[7,14]

Enamel matrix proteins

Two enamel matrix proteins ENAM and AMBN are coded by genes located on human chromosome 4q13.3. The human *AMBN* gene has 13 exons coding for a small protein of 447 residues. The human *ENAM* gene has eight exons coding for a protein that is rich in prolines. The third protein believed to be involved