Dev Genes Evol (2009) 219:509–521
DOI 10.1007/s00427-009-0313-9

ORIGINAL ARTICLE

Molecular evolution of SPARC: absence of the acidic module and expression in the endoderm of the starlet sea anemone, *Nematostella vectensis*

Anne Koehler · Sherwin Desser · Belinda Chang · Jacqueline MacDonald · Ulrich Tepass · Maurice Ringuette

Received: 20 August 2009 / Accepted: 2 December 2009 / Published online: 31 December 2009

Abstract The matricellular glycoprotein SPARC is composed of three functional domains that are evolutionarily conserved in organisms ranging from nematodes to mammals: a Ca^{2+}-binding glutamic acid-rich acidic domain at the N-terminus (domain I), a follistatin-like module (domain II), and an extracellular Ca^{2+}-binding (EC) module that contains two EF-hands and two collagen-binding epitopes (domain III). We report that four SPARC orthologs (designated nvSPARC1-4) are expressed by the genome of the starlet anemone *Nematostella vectensis*, a diploblastic basal cnidarian composed of an ectoderm and endoderm separated by collagen-based mesoglea. We also report that domain I is absent from all *N. vectensis* SPARC orthologs. *In situ* hybridization data indicate that *N. vectensis* SPARC mRNAs are restricted to the endoderm during post-gastrula development. The absence of the Ca^{2+}-binding N-terminal domain in cnidarians and conservation of collagen-binding epitopes suggests that SPARC first evolved as a collagen-binding matricellular glycoprotein, an interaction likely to be dependent on the binding of Ca^{2+}-ions to the two EF-hands in the EC domain. We propose that further Ca^{2+}-dependent activities emerged with the acquisition of an acidic N-terminal module in triplobastic organisms.

Keywords SPARC · *Nematostella vectensis* · Gene duplication · Calcium binding · Collagen binding · Endoderm

Communicated by M.Q. Martindale

Electronic supplementary material The online version of this article (doi:10.1007/s00427-009-0313-9) contains supplementary material, which is available to authorized users.

A. Koehler
Laboratory of Tissue Repair and Regeneration,
Matrix Dynamics Group, University of Toronto,
150 College Street,
Toronto, ON, Canada M5S 3E2

S. Desser · U. Tepass · M. Ringuette (✉)
Department of Cell and Systems Biology, University of Toronto,
25 Harbord Street,
Toronto, ON, Canada M5S 3G5
e-mail: ringuet@csb.utoronto.ca

B. Chang
Department of Cell and Systems Biology,
Department of Ecology and Evolutionary Biology,
Centre for Analysis of Genome Evolution and Function,
University of Toronto,
25 Harbord Street,
Toronto, ON, Canada M5S 3G5

J. MacDonald
Department of Chemical Engineering & Applied Chemistry,
University of Toronto,
200 College St.,
Toronto, ON, Canada M5S 3E5

Introduction

The evolution of multicellular organisms and their adaptation to different environments is dependent on a complex, heterogeneous array of extracellular matrix (ECM) macromolecules. Collagens and other polymer-forming components assemble into elaborate scaffolds that function as the prime determinants of tissues and organ shape, stability, and biomechanical behaviors. In contrast, matricellular glycoproteins are not integrated within matrix scaffolds, but function as transient, diffusible components of matrices with potent effects on cell–cell and cell–matrix interactions, which in turn affect cellular behaviors and physiological

functions. Matricellular glycoproteins, which include the SPARC family, represent a diverse group of structurally and functionally distinct matrix elements, several of which are expressed at high levels during early embryonic development, decreasing to near basal levels in mature tissues (Damjanovski et al. 1998; Martinek et al. 2007; Midwood et al. 2004). During embryonic development, they act in concert to promote proper assembly and maturation of matrix scaffolds, modulate growth factors and cytokine signaling, and regulate several other aspects of development. The expression of many matricellular glycoproteins is re-activated in tissues undergoing remodeling and in response to injury. A hallmark of loss in cell-stroma homeostasis, such as in cancer, is the de-regulation of matricellular glycoproteins (Bradshaw and Sage 2001; Brown et al. 1999; Sodek et al. 2008).

SPARC is a small (33 kDa) Ca^{2+}- and collagen-binding glycoprotein whose trimodular organization is highly conserved in animal phyla, as far ranging as nematodes to mammals, examined to date. The three domains/modules of SPARC include a glutamic acid-rich N-terminal domain (I) that can bind up to eight Ca^{+2} ions with low affinity [K_D= 10^{-3}–10^{-5}M], rendering SPARC sensitive to changes in ECM Ca^{+2} levels as well as promoting its enrichment within calcified tissues (Maurer et al. 1995). Indeed, SPARC was first identified as osteonectin, a major non-collagenous component of bone (Termine et al. 1981). The second domain (II) contains a follistatin-like (FS) module with 10 conserved cysteine residues and an N-linked glycosylation site in triploblastic organisms. The C-terminal domain (III) is an alpha helix-rich region containing two high-affinity Ca^{2+}-binding EF-hands (EF-hand1 and EF-hand2); hence, it is also referred to as the extracellular calcium domain (EC domain). The C-terminal EF-hand2 is stabilized by a disulfide bridge, a common feature of the SPARC family. High-affinity binding of Ca^{2+} to EF-hand1 is dependent on cooperative interactions between the FS and EC domain pair (Maurer et al. 1995). The affinity of SPARC for fibril-forming collagens (collagen I, II, III, and V) and network-forming collagen IV is dependent on two conserved epitopes in the EC domain (Sasaki et al. 1997, 1998). The presence of an FS and EC domain is a defining feature of the SPARC family, which includes SPARC-like 1 (SC1/Hevin and Mast9), Testican-1, -2, -3, and SMOC (SPARC-related modular calcium-binding)-1 and -2 (Johnston et al. 1990; Vannahme et al. 1999, 2002, 2003).

Mammalian SPARC has been demonstrated *in vitro* to have diverse biological functions, including inhibition of cell–matrix adhesion, G1 to S phase cell cycle progression, modulation of VEGF, PDGF, bFGF, and TGF-β1 signaling, regulation of matrix remodeling metalloproteinase expression, and the expression of other matrix glycoproteins (Bassuk et al. 2000; Francki et al. 1999; Lane and Sage 1994; Motamed et al. 2002, 2003). Subsets of these diverse biological activities are replicated by mimetic synthetic peptides (20mers) corresponding to distinct functional domains. For example, several of these peptides mimic the pro-angiogenic, counteradhesive, and anti-proliferative activities of native SPARC (Huynh et al. 1999; Lane and Sage 1990). However, some peptides have opposing activities to native SPARC, raising the possibility that proteolytic fragments of SPARC generated during tissue remodeling may unmask activities that are not exhibited by full-length SPARC or that SPARC may have opposing activities in different tissues during development (Lane and Sage 1994; Tai and Tang 2008).

Despite the knowledge gained from in vitro studies, the precise morphogenetic functions of SPARC during development are poorly understood. Smaller, more uniform collagen I fibrils are observed in the dermis of *Sparc*-null mice relative to wild-type mice. Within a few months, cataracts become visible in *Sparc*-null mice, a deformity hypothesized to be attributed in part to decreased collagen IV stability in lens capsules (Greiling et al. 2009; Norose et al. 1998). SPARC is enriched in basal laminae of *Drosophila melanogaster* embryos, and in the absence of SPARC collagen IV, incorporation is de-regulated, consistent with the finding of compromised collagen maturation and/or stability in *Sparc*-null mice (Martinek et al. 2008; Rentz et al. 2007). In both *Ceanorhabditis elegans* and *D. melanogaster*, the absence of SPARC proves embryonic lethal (Fitzgerald and Schwarzbauer 1998; Martinek et al. 2007; Schwarzbauer et al. 1994; Schwarzbauer and Spencer 1993). These data combined with in vitro biochemical studies demonstrate that the binding of SPARC to several types of collagens (Hohenester et al. 1997, 2008) may be the basis for a functional relationship between SPARC and collagens that is conserved throughout evolution.

Despite the conservation and functional importance of SPARC across evolutionarily distant phyla ranging from nematodes to mammals, the origins and historical functions of this protein remain poorly resolved. Hence, the genome of the basal cnidarian, *Nematostella vectensis* was targeted to illuminate early SPARC evolution and function (Darling et al. 2005; Martindale 2005; Putnam et al. 2007). We report that the genome of this diploblastic cnidarian codes for four SPARC orthologs. With the exception of a SPARC-like paralog, which first appears in chordates (Kawasaki and Weiss 2006; Vannahme et al. 2003), all members of the SPARC family are encoded by the genome of *N. vectensis*, indicating that this family of matricellular glycoproteins arose early during the evolution of Metazoa. In contrast to SPARC in triploblastic organisms, all four *N. vectensis* SPARC orthologs lack an acidic N-terminal domain I, and the expression of SPARC genes is restricted to the endoderm during early *N. vectensis* development.

Materials and methods

In silico identification of *Sparc* genes

N. vectensis Sparc genes (*nvSparc*1, NEMVEDRAFT_v1g 238351; and *nvSparc*2, NEMVEDRAFT_v1g215469) were retrieved from the KOG classification section (Function ID: W, KOG 4004) of the Joint Genome Institute (JGI) genome server (http://genome.jgi-psf.org/cgi-bin/kogBrowser?expandclass=W&type=KOG&db=Nemve1). BLASTp searches of the Genbank database (http://blast.ncbi.nlm.nih.gov/Blast.cgi), using the amino acid sequences of nvSPARC1and nvSPARC2, identified two additional potential SPARC genes (*nvSparc*3, NEMVEDRAFT_v1g1 96727; and *nvSparc*4, NEMVEDRAFT_v1g245074) in the *N. vectensis* genome. nvSPARC orthologs had *E* values ranging from 30 to 7e−21 relative to each other and *E* values ranging from 1e−16 to 1e−6 relative to several invertebrate SPARC. While *E* values $\leq 10^{-10}$ are considered an experimental confirmation of gene identity, the higher *E* value of approximately 10^{-6} between cnidarian and non-cnidarian SPARC reflects the high conservation of the number and position of cysteine residues within the FS–EC domain pairs, a defining feature of SPARC.

5′RACE of nvSPARC3

As the JGI published gene of *nvSparc*3 appeared to be lacking a signal sequence we performed 5′ RACE (Invitrogen, Burlington, Ontario) according to manufacturer's specifications. Reverse transcription of total RNA from adult *N. vectensis* was primed using nvSPARC3-R1 (see Electronic supplementary materials, Table 1). Resultant cDNAs were PCR-amplified using primers APP (Invitrogen, Burlington, Ontario) and nvSPARC3-R2. An 800-bp amplicon was cloned into pCRII®TOPO® vector (Invitrogen, Burlington, Ontario), sequenced (AGTC, Toronto, Ontario), and verified by BLASTn on the *N. vectensis* JGI genome server.

Sequence alignments and phylogenetic analysis

SPARC protein sequences were aligned using ClustalW (Thompson et al. 1994; Larkin et al. 2007) and subjected to phylogenetic analyses using maximum likelihood and Bayesian methods (Guindon and Gascuel 2003; Ronquist and Huelsenbeck 2003). Maximum likelihood phylogenetic methods were implemented in the program PHYML 3.0 (Guindon and Gascuel 2003; Guindon et al. 2005), using the LG amino acid replacement matrix (Le and Gascuel 2008). For the likelihood analyses, bootstrapping methods were used to assess the degree of confidence in nodes of the phylogeny (Felsenstein 1985). Bayesian inference was performed in MrBayes 3.1.2, using a model that allows for jumping among fixed amino acid substitution rate matrices (Ronquist and Huelsenbeck 2003), with all of the SPARC sequence data in a single partition. Two Markov chain Monte Carlo runs were performed, with four chains each (three heated and one cold) for 10 million generations, sampling trees (and parameters) every 1,000 generations. Convergence was assessed using a number of methods. The average standard deviation of split frequencies, as calculated in MrBayes, were all well below 0.01 at stationarity. Also implemented in MrBayes, a convergence diagnostic for branch length posterior probabilities, is the potential scale reduction factor that roughly approached 1 as the runs converged (Gelman and Rubin 1992). Convergence to stationarity was also assessed by plotting log-likelihood scores and other parameter values in the program Tracer 1.4.1 (http://beast.bio.ed.ac.uk/Tracer; Drummond and Rambaut 2007), to ensure that there were no trends in the data post burn-in. Finally, adequacy of mixing was assessed by examining acceptance rates for parameters in MrBayes and by calculating in Tracer the effective sample sizes, the number of independent samples from the marginal posterior distribution for each parameter, with higher values being indicative of better sampling from the posterior distribution. These values were all well above 100. By these measures, convergence was achieved within the first 25% of trees sampled, which were discarded as burn-in, and remaining trees were taken as representative of the posterior probability distribution. For accession and ID numbers of all proteins used in the analyses, see Table 2 (Electronic supplementary materials).

Gene expression

Whole mount *in situ* hybridizations were performed as previously described by Martindale et al. (2004) with the following changes. Total RNA was isolated from mature polyps using the RNeasy mini kit (Qiagen, Mississauga, Ontario) followed by oligo-dT-primed cDNA synthesis (Fermentas, Burlington, Ontario). Gene-specific primers were designed using Primer3 (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (see Electronic supplementary materials, Table 1), and coding domains larger than 500-bp were amplified. The fragments were cloned into the pCRII®TOPO® vector (Invitrogen, Burlington, Ontario), sequenced (AGTC, Toronto, Ontario), and verified by BLASTn (Altschul et al. 1990). Due to the fact that NvSPARCs have only 32–35% amino acid sequence identity, BLASTx identified only one hit for each ortholog. Digoxygenin-labeled RNA sense and antisense probes were generated by *in vitro* transcription using the SP6/T7 transcription kit (Roche Diagnostics, Laval, Quebec) and quantified according to manufacturer's directions. Hybrid-

izations were performed under high stringency conditions (65°C for at least 24 h at probe concentrations of 1 ng/µl) on all stages of larval development. Probes were detected by incubation with an anti-digoxygenin antibody conjugated to alkaline phosphatase (Roche Diagnostics, Laval, Quebec) and a colorimetric reaction using the substrate NBT/BCIP. As a control, sense probes were used alongside the experimental procedure. Specimens were examined in DIC mode on a Zeiss Axioplan 2 and documented using a digital camera. Images were assembled and annotated in Adobe Photoshop CS3.

Results

Domain I of SPARC is absent in diploblastic organisms

Previous studies have shown that the trimodular organization of SPARC is conserved in triploblastic animals (Fig. 1). To determine whether the trimodular organization of SPARC also occurs in more basal metazoans, we conducted a BLASTp search of the recently completed and annotated draft of the genome of the starlet sea anemone. *N. vectensis* is a member of the class Anthozoa (Martindale 2005), which in addition to anemones includes corals and sea pens. We identified four EST entries derived from different genetic loci in the *N. vectensis* database having significant sequence similarity to the FS–EC domain pair of invertebrate and vertebrate SPARC proteins. While the acidic character of domain I is maintained in all triploblastic animals, the position of the acidic residues is not conserved. Consequently, no sequence similarity is detected by BLASTp using domain I from different animal phyla. For nvSPARC, BLASTp searches of individual paralogs indicate *E* values ranging from e−30 to 7e−21 relative to each other and *E* values ranging from 1e−16 to 1e−6 for several invertebrate SPARCs. While *E* values below 10^{-10} are considered as experimental confirmation of gene identity, the higher *E* value of approximately 10^{-6} used to identify SPARCs in *N. vectensis* is based primarily on the number and position of the cysteine residues that are signature amino acids within the FS–EC domain pairs. For simplicity, we designated the four ESTs as nvSPARC1-4. EST data analysis indicated that the EST for nvSPARC3 was incomplete at the 5′end. Hence, 5′ RACE amplification was performed. DNA sequence analysis of the 5′ extension revealed a start codon followed by a 20 amino acid sequence at the N-terminus with the classical features of a signal peptide and with sequence similarity to the signal peptides of cnidarian SPARC paralogs and non-cnidarian SPARC orthologs. In contrast to the three other nvSPARC proteins with no domain I, nvSPARC3 has a short 17 amino acid sequence overlapping with domain I of higher Metazoa. Gene architecture analysis using the JGI *N. vectensis* v1.0 genome browser confirmed that domain I is not encoded by the *N. vectensis* genome (Electronic supplementary materials, S1).

Several other properties within the four distinct nvSPARC EST entries indicate that these are bona fide SPARC orthologs and not representatives of other SPARC

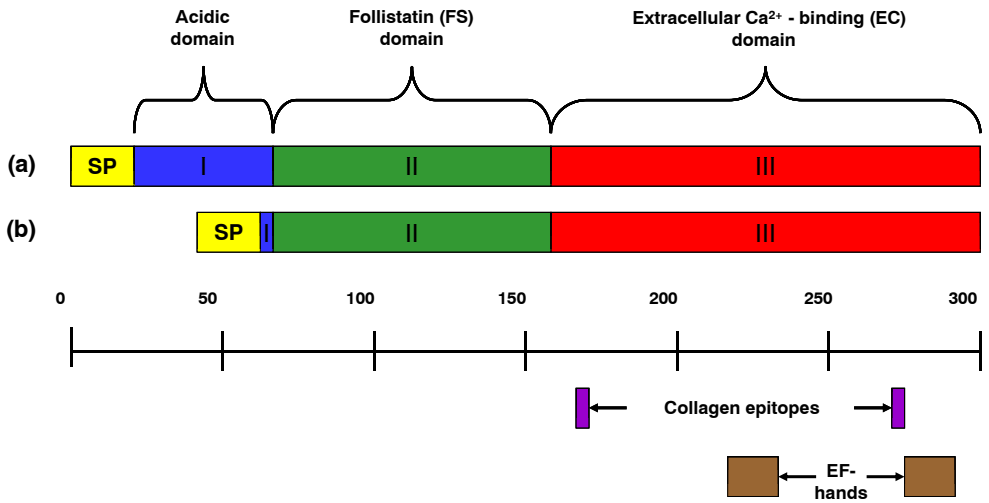

Fig. 1 Modular organization of SPARC. Schematic of the trimodular organization of SPARC in triploblastic organisms (**a**) and bimodular organization in diploblastic cnidarians (**b**). A signal peptide is present at the N-termini of SPARC orthologs in all animal phyla. The acidic N-terminal domain varies in length, ranging from 52 to 112 amino acids for *H. sapiens* and *C. intestinalis* SPARC, respectively. Domain I is absent from three of the four *N. vectensis* SPARC orthologs. The exception is nvSPARC3, which has a 17 amino acid-long domain I. The FS–EC domain pair is conserved in all animal phyla. The positions of the evolutionarily conserved collagen-binding epitopes and EF-hands within the EC domain are shown as *boxes*. For simplicity, the *scale bar* is representative of human SPARC

family members. Previous studies have demonstrated that the FS–EC domain pair of *N. vectensis* SMOC1 and SMOC2 are separated by two thyroglobulin modules and a SMOC module (Novinec et al. 2006). The C-terminal domain of *N. vectensis* Testican has a thyroglobulin module, a conserved feature of the Testican subfamily. While SPARC-like (SC1/Hevin) has a trimodular organization similar to SPARC, domain I of SPARC-like is approximately 12 times longer than domain I of SPARC (Johnston et al. 1990; Vannahme et al. 1999, 2002, 2003). Moreover, a recent comprehensive EST and genomic bioinformatic analysis of the evolution of glycoproteins associated with vertebrate mineralized tissues demonstrated that SPARC like evolved from SPARC and was restricted to the genomes of amniotes (Kawasaki et al. 2004, 2007; Kawasaki and Weiss 2006).

Sequence conservation and divergence of SPARC during evolution

Alignment of SPARC sequences derived from representative invertebrate and vertebrate taxa revealed that the Ca^{2+}-binding acidic N-terminal (domain I) is absent from the nvSPARC1-4 (Fig. 2). Consequently, nvSPARCs are on average approximately 10-fold less acidic (p*I* values ranging from 5.1 to 6.1). Despite its relatively long acidic domain I (112 amino acids), *Ciona intestinalis* SPARC has a p*I* value of 4.0, which is comparable to other SPARC orthologs. Hence, domain I of *C. intestinalis* SPARC is expected to have an affinity for several Ca^{2+} ions at neutral pH.

In all animal phyla, the FS domain (human SPARC, amino acids 70–154; nvSPARC1, amino acids 26–85) contains 11 cysteine residues, seven of which are aligned (Martinek et al. 2002, 2007). The non-globular organization of domain II has been shown in human SPARC to be in part stabilized by a small hydrophobic core and five disulfide bridges organized in two non-overlapping sets: bonds between N-terminal cysteines 1–3 and 2–4 form a follistatin subdomain, whereas bonds between cysteines 5–9, 6–8, and 7–10 give rise to a globular subdomain similar to Kazal proteinase inhibitor. The conservation of cysteine residues and distribution pattern may reflect that the disulfide-bridge linkage pattern is also conserved across phyla. A potential N-glycosylation acceptor site (NXT/S) is found in all SPARC orthologs, except for nvSPARC3, which has two potential acceptor sites, while nvSPARC2 has none. Moreover, the positions of potential N-glycosylation sites in nvSPARC are more variable relative to non-cnidarian SPARCs.

As with non-cnidarian SPARCs, domain III (extracellular calcium (EC) domain) of nvSPARC1-4 (e.g., nvSPARC1, amino acids 80–211) is the most evolutionarily conserved. Identity search by BLASTp indicated that within this region, nvSPARC1 has 26% sequence identity (44% similarity) with sea urchin SPARC and 28% sequence identity (44% positive) with human SPARC. The *E* values generated using nvSPARC2-4 as queries were comparable to nvSPARC1. Sequence identity is highest within the last four of the six α-helices of the EC domain. These four α-helices form a pair of EF-hands: EF–hand1 (amino acids 227–251 for human SPARC, amino acids 171–199 for nvSPARC1) and EF-hand2 (amino acids 272–291 for human SPARC; amino acids 171–199 for nvSPARC1). For simplicity, only the amino acid positions of nvSPARC1 are listed.

The binding of a Ca^{2+} ion with high affinity to EF-hand1 (K_D=470 nm) and EF-hand2 (K_D=57 nM) for human SPARC is dependent on cooperative interactions between the two EF-hands. The high sequence conservation of both EF-hands in all animal phyla may reflect evolutionary pressure to maintain the positive cooperativity for the proper folding of SPARC. Mutagenesis studies with human SPARC have demonstrated that the substitution of Asp^{251} for a Lys residue within the αE helix of EF-hand1 dramatically decreases its affinity for Ca^{2+} (Sasaki et al. 1998). An Asp residue is conserved at this position in all animal phyla, except for nvSPARC1 where a Glu residue is present. However, gaps exist in the alignment within the EF1 domain as a result of the algorithm being set to maximize sequence alignments. Hence, X-ray crystallography studies are required to determine whether if a bona fide EF-hand1 exists in cnidarians.

Two amino acids, Asp^{257} and Glu^{268} in human SPARC, required for the high-affinity binding of a Ca^{2+} ion by EF-hand2 are conserved across all animal phyla. EF-hand2 of human SPARC is stabilized by a disulfide bridge between Cys^{274} and Cys^{289} residues (Pottgiesser et al. 1994; Sasaki et al. 1998). These two signature Cys residues are conserved in all animal phyla except in *C. intestinalis* SPARC where Ala residues are found at these positions. Alanine residues are also present in the closely related *Ciona savignyi* SPARC ortholog (data not shown), making it unlikely that sequencing errors account for the Cys to Ala substitutions. Furthermore, the highly conserved amino acid residue upstream of the Trp in helix A of the first collagen epitope is substituted for a Ser residue. Hence, for sea squirt SPARC, amino acid substitutions are found at two key sites.

In addition to the conservation of the EF-hands, the EC domain contains two low-affinity binding sites for several fibrillar collagens (K_D~0.4–1.4 nM) and network-forming collagen IV (K_D~1.1 nM) (Sasaki et al. 1997, 1998). Site-directed mutagenesis of a constitutively active form of the human EC domain has demonstrated that amino acids within two loci are critical for collagen binding: amino acids Arg^{169} and Asn^{177} in helix αA and Leu^{259}, Met^{262}, and Glu^{263} in the loop-region linking EF-hand1 and EF-

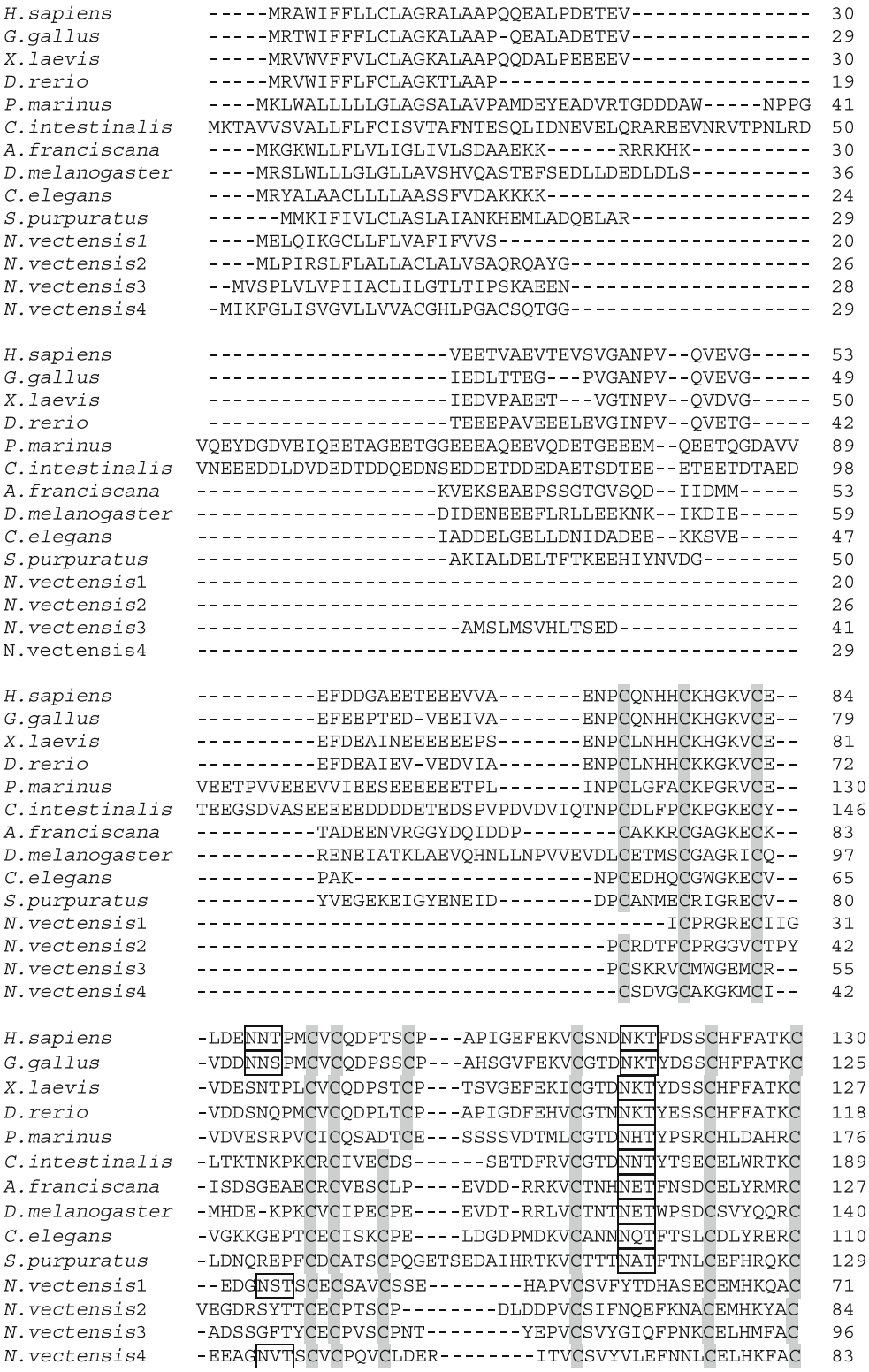

Fig. 2 CLUSTALW alignment of metazoan SPARC proteins. Amino acid sequences from representative triploblastic organisms and the four *N. vectensis* orthologs aligned with ClustalW. Cysteine residues are *shaded*. Potential N-glycosylation sites are *boxed*. The two collagen-binding epitopes within the EC domain are *underlined*. EF-hand1 and EF-hand2 are in *bold*

hand2 (Fig. 3) (Sasaki et al. 1998). As with the EF-hands, high sequence conservation is observed at these two collagen-binding sites across all animal phyla. Two critical amino acids, Arg and Glu are 100% conserved in all animal phyla examined, whereas conservative amino acid substitutions are found at the critical Leu residue. The fourth critical amino acid, a Gln residue in helix A, shows the most divergence.

Fig. 2 (continued)

```
H.sapiens          TLEGTKKG-------HKLHLDYIGPCKYIPPCL-DSELTEFPLRMRDWLK  172
G.gallus           TLEGTKKG-------HKLHLDYIGPCKFIPACL-DTELTEFPLRMRDWLK  167
X.laevis           TLEGTKKG-------HKLHLDYIGPCKYIAPCL-DNELSEFPLRMRDWLK  169
D.rerio            SLEGTKKG-------HKLHMDYIGPCKYIAPCL-ENELNEFPLRMRDWLK  160
P.marinus          ALDGTKKG-------RHLHLDYIGPCKEITPCL-DVELTEFPLRMRDWLK  218
C.intestinalis     IMKQNKAKGV-----QHLRLDYYGDCKEIQPCG-EHELSEYPTRMRSWIK  233
A.franciscana      LCTTGSQECLG-PKYSHAHIEYYGECRDMPECS-EQEMDDFPRRMRDWLF  175
D.melanogaster     WCDSGEPGCTN-PDNAHMHIDYYGACHEPRSCE-GEDLKDFPRRMRDWLF  188
C.elegans          LCKRKSKECSK-AFNAKVHLEYLGECKKLDECT-EEHMAQFPERMADWLF  158
S.purpuratus       MEVD----------LMEVHVDYYGECAEMGSCS-AEDLREYPERMTNWFI  168
N.vectensis1       DFGFFTAVKHKGKCTKKG------SNIIGRQCP-VERLLDFHSRYLEWLL  114
N.vectensis2       KLQMSMAIKNAGACVKPT----DPKMMGMTDCP-KWMLEQFPYRFLDWLY  129
N.vectensis3       IEGVNIGVKNKGPCVNKTGGSDPEPDARTLVCPGIDHMSQFRDRYLEWAH  146
N.vectensis4       RYEIPIGVASMGPC--DVQALRQSSGQNVSECA-RDNLLQFGERYLEWIL  130

H.sapiens          NVLVTLYERDEDNN-LLTEKQKLRVKKIHENEKRLE-AGDHPVELLARDF  220
G.gallus           NVLITLYERDEDNN-LLTEKQKLKVKNIHENEKRLE-AGDHTVELLARDF  215
X.laevis           NVLVSLYERDENNN-LLNEKQKLRVKKIHENEKRLE-AGDHTVELLVRDF  217
D.rerio            NVLVTLYERDEDNN-LLTEKQKLRVKKIYENEKRLQ-AGDHSLDLLALDF  208
P.marinus          NVVVQMYERDEERGELLTDKQRRKLRKIVSDERRLR-QGDHTRELLQRDF  267
C.intestinalis     NIYLQMYDEAEDMG-GLNEKQRFHGRKLVANRDRLQTYEEHHIDLMSREF  282
A.franciscana      NIMRDLAARHELS-------------------------PHYLKLEKEAE  199
D.melanogaster     NVMRDLAERDELT-------------------------EHYMQMELEAE  212
C.elegans          QVMKELKKRRELHK-----------------------LEWEELLSEAE  183
S.purpuratus       KSLALIRNRPEEHG-----------------------GLSQKEKETL  192
N.vectensis1       VAYKESKHAVMDSR-------------------------------VALE  132
N.vectensis2       LTRELSRD-------------------------------PRFEIAKRGE  147
N.vectensis3       VAKEQTKDQNYTLT-------------------------------RRIH  164
N.vectensis4       ISRRRLVNRYYQVK-----------------------------ESEKGH  150

H.sapiens          EKNYNM**YIFPVHWQFGQLDQHPIDGYLSHTE**LAPLRAPLIPMEHCTTRFF  270
G.gallus           EKNYNMYIFPVHWQFGQLDQHPIDGYLSHTELAPLRAPLIPMEHCTTRFF  265
X.laevis           EKNYNMYIFPVHWQFGQLDQHPIDGYLSHTELSPLRAPLIPMEHCTTRFF  267
D.rerio            EKNYNMYIFPVHWQFGQLDQHPIDGFLSHTELAPLRAPLIPMEHCTTSFF  258
P.marinus          VKNYRMYIYPVHWQFAQLDSRPADRYLSHSELSPLRAPLVPMEHCTTRFF  317
C.intestinalis     QKFYPLYKYPIHWKFGILDVNPTDKYLSKRELEPMRAPLVPLEHCTDTFF  332
A.franciscana      VEQSKRWANAAIWKFCDLDGHPHDRKVSRHELFPIKAPLMALEHCISPFL  249
D.melanogaster     TNNSRRWSNAAVWKWCDLDG-DTDRSVSRHELFPIRAPLVSLEHCIAPFL  261
C.elegans          NDDEKKHVYPVIWKFCELDTKPHDKSVSHHELIPITAPVIPMESCIKPFL  233
S.purpuratus       QQEWSQSTHAIFWKFDRLDKNPSDHYLDFAELEGLRAPIVLFEPCTKPFL  242
N.vectensis1       GILDEEIKRIIMWEFDNQDANRNDQLDEHEVKS----MIDLREPCMVGFM  178
N.vectensis2       QLTGDEIKEVLAWEFNNWDLNKNGIWDK-EELDLVQTILMPIEGCSFGFL  196
N.vectensis3       DMTMRQKVAIAKWEFDRNDLDKDNALEGAEIDN--ILAMMIYEPCAAGFL  212
N.vectensis4       NLPPDMRKELAQWEFDRVDFNRDGVLSG-REINSIIGPLLFHERCIYGFV  199

H.sapiens          E**TCDLDNDKYIALDEWAGCFG**IKQKDIDKDLVI----------------  303
G.gallus           EACDLDFDKYIALEEWASCFGIKEQDIDKDLVI----------------  298
X.laevis           DECDIDDDKYIALEEWAKCFGIKEQDVDKDMIV----------------  300
D.rerio            EQCDADQDKYIALEEWANCFGIKEQDVDKDLVI----------------  291
P.marinus          LRCDADGDRLISLPEWGACFGLLPDDVDEALLF----------------  350
C.intestinalis     STADANKDHLISLYEWAEALGLQEDDIDSTLTTV----------------  366
A.franciscana      NKCDVDDDHFITLREWGKCLEIPEDELEDKCEDVRGVAFNEI--------  291
D.melanogaster     ESCDSNKDHRITLVEWGACLELDPEDLKERCDDVQRAQPHLLG-------  304
C.elegans          EGCDANNDGNISIKEWGKCLGLKEGEIQERC--------------------  264
S.purpuratus       QACDVDNDTLISAVEFGRCLNLSDEQVPY--------------------- 271
N.vectensis1       DACDFDGHPGITRHEWNACFPTRLEGQKEGMVV----------------  211
N.vectensis2       KSCDLDGKGDMDRREWMMCFPEGVIDISTGSRRH----------------  230
N.vectensis3       WSCDLNRKQGISRSEWDMCFTLAGNWSREGINNQAKPSKAEFSHPWYNFF  262
N.vectensis4       MSCDVNKNSVIDKQEWLLCFPAFARSSETTEWRRQY-------------  235
```

Phylogenetic analysis of the FS–EC domain of SPARC

Since domain I is absent from cnidarian SPARC1-4, an amino acid alignment of the FS–EC domains of SPARC sequences was used to construct a phylogenetic tree using likelihood and Bayesian methods. The FS–EC domains of Testican sequences were also included as an outgroup. Within the SPARC gene family, the phylogeny is consistent with accepted taxonomic relationships, with fairly well-supported clades representing Cnidaria, Protostomia, and

		Helix A		Link
Triplobast organisms				
H. sapiens	169-	RmrdwlkN-177	259-	LipmE-263
G. gallus		RmrdwlkN		LipmE
X. laevis		RigdwlkN		LipmE
D. rerio		RmrdwlkN		LipmE
D. melanogaster		RmrdwlfN		LvslE
C. elegans		Rmadwlfq		VipmE
S. purpuratus		Rmtnwfik		ivlfE
C. intestinalsis		RmrswikN		LvplE
Diploblast organisms				
H. magnipapillata		Rfjewmvh		LmsmE
N. vectensis1		Rylewlla		idlrE
N. vectensis2		Rfldwylt		LmpiE
N. vectensis3		Rylewahv		mmiyE
N. vectensis4		RylewilL		LlfhE
		* *		: *

Fig. 3 Evolutionary conservation of amino acids critical for the binding of SPARC to the triple helices of fibrillar collagens and network-forming collagen. For simplicity, the amino acid positions of the collagen-binding epitopes are shown only for human SPARC. *Shaded letters* highlight the five amino acids in the EC domain found to be critical for the binding of SPARC to the triple helices of several fibrillar collagens and network-forming collagen IV. *Asterisks* fully conserved residues, *colons* conservation of strong groups

Deuterostomia (Fig. 4). Even within each of these groups, the topology is largely congruent with organismal relationships. For example, within deuterostomes, vertebrates form a well-supported monophyletic group sister to the tunicate *C. intestinalis*, with the echinoderm *Strongylocentrotus purpuratus* as the most basally branching deuterostome. The one exception is within the protostomes, the failure to recover the lophotrochozoan (which should consist of the annelid *Capitella* sp. and mollusk *Lottia gigantea*) and ecdysozoan clades; however the lack of statistical support for these basal divergences within protostomes may reflect the poor species sampling within this highly diverse group. The monophyletic grouping of the *N. vectensis* SPARC sequences separately from the hydra *Hydra magnipapillata* suggests that the four *N. vectensis* copies are the result of two rounds of gene duplication in the anthozoan lineage. From an evolutionary standpoint, the lack of domain I in the cnidarian SPARC sequences is intriguing. This domain, while not found in Testicans and SMOCs, is present in SPARC-like (SC1/Hevin). This would suggest that domain

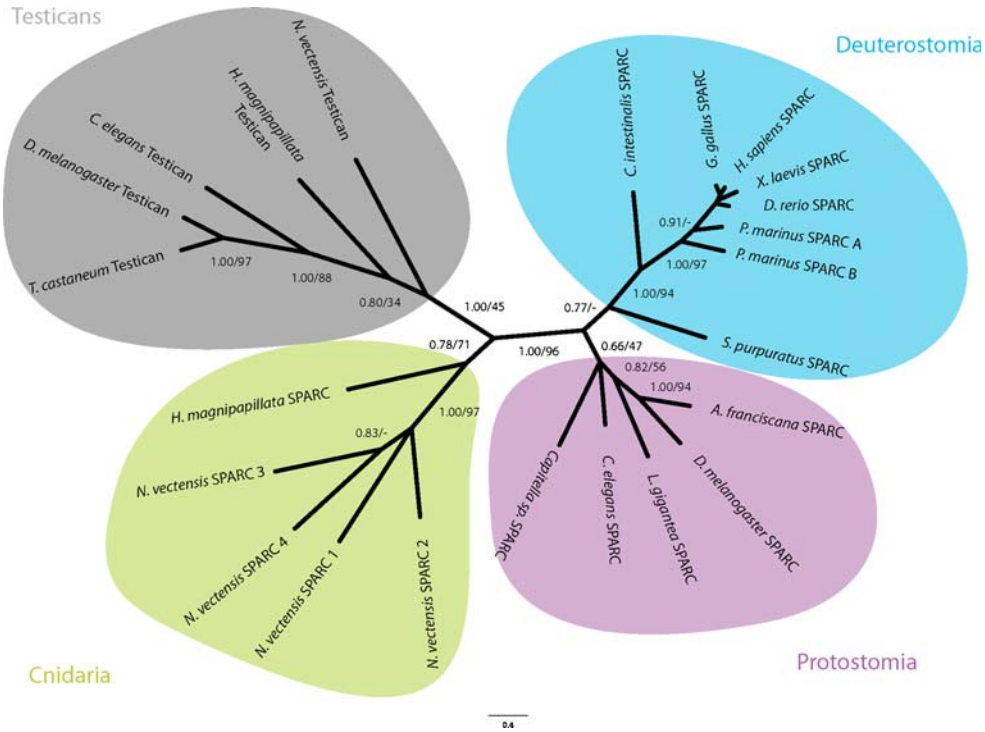

Fig. 4 Bayesian phylogeny of metazoan SPARC FS–EC domains, with Testican FS–EC domains included as outgroup sequences. *Numbers above each node* represent Bayesian posterior probabilities, followed by likelihood bootstrap percentages (100 replicates), with nodes having <50% bootstrap support (not represented in the bootstrap consensus tree) *marked with a dash*

I might have been a later addition in SPARC evolution in metazoans, after the emergence of bilaterians. Alternatively, domain I may have arisen early in metazoans and then been subsequently lost in the cnidarian lineage, although this would be a less parsimonious explanation. Sequences from more basal metazoans such as sponges, or even placozoans, would be necessary in clarifying this issue. Preliminary inspections of the sponge and placozoan genomes did not reveal any SPARC sequences, although resolution of this issue may need to await better annotation of these genomes.

Expression of SPARC in the endoderm of *N. vectensis* early developmental stages

In contrast to triploblastic animals that form three germs layers during gastrulation, diploblastic Cnidaria such as *N. vectensis* gastrulate without the formation of a mesoderm. The resultant bi-layered adult organisms are composed of an outer ectoderm and inner endoderm separated by a collagen and proteoglycan-rich mesoglea. Whole mount *in situ* hybridization demonstrates that nvSPARC1 and 2 transcripts are restricted to the endoderm at the early planula, metamorphosing planula and polyp stages (Fig. 5). Transcripts were first detected after gastrulation (data not shown).

In planula embryos, slightly higher expression of nvSPARC1 was observed in the endoderm adjacent to the oral and aboral regions, with a mainly uniform distribution throughout the remainder of the endoderm. In metamorphosing planula, expression was detectable in the endoderm of tentacle buds. By polyp stage, a higher level of expression was evident within the tentacle buds and the aboral endoderm lining of the gastric cavity. In comparison, nvSPARC2 expression at the early polyp stage is restricted to the pharyngeal endoderm and tentacle buds. The expression profile of nvSPARC3 was similar to nvSPARC2 (data not shown).

Discussion

Bioinformatic analyses of completed and annotated genomes have demonstrated that the trimodular organization of SPARC, first described in mammals (Termine et al. 1981), is conserved in triploblastic organisms. In order to provide new insights into the molecular evolution and function of SPARC in more ancestral organisms, we data-mined a draft version of the completed and partially annotated genome of the basal cnidarian *N. vectensis*. We report that the *N. vectensis* genome codes for novel SPARC genes, composed of a signal peptide followed by the signature FS–EC domain pair. We hypothesize that an ancestral gene lacking exon(s) coding for the N-terminal acidic domain I module underwent two cnidarian-specific duplication events. Expression of the nvSPARCs is restricted to the endoderm during post-gastrulation development. Moreover, the evolutionary conservation of the two collagen-binding epitopes found in the EC module raises the possibility that a functional relationship between SPARC and collagen first arose in diploblastic organisms and that this relationship has been maintained throughout the Metazoa. Database EST and BLASTp searches indicate that two other members of the SPARC family, SMOC and Testican, are encoded by the genome of *N. vectensis*.

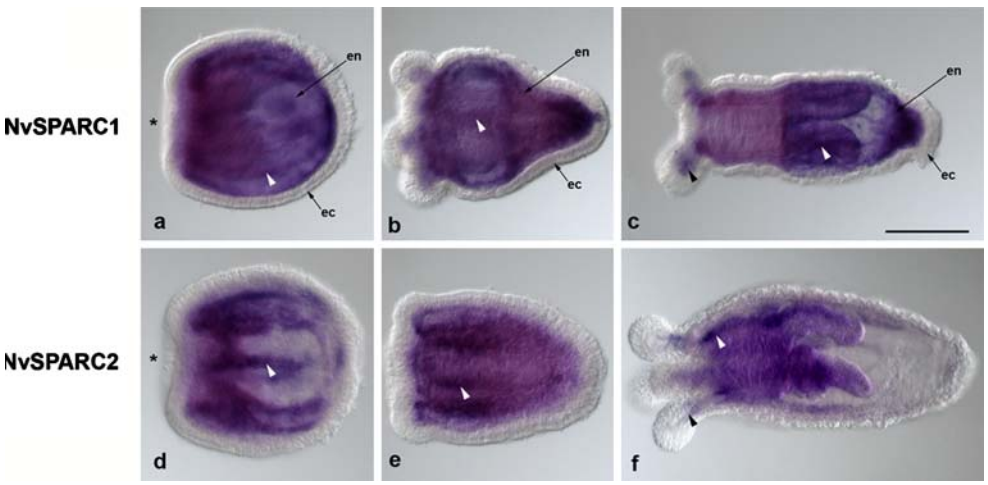

Fig. 5 Expression of SPARC is restricted to the developing endoderm during post-gastrula development in *N. vectensis*. SPARC transcripts were not detectable until post-gastrula development for nvSPARC1 and 2. Expression is restricted to the endoderm (*en*) in early planula (**a**, **d**) and remains restricted to the endoderm (*en*) during metamorphosing planula (**b**, **e**) and primary polyp stages (**c**, **f**). The ectoderm (*ec*) is devoid of staining in all stages. Higher level staining is observed within developing mesenteries (*white arrowheads*). In primary polyps, staining is also observed in the tenticular endoderm (*arrowheads*). All specimens are oriented with the oral end (*asterisks*) to the left

Hence, with the exception of SPARC-like (SC1/Hevin), the SPARC family appeared early during metazoan evolution.

The enrichment of SPARC within mineralized tissues in vertebrates has been hypothesized to be due to a combination of its affinity for fibril-forming collagens and high Ca^{2+}-binding potential of the acidic N-terminal domain I (Giudici et al. 2008; Hohenester et al. 2008). An observation on SPARC Ca^{2+}-phosphate crystal growth in vitro raises the possibility that it may serve to inhibit tissue mineralization in vivo (Hunter et al. 1994). However, in osteoblast tissue culture models, a correlation was not observed between the activation of SPARC expression and the onset of mineralization. This is consistent with the zygotic expression of SPARC in mice and amphibians several weeks after the appearance of mineralized tissues (Damjanovski et al. 1994; Kawasaki et al. 2004; Mason et al. 1986). If domain I has no role in promoting tissue mineralization, what is its function in triploblastic organisms? Prominent among the several activities of mammalian SPARC demonstrated in vitro is its ability to inhibit the spreading of newly plated cells and to promote the rounding of endothelial cells and fibroblasts by decreasing the number of focal adhesions. This effect of native SPARC is mimicked by synthetic peptides to different domains of murine SPARC, including peptide 1.1 (QTEVAEEIVE EETVVEETGV; amino acids 5–23 of human SPARC), which corresponds to the N-terminal sequence of domain I (Lane and Sage 1990; Nie et al. 2008). While it is tempting to suggest that domain I evolved to decrease cell–matrix adhesion, synthetic peptides overlapping all three domains also have counteradhesive effects in vitro (Lane and Sage 1990). Hence, counteradhesive effects cannot be excluded for cnidarian SPARCs. In contrast, inhibition of cell cycle progression has only been demonstrated with synthetic peptides corresponding to the FS and EC domains, raising the possibility that an ancestral function of SPARC, conserved during metazoan evolution, is to induce the withdrawal of cells from the cell cycle and to promote their differentiation by modulating the activity of growth factors. The binding of SPARC to several growth factors, including PDGF and VEGF, supports this hypothesis (Bradshaw and Sage 2001).

The supramolecular organization, biomechanical properties and functions of collagens are dependent on homo- and heterotypic associations between collagen molecules as well as interactions with transmembrane receptors and non-collagenous matrix molecules. While the precise functional relationship between SPARC and collagens is poorly understood, collagen fibrils are smaller and more uniform in size in the dermis of *Sparc*-null mice relative to their wild-type counterpart. Consequently the skin of *Sparc*-null mice lacks tensile strength (Bradshaw 2009). *Sparc*-null mice develop cataracts, due in part to collagen IV instability in the lens capsule (Norose et al. 1998). While SPARC immunostaining is not readily observed within basal laminae of vertebrates (presumably due to epitope masking), intense SPARC immunostaining is observed in embryonic basal laminae of nematodes and insects. In *Sparc*-mutant *D. melanogaster* embryos, collagen IV immunostaining is not observed, resulting in discontinuous basal laminae in late stage embryos.

The affinity of SPARC for collagen helices is only partly reduced by the absence of domain I (SPARC αI), indicating that conformational changes associated with the binding of Ca^{2+} ions by the EC domain EF-hands also plays a role in the Ca^{2+}-dependent binding to collagen helices. The deletion of a steric constraint (helix βC, amino acids 196–203 human SPARC; FS–EC αβC) in the EC domain results in a 10-fold increase in the binding of recombinant SPARC (rSPARC, also referred to as "activated" SPARC) to collagen helices. Site-directed mutagenesis studies conducted with rSPARC revealed that five residues within two epitopes were critical for collagen binding (Sasaki et al. 1998). Sequence alignment revealed a high conservation of these critical amino acids from cnidarians to mammals. Previous studies demonstrated that the EC domain of *C. elegans* SPARC displays a comparable affinity for collagens as mammalian SPARCs (Sasaki et al. 1998), indicating that minor substitutions of critical amino acids have minimal impact on collagen binding. Although it remains to be determined experimentally, the high amino acid conservation at these two critical loci suggests that the interaction of SPARC and collagens is evolutionarily conserved throughout the Metazoa. Even though SPARC binds to the entire triple helical domain of collagen I, several preferential binding sites exist, including an α1β1 integrin epitope (Giudici et al. 2008). This, combined with the observation that SPARC inhibits collagen I fibrillogenesis *in vitro* when added prior to the lag phase, but not when added in the middle of the growth phase, has led to the suggestion that SPARC may play a critical role in regulating early stages of collagen fibril assembly by modulating integrin–collagen interactions.

Our laboratory and others have demonstrated that a mimetic peptide corresponding to the unique C-terminal disulfide bridge of EF-hand2 of SPARC (peptide 4.2) has potent anti-proliferative and counteradhesive activities, both *in vitro* and *in vivo* (Damjanovski et al. 1997; Lane and Sage 1990). While other mimetic peptides corresponding to other domains of SPARC have biological activity in vitro, to date only peptide 4.2 has been shown to mimic these effects in vivo. Disruption of the disulfide bridge in peptide 4.2, by substituting the Cys for an Ala residue, abrogates in vitro and in vivo biological activities, indicating a critical requirement of this disulfide bridge. It is interesting to note that like native SPARC, disulfide-bridged peptide 4.2 binds to VEGF, inhibiting VEGF-stimulated microvascular-

derived endothelial cell proliferation. In addition, peptide 4.2 also binds collagens (Lane and Sage 1990). Hence, the striking sequence conservation of EF-hand2 suggests that the C-terminal domain endows SPARC with activities that are conserved in Metazoa. Surprisingly, both the Cys residues involved in the formation of the disulfide bridge are absent in urochordate SPARCs. It is possible that the loss of the highly conserved Cys residues reflects a unique activity of SPARC in promoting the metamorphosis of ascidians from non-feeding motile tadpoles into filter-feeding, fixed juveniles.

Zygotic expression of SPARC is first detected in the mesoderm of *Xenopus laevis* and *D. melanogaster* embryos. We report in this study that nvSPARC1-3 expression is restricted to the endoderm from planula to end-metamorphosis in *N. vectensis*. Several evolutionary scenarios have been proposed for the evolution of the mesoderm in triploblastic organisms, including the hypothesis that the mesoderm is derived from endoderm. Consistent with this hypothesis is that genes involved in the early specification and differentiation of the mesoderm in triploblastic organisms are expressed by the endoderm in diploblastic organisms, e.g., transcription factors such as Twist, Snail, and Brachyury (Martindale 2005; Martindale et al. 2004). Thus, the early expression of SPARC in the mesoderm of triploblastic organisms and the endoderm in diploblastic organisms is consistent with a proposed evolutionary origin of the mesoderm. The timing of endoderm and mesoderm expression makes it unlikely that SPARC functions as a germ layer inducer. It most likely affects cellular differentiation by promoting the withdrawal of cells from the cell cycle and contributes to modulate cell–cell and cell–matrix adhesion.

In vertebrates, separate basal laminae and interstitial matrices are clearly visible by electron microscopy, in contrast to cnidarians where the basal laminae and interstitial matrices are not as distinct. Instead, the outer ectoderm and inner endoderm (gastrodermis) are separated by at thick, sheet-like transparent composite matrix (mesoglea). Ultrastructural analyses revealed that the mesoglea is composed of a basal lamina-like matrix juxtaposed to the epithelial sheets of the two germ layers, forming a continuum with an intermediate mucopolysaccharide-rich matrix, reinforced with radial elastic fibrillin-rich fibers, and collagen microfibers (Davis and Haynes 1968; Schmid et al. 1999). Based on the conservation of the collagen-binding epitopes, it is conceivable that SPARC plays a role in promoting collagen assembly and maturation in the mesoglial matrix. The presence of four SPARC orthologs in a relatively primitive diploblastic organism is intriguing, considering that in more advanced triploblastic organisms SPARC is encoded by a single gene, with the exception of *Petromyzon marinus*. However, the presence of four orthologs is consistent with emerging evidence indicating a surprising genetic complexity of ancestral organisms.

Our data raise the possibility that SPARC first evolved as a protein composed of an FS–EC domain pair, whose ancestral gene underwent two rounds of duplication. High-resolution ultrastructural data indicate that the acidic module of mammalian SPARC extends from the globular FS–EC domain pairs (Maurer et al. 1992), potentially facilitating access and interaction with other ECM components and therefore adding to the functional complexity of SPARC in more advanced organisms. It remains to be determined if SPARC is encoded by the genome of more primitive animals and, if so, whether domain I is absent. Like collagens, it appears that SPARC and two other family members (SMOCs and Testicans) represent a subset of the ECM components associated with higher organisms that have important functions throughout the Metazoa.

Acknowledgments This work was supported by Discovery grants to B.C., M.R., S.D., and U.T. from National Sciences and Research Council of Canada and an Early Researcher Award to B.C.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bassuk JA et al (2000) Induction of TGF-beta1 by the matricellular protein SPARC in a rat model of glomerulonephritis. Kidney Int 57:117–128

Bradshaw AD (2009) The role of SPARC in extracellular matrix assembly. J Cell Commun Signal (in press) doi:10.1007/s12079/009-0062-6

Bradshaw AD, Sage EH (2001) SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury. J Clin Invest 107:1049–1054

Brown TJ, Shaw PA, Karp X, Huynh MH, Begley H, Ringuette MJ (1999) Activation of SPARC expression in reactive stroma associated with human epithelial ovarian cancer. Gynecol Oncol 75:25–33

Damjanovski S, Malaval L, Ringuette MJ (1994) Transient expression of SPARC in the dorsal axis of early Xenopus embryos: correlation with calcium-dependent adhesion and electrical coupling. Int J Dev Biol 38:439–446

Damjanovski S, Karp X, Funk S, Sage EH, Ringuette MJ (1997) Ectopic expression of SPARC in Xenopus embryos interferes with tissue morphogenesis: identification of a bioactive sequence in the C-terminal EF hand. J Histochem Cytochem 45:643–655

Damjanovski S, Huynh MH, Motamed K, Sage EH, Ringuette M (1998) Regulation of SPARC expression during early Xenopus development: evolutionary divergence and conservation of DNA regulatory elements between amphibians and mammals. Dev Genes Evol 207:453–461

Darling JA et al (2005) Rising starlet: the starlet sea anemone, Nematostella vectensis. Bioessays 27:211–221

Davis LE, Haynes JF (1968) An ultrastructural examination of the mesoglea of Hydra. Z Zellforsch Mikrosk Anat 92:149–158

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214

Felsenstein J (1985) Confidence limits on phlyogenies: an approach using the bootstrap. Evolution 39:783–791

Fitzgerald MC, Schwarzbauer JE (1998) Importance of the basement membrane protein SPARC for viability and fertility in Caenorhabditis elegans. Curr Biol 8:1285–1288

Francki A, Bradshaw AD, Bassuk JA, Howe CC, Couser WG, Sage EH (1999) SPARC regulates the expression of collagen type I and transforming growth factor-beta1 in mesangial cells. J Biol Chem 274:32145–32152

Gelman A, Rubin DB (1992) Inference from interative simulation using multiple sequences. Statistical Science 7:457–511

Giudici C et al (2008) Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. J Biol Chem 283:19551–19560

Greiling TM, Stone B, Clark JI (2009) Absence of SPARC leads to impaired lens circulation. Exp Eye Res 89:416–425

Guindon S, Gascuel O (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Guindon S, Letheic F, Duroux P, Gascuel O (2005) PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:W557–559

Hohenester E, Maurer P, Timpl R (1997) Crystal structure of a pair of follistatin-like and EF-hand calcium-binding domains in BM-40. EMBO J 16:3778–3786

Hohenester E, Sasaki T, Giudici C, Farndale RW, Bachinger HP (2008) Structural basis of sequence-specific collagen recognition by SPARC. Proc Natl Acad Sci USA 105:18273–18277

Hunter GK, Kyle CL, Goldberg HA (1994) Modulation of crystal formation by bone phosphoproteins: structural specificity of the osteopontin-mediated inhibition of hydroxyapatite formation. Biochem J 300(Pt 3):723–728

Huynh MH, Sage EH, Ringuette M (1999) A calcium-binding motif in SPARC/osteonectin inhibits chordomesoderm cell migration during Xenopus laevis gastrulation: evidence of counter-adhesive activity in vivo. Dev Growth Differ 41:407–418

Johnston IG, Paladino T, Gurd JW, Brown IR (1990) Molecular cloning of SC1: a putative brain extracellular matrix glycoprotein showing partial similarity to osteonectin/BM40/SPARC. Neuron 4:165–176

Kawasaki K, Weiss KM (2006) Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. J Exp Zool B Mol Dev Evol 306:295–316

Kawasaki K, Suzuki T, Weiss KM (2004) Genetic basis for the evolution of vertebrate mineralized tissue. Proc Natl Acad Sci USA 101:11356–11361

Kawasaki K, Buchanan AV, Weiss KM (2007) Gene duplication and the evolution of vertebrate skeletal mineralization. Cells Tissues Organs 186:7–24

Lane TF, Sage EH (1990) Functional mapping of SPARC: peptides from two distinct Ca+(+)-binding sites modulate cell shape. J Cell Biol 111:3065–3076

Lane TF, Sage EH (1994) The biology of SPARC, a protein that modulates cell–matrix interactions. Faseb J 8:163–173

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol 25:1307–1320

Martindale MQ (2005) The evolution of metazoan axial properties. Nat Rev Genet 6:917–927

Martindale MQ, Pang K, Finnerty JR (2004) Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). Development 131:2463–2474

Martinek N, Zou R, Berg M, Sodek J, Ringuette M (2002) Evolutionary conservation and association of SPARC with the basal lamina in Drosophila. Dev Genes Evol 212:124–133

Martinek N, Shahab J, Sodek J, Ringuette M (2007) Is SPARC an evolutionarily conserved collagen chaperone? J Dent Res 86:296–305

Martinek N, Shahab J, Saathoff M, Ringuette M (2008) Haemocyte-derived SPARC is required for collagen-IV-dependent stability of basal laminae in Drosophila embryos. J Cell Sci 121:1671–1680

Mason IJ, Murphy D, Munke M, Francke U, Elliott RW, Hogan BL (1986) Developmental and transformation-sensitive expression of the Sparc gene on mouse chromosome 11. EMBO J 5:1831–1837

Maurer P et al (1992) High-affinity and low-affinity calcium binding and stability of the multidomain extracellular 40-kDa basement membrane glycoprotein (BM-40/SPARC/osteonectin). Eur J Biochem 205:233–240

Maurer P, Hohenadl C, Hohenester E, Gohring W, Timpl R, Engel J (1995) The C-terminal portion of BM-40 (SPARC/osteonectin) is an autonomously folding and crystallisable domain that binds calcium and collagen IV. J Mol Biol 253:347–357

Midwood KS, Williams LV, Schwarzbauer JE (2004) Tissue repair and the dynamics of the extracellular matrix. Int J Biochem Cell Biol 36:1031–1037

Motamed K, Funk SE, Koyama H, Ross R, Raines EW, Sage EH (2002) Inhibition of PDGF-stimulated and matrix-mediated proliferation of human vascular smooth muscle cells by SPARC is independent of changes in cell shape or cyclin-dependent kinase inhibitors. J Cell Biochem 84:759–771

Motamed K et al (2003) Fibroblast growth factor receptor-1 mediates the inhibition of endothelial cell proliferation and the promotion of skeletal myoblast differentiation by SPARC: a role for protein kinase A. J Cell Biochem 90:408–423

Nie J et al (2008) IFATS collection: combinatorial peptides identify alpha5beta1 integrin as a receptor for the matricellular protein SPARC on adipose stromal cells. Stem Cells 26:2735–2745

Norose K et al (1998) SPARC deficiency leads to early-onset cataractogenesis. Invest Ophthalmol Vis Sci 39:2674–2680

Novinec M, Kordis D, Turk V, Lenarcic B (2006) Diversity and evolution of the thyroglobulin type-1 domain superfamily. Mol Biol Evol 23:744–755

Pottgiesser J et al (1994) Changes in calcium and collagen IV binding caused by mutations in the EF hand and other domains of extracellular matrix protein BM-40 (SPARC, osteonectin). J Mol Biol 238:563–574

Putnam NH et al (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317:86–94

Rentz TJ, Poobalarahi F, Bornstein P, Sage EH, Bradshaw AD (2007) SPARC regulates processing of procollagen I and collagen fibrillogenesis in dermal fibroblasts. J Biol Chem 282:22062–22071

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Sasaki T et al (1997) Limited cleavage of extracellular matrix protein BM-40 by matrix metalloproteinases increases its affinity for collagens. J Biol Chem 272:9237–9243

Sasaki T, Hohenester E, Gohring W, Timpl R (1998) Crystal structure and mapping by site-directed mutagenesis of the collagen-binding epitope of an activated form of BM-40/SPARC/osteonectin. EMBO J 17:1625–1634

Schmid V, Ono SI, Reber-Muller S (1999) Cell–substrate interactions in cnidaria. Microsc Res Tech 44:254–268

Schwarzbauer JE, Spencer CS (1993) The Caenorhabditis elegans homologue of the extracellular calcium binding protein SPARC/osteonectin affects nematode body morphology and mobility. Mol Biol Cell 4:941–952

Schwarzbauer JE, Musset-Bilal F, Ryan CS (1994) Extracellular calcium-binding protein SPARC/osteonectin in Caenorhabditis elegans. Methods Enzymol 245:257–270

Sodek KL et al (2008) Identification of pathways associated with invasive behavior by ovarian cancer cells using multidimensional

protein identification technology (MudPIT). Mol Biosyst 4:762–773

Tai IT, Tang MJ (2008) SPARC in cancer biology: its role in cancer progression and potential for therapy. Drug Resist Updat 11:231–246

Termine JD, Kleinman HK, Whitson SW, Conn KM, McGarvey ML, Martin GR (1981) Osteonectin, a bone-specific protein linking mineral to collagen. Cell 26:99–105

Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vannahme C et al (1999) Molecular cloning of testican-2: defining a novel calcium-binding proteoglycan family expressed in brain. J Neurochem 73:12–20

Vannahme C et al (2002) Characterization of SMOC-1, a novel modular calcium-binding protein in basement membranes. J Biol Chem 277:37977–37986

Vannahme C, Gosling S, Paulsson M, Maurer P, Hartmann U (2003) Characterization of SMOC-2, a modular extracellular calcium-binding protein. Biochem J 373:805–814