GBE

Maintenance of a Protein Structure in the Dynamic Evolution of TIMPs over 600 Million Years

Aldo Nicosia[1], Teresa Maggio[2], Salvatore Costa[3], Monica Salamone[1], Marcello Tagliavia[1], Salvatore Mazzola[1], Fabrizio Gianguzza[3,*], and Angela Cuttitta[1,*]

[1]Laboratory of Molecular Ecology and Biotechnology, National Research Council-Institute for Marine and Coastal Environment (IAMC-CNR) Detached Unit of Capo Granitola, Torretta Granitola, Trapani, Sicily, Italy

[2]Institute for Environmental Protection and Research-ISPRA, Palermo, Sicily, Italy

[3]Dipartimento Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche, University of Palermo, Sicily, Italy

*Corresponding author: E-mail: fabrizio.gianguzza@unipa.it; angela.cuttitta@iamc.cnr.it.

Abstract

Deciphering the events leading to protein evolution represents a challenge, especially for protein families showing complex evolutionary history. Among them, TIMPs represent an ancient eukaryotic protein family widely distributed in the animal kingdom. They are known to control the turnover of the extracellular matrix and are considered to arise early during metazoan evolution, arguably tuning essential features of tissue and epithelial organization. To probe the structure and molecular evolution of TIMPs within metazoans, we report the mining and structural characterization of a large data set of TIMPs over approximately 600 Myr. The TIMPs repertoire was explored starting from the Cnidaria phylum, coeval with the origins of connective tissue, to great apes and humans. Despite dramatic sequence differences compared with highest metazoans, the ancestral proteins displayed the canonical TIMP fold. Only small structural changes, represented by an α-helix located in the N-domain, have occurred over the evolution. Both the occurrence of such secondary structure elements and the relative solvent accessibility of the corresponding residues in the three-dimensional structures raises the possibility that these sites represent unconserved element prone to accept variations.

Key words: cnidarians, transcriptome wide analysis, TIMPs, homology modeling.

Introduction

It is generally admitted that three-dimensional (3D) structures of proteins are more conserved than sequences. Indeed structures change at a slower rates than the sequences from which they are derived.

As a result of divergent and convergent evolutionary mechanisms, proteins with low sequence identities (<20%) can share almost identical 3D structures (Flores et al. 1993). Concurrently, similar structures, without sequence or functional similarities, may arise both from processes of functionally divergent or structurally convergent evolution (Salemme et al. 1977).

It has been pointed out that changes in protein evolution can accumulate in such a way that the pattern of possible and allowable amino acid substitutions satisfy restraints arising from local structural environments (Blundell and Wood 1975; Gong et al. 2009), including solvent accessibility of side chain (relative solvent accessibility [RSA]) and occurrence of secondary structure elements (SSEs).

Understanding the forces that drive structure evolution represents a challenge especially for protein families, which show complex evolutionary history (Salemme et al. 1977). Among them, TIMPs are an ancient eukaryotic protein group, widely distributed in the animal kingdom, which are considered to be coeval with the origins of connective tissue in Metazoa (Brew and Nagase 2010).

TIMPs, are responsible for the control of turnover of the extracellular matrix (ECM) as they form 1:1 complex with matrix metalloproteinases (MMPs) and inhibit their activity. Therefore, MMPs activity is critically and tightly regulated by TIMPs to maintain a balance in ECM remodeling and degradation (Okamoto et al. 1977; Montagnani et al. 2007; Murphy and Nagase 2008).

Humans possess four TIMP isoforms (-1, -2, -3, and -4), which are the most characterized respect to structure, activity, and function. They comprise 184–194 amino acids stabilized by six disulfide bonds with an N- and C-terminal domain

(Docherty et al. 1985; Stetler-Stevenson et al. 1990; Pavloff et al. 1992; Williamson et al. 1994; Gasson et al. 1995; Greene et al. 1996).

In addition, TIMPs carry out several other functions, with an MMP-independent mechanism, as in angiogenesis, cell survival pathways, and tumor progression (Seo et al. 2003; Jung et al. 2006; D'Alessio et al. 2008).

Because of their implication in numerous processes of basal importance for whole organism systems, TIMPs activity and diversity has been described in several studies (Brew et al. 2000, 2010; Murphy 2011), and double-domain proteins were identified in different branches of the tree of life including vertebrates, as Japanese flounder *Paralichthys olivaceus* (Kubota et al. 2003), ghostshark *Callorhinchus milii*, and grass carp *Ctenopharyngodon idella* (Xu et al. 2011), and invertebrates, such as *Drosophila melanogaster* (Pohar et al. 1999) and blood clam *Tegillarca granosa* (Wang et al. 2012).

Single-domain TIMPs, which possess a different architecture to those of their mammalian relatives, were also found in certain invertebrates as occurred in the nematodes *Caenorhabditis elegans*.

On the basis of TIMPs domain organization and evolutionary relationships among species, it has been proposed that TIMPs evolution does not appear to have been a linear process (Brew et al. 2000).

Furthermore, there is little characterization of TIMPs in ancestral organisms as in cnidarians (Reitzel et al. 2008).

The Cnidaria phylum is sister taxa of Bilateria within the Eumetazoan and some studies demonstrated that time of divergence between Cnidarians and Bilaterians dates back to 600 Ma in the Precambrian era (Ball et al. 2004; Morris 2006). The phylum split into two major lineages: Anthozoa and its sister group, the Medusozoa, comprising three classes Scyphozoa (true jellyfishes), Cubozoa (box jellyfishes), and Hydrozoa (Bridge et al. 1992; Collins 2002). Within the Cnidaria, the Class Anthozoa is phylogenetically basal, and both morphological and molecular evidence are consistent with the idea that its members best represent the primitive cnidarians and are therefore least changed from the Ureumetazoan.

Cnidarians are very simple aquatic organisms, characterized by radial symmetry, and a single digestive opening. They show a basic diblastic organization consisting in two cell layers, the ectoderm and endoderm, separated by the mesoglea, which constitutes an ECM resembling the basal lamina with type IV collagen (Sarras and Deutzmann 2001; Shimizu et al. 2008). Interestingly, cnidarians are known to represent the first organisms whose cells become differentiated and organized into tissues and mechanisms of mesoglea remodeling MMPs dependent have been described, at least in hydra (Leontovich et al. 2000; Yan et al. 2000; Fujisawa 2003; Sarras 2012). Thus, as sister group of the Bilateria, they are of great interest for the identification of mechanisms leading to bilaterian traits (Martindale et al. 2004; Burton 2008).

Moreover, despite the simple anatomy, recent genomic and transcriptional surveys revealed that the complexity of some cnidarian gene families was similar to that found in vertebrates (Kortschak et al. 2003; Yang et al. 2003; Steele et al. 2011; Technau and Steele 2011).

Although a few computationally annotated TIMP sequences exist for the anthozoan *Nematostella vectensis* and for hydrozoan *Hydra magnipapillata*, no TIMPs have been studied in basal cnidarian species as the stony coral *Acropora millepora* or in the Mediterranean snakelocks *Anemonia viridis*.

In the last decade, several transcriptome platforms for these species have been produced (Kortschak et al. 2003; Putnam et al. 2007; Meyer et al. 2009; Sabourault et al. 2009; Chapman et al. 2010), thus allowing to perform computational analysis.

To explore the structural diversity of TIMPs in such basal metazoans, we performed an extensive transcriptional survey for the mining of expressed TIMPs from *A. millepora*, *An. viridis*, *N. vectensis*, and *H. magnipapillata*.

Exploiting the advances in homologs detection and alignment implemented in PSI-BLAST (Altschul et al. 1997), hidden Markov models (Krogh et al. 1994), and homology protein modeling (Kelley et al. 2015; Salamone et al. 2015), theoretical structure prediction methods were herein extensively used to characterize the structures and molecular evolution of TIMPs.

We explored the TIMPs repertoire over 600 Myr, spanning from cnidarians to great apes and humans, to unveil mechanisms of TIMP evolution. Thus, we compare the cnidarian structures with TIMPs from phylogenetically different organisms, representative of the animal evolution history, and discuss them in the light of the relationship between the evolutionary rate of a region and structural constraints, including SSEs and RSA.

Despite dramatic sequence diversity, evidence of an evolutionary conservation of a prototypic structure emerged; while differences in conservation of a specific SSE, represented by an unusual α-helix located in the N-domain, were also observed.

Materials and Methods

Data Mining

The characterized TIMPs reported in literature were used initially to retrieve their corresponding sequences from the publicly available database at the National Centre for Biotechnology Information. To obtain more putative TIMP homologs, these sequences were used as query to perform extensive BLASTP, BLASTX, and TBLASTN searches until no novel putative matches could be retrieved.

The putative matching sequences from *A. millepora*, *An. viridis*, *N. vectensis*, and *H. magnipapillata* were obtained from public available databases under the following accession

numbers: FK740554.1 (*AvTIMP*); XP_001630935.1 (NvTIMPa); XP_001641739.1 (NvTIMPb); XP_001641336.1 (NvTIMPc); XP_002156125.2 (HmTIMP3a); XP_002160236.1(HmTIMP3b); XP_004210412.1 (HmTIMPa); XP_002165494.1 (HmTIMPb); GB_EZ032238.1 (AmTIMPa); GB_JR979316.1 (AmTIMPb); GB_JR985821.1 (AmTIMPc); GB_JR990988.1 (AmTIMPd); GB_JT010482.1 (AmTIMPe); GB_JR989567.1 (AmTIMPf); and GB_JT010782.1 (AmTIMPg). The protein sequence XP_001625448.1, used as TIMP from *N. vectensis* has not been included in this study as it is not supported by expressed sequence tag (EST) evidence.

Sequence and Structure Evolution of TIMPs

Signal peptides, functional sites, and domains in the predicted amino acid sequences were predicted using the Simple Modular Architecture Research Tool program, the InterPro database, the Pfam database, the PROSITE program, and the Eukaryotic Linear Motif resource for Functional Sites in Proteins. N-glycosylating sites were predicted by NetNGlyc 1.0 Server software. Alignments were generated using ClustalW (Thompson et al. 1994) and Tcoffe (Notredame et al 2000) as implemented in Molecular Evolution Genetics Analyses (MEGA) software version 6 (http://www.megasoftware.net/mega.php) (Tamura et al. 2013). Alignments were manually checked for substitution matrix and rendered using the ESPript 3.0 server (http://espript.ibcp.fr/ESPript/ESPript/).

To reconstruct the molecular evolution of the multifunctional TIMP family, evolutionary diversification and phylogenetic relationship were studied at two different levels, among cnidarians and between cnidarians and other Eumethazonas, as listed in supplementary table S1, Supplementary Material online.

In particular, phylogenetic analyses were conducted on the amino acid sequences of the TIMP domain using a neighbor-joining (NJ) method, implemented in MEGA version 6.0, in which Poisson correction, pairwise deletion, and bootstrapping (1,000 replicates) were considered as parameters, to reconstruct the evolutionary diversification and the molecular evolution of TIMP proteins family in Cnidarians and different Eumetazoan groups. Moreover, the 3D structures of TIMP proteins were reconstructed by homology modeling via the Protein Homology/analogY Recognition Engine 2.0 (Phyre 2) software (Kelley et al. 2015) using the intensive modeling mode. Candidate structures for homology modeling were selected according to pairwise alignment. At least two different structures were used as template for each generated structures, and homology models were built for all of the sets of proteins. Validation of the structural protein models was performed by assessing the Ramachandran plots. Cycles of clash minimization were also performed for the refinement of structures.

Secondary structures assignments and relative solvent accessibility (RSA) were calculated by the DSSP program (Kabsch et al. 1983; Touw et al. 2015) as implemented in ENDscript, http://endscript.ibcp.fr (Robert and Gouet 2014).

The generated models for cnidarians TIMPs were submitted to SuperPose server for structural superposition (Maiti et al. 2004) with TIMPs 3D structures (deriving from PDB databank, TIMP-1 [PDB:1O09], TIMP-2 [PDB:1BR9], and TIMP-3 [PDB:3CKI] or computationally generated) of all other Eumethazoa to evaluate changes in the 3D structures during the evolution. Molecular graphics and analyses, including root mean square deviation (RMSD) measurements after superposition, were performed with the UCSF Chimera package (Pettersen et al. 2004).

Results

Identification of TIMPs and TIMP-Related Proteins in Cnidarians

First, we performed comprehensive analyses of the TIMP homologs in four cnidarian species representative of the ancestral condition of metazoans. The availability of large-scale transcriptional data sets allowed us to carry out transcriptome surveys in the stony coral *A. millepora*, in sea anemones *An. viridis* and *N. vectensis*, and in the hydrozoan *H. magnipapillata*.

To identify the expressed TIMP-related proteins, we implemented TBLASTN, BLASTX, and BLASTP searches. Collectively, 15 putative TIMP homologs were retrieved in such cnidarians (table 1). Given the high divergence of TIMP-related proteins, their identification was checked manually and the matching sequences were reconfirmed by comparative analysis. Several other predicted proteins were detected in the database but were not subjected to further analysis as they contained substantial truncations, showed other domains in addition to the TIMP, or they were considered likely to be an artifact because of the absence of any supporting ESTs.

To avoid confusion in nomenclature, we used the names provided in the databases or in previous reports. In addition, sequences clearly representing orthologs of mammalian TIMPs were named accordingly, whereas the newly identified sequences with no sequence similarity to the known mammalian TIMPs were alphabetically ordered.

Our survey retrieved three, four, and seven putative TIMP homologs in *N. vectensis* (herein named as NvTIMPa, NvTIMPb, and NvTIMPc), *H. magnipapillata* (herein named as HmTIMP3a, HmTIMP3b, HmTIMPa, and HmTIMPb), and *A. millepora* (herein named as AmTIMPa, AmTIMPb, AmTIMPc, AmTIMPd, AmTIMPe, AmTIMPf, and AmTIMPg), respectively, whereas exclusively one matching sequence encoding putativeTIMP homolog (AvTIMP) was found among over 39,000 ESTs for *An. viridis*.

Table 1

Putative TIMPs from Cnidarians

Organism	Accession Number	Protein	Protein Length[a]	Molecular Weight[b]	pI[c]
Acropora millepora	222803850	AmTIMPa	159	18,596.5	9.03
A. millepora	379081647	AmTIMPb	237	27,323.9	9.44
A. millepora	379088152	AmTIMPc	221	25,040.1	9.26
A. millepora	379093319	AmTIMPd	158	17,163.8	9.40
A. millepora	379112812	AmTIMPe	161	18,376.1	9.69
A. millepora	379091898	AmTIMPf	180	21,006.2	9.51
A. millepora	379113112	AmTIMPg	179	20,637.6	8.77
Anemonia viridis	FK740554.1	AvTIMP	263	30,489.7	10.29
Nematostella vectensis	156378005	NvTIMPa	296	34,389.3	10.48
N. vectensis	156408189	NvTIMPb	139	15,228.8	9.21
N. vectensis	156407007	NvTIMPc	290	32,745.0	9.74
Hydra magnipapillata	449680372	HmTIMP3a	223	25,559.7	8.31
H. magnipapillata	221128951	HmTIMP3b	224	25,918.3	8.70
H. magnipapillata	449683625	HmTIMPa	313	35,627.4	8.59
H. magnipapillata	221116789	HmTIMPb	209	23,423.5	9.13

[a]Length (no. of amino acids) of the deduced protein.
[b]Molecular weight of the deduced polypeptide in Dalton.
[c]Isoelectric point of the deduced protein.

Since the absence of well-defined upstream elements in TIMP from the EST data set of *An. viridis*, the full-length cDNA was obtained by assembling the 5′ RACE product with the original EST (data submitted to another journal).

Conserved Structural Features of Cnidarian Proteins

On the basis of the computational analysis, we determined the key features of TIMP homologs in cnidarians. They consist of an N-terminal signal peptide, a TIMP domain organized in an N-domain and often C-terminal region, a pattern of conserved cysteine residues required for disulfide bond formation and putative N-Linked glycosylation sites, which are involved in the inhibitor activities (Nancy et al. 1998). A schematic representation of these proteins is reported in figure 1.

With the exception of NvTIMPb, AmTIMPa, and AmTIMPd, TIMPs from cnidarians are synthesized as inactive precursors with a typical signal peptide and a cleavage site, generally located within the first 30 amino acidic residues, suggesting they represent secreted proteins (Brew et al. 2000; Zhang et al. 2003). Interestingly, predictions of transmembrane domains in the N-termini of cnidarian TIMP homologs could be extended to other members of such family including human TIMP-4. Thus, it is possible to suppose that such structures may play a role in the protein delivery and sorting into the membrane or to the secretory pathway (Zhang et al. 2003).

TIMPs from the coral *A. millepora* consist of members both showing the two domains and the single-domain organization. AmTIMPa-b-c and -d contain N-terminal domain with OB-fold and a C-terminal region (amino acid residues 1–146 in AmTIMPa, 24–224 in AmTIMPb, 28–214 in AmTIMPc, and 15–150 in AmTIMPd) (fig. 2). Conversely, AmTIMPe,

AmTIMPf, and AmTIMPg are organized as single-domain proteins in which the TIMP/NTR domain is located between amino acid residues 25–159 (AmTIMPe), 37–171 (AmTIMPf), and 31–170 (AmTIMPg). Additionally, AmTIMPa-d-f and-g lack the canonical Cys-X-Cys motif downstream the cleavage/activation site, which has been considered as a key feature for protease inhibition.

AvTIMP contains a putative TIMP/NTR domain (amino acid residues 73–245) consisting of an N-terminal domain with the OB-fold (amino acid residues 73–171) and a smaller C-terminal region (amino acid residues 172–246). The deduced protein presents a putative N-terminal signal peptide (residues 1–24), a cleavage site for proteolytic activation located between residues 24 and 25 and contains a transmembrane domain with high hydrophobic moment (amino acid residues 2–19) partially overlapped with the N-terminal signal sequence. Generally, putative homologs from *N. vectensis* are single-domain proteins showing the N-inhibitory domain located between amino acid residues 166 and 290 in NvTIMPa and amino acid residues 4–138 in NvTIMPb. Conversely, NvTIMPc is representative of the bipartite TIMP/NTR proteins with an N-domain (18–189) and the C-terminal structure (190–262).

As described above, for some members from the stony coral, proteins from sea anemones *An. viridis* and *N. vectensis* lack the canonical Cys-X-Cys pattern usually required for inhibition of MMPs. Computational analyses also predicted the presence of putative N-glycosylation sites in all the proteins including NvTIMPb, AmTIMPa, and AmTIMPd, which lack the N-terminal signal peptide. As N-glycosylation occur in the ER and Golgi, only proteins with a signal peptide are expected to be glycosylated. Thus, even though potential motifs have been
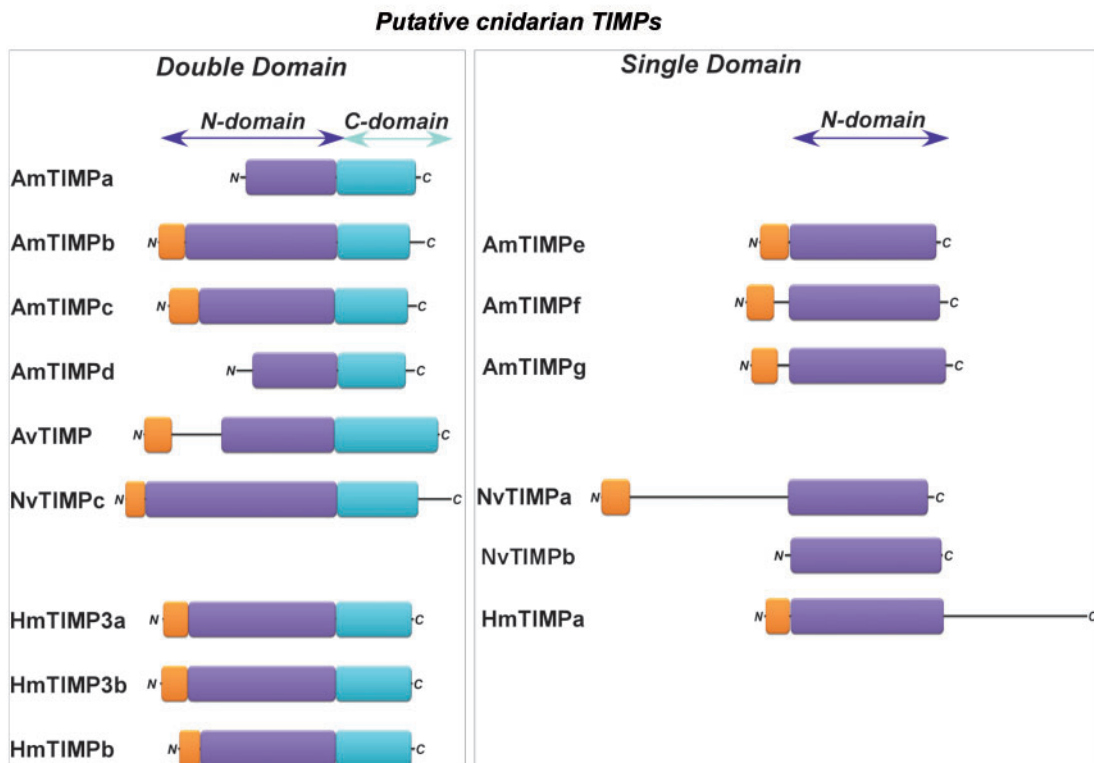
Putative cnidarian TIMPs

Fig. 1.—Architecture of conserved protein domains of putative TIMP homologs in cnidarians. Putative TIMPs from cnidarians consist of members both showing the two domains and the single-domain organization. Signal peptides are shown in orange, N-domains are shown in purple, and C-domains are shown in pale blue.
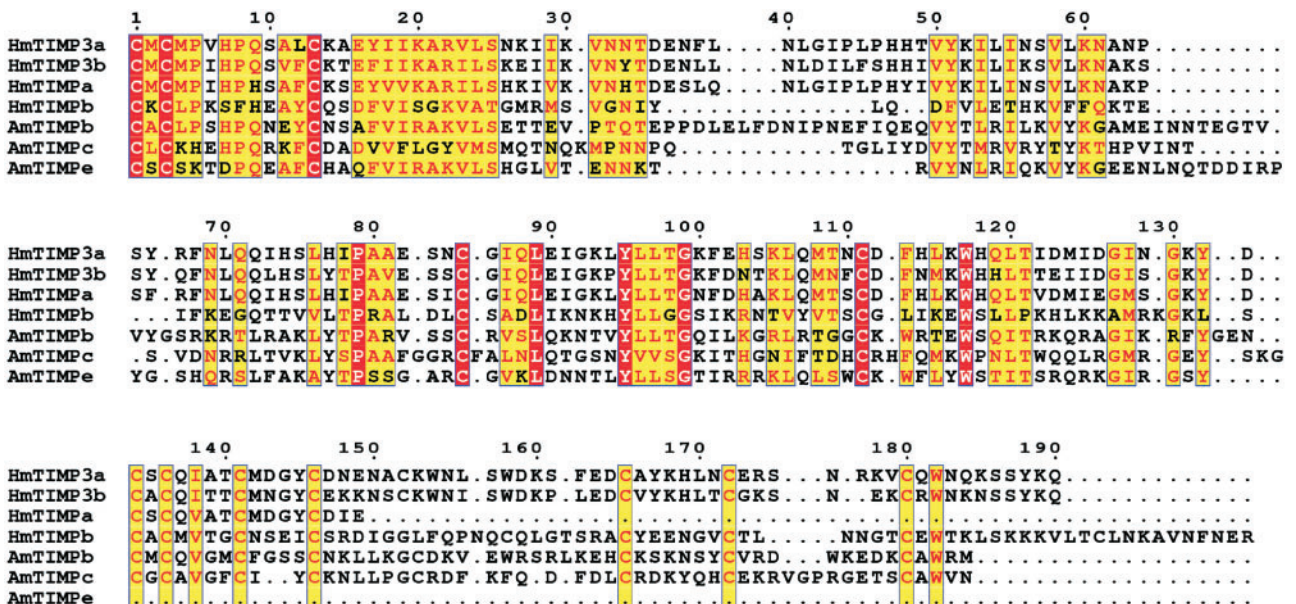

Fig. 2.—Multiple sequence alignment of the selected TIMPs in cnidarians. Alignment was performed with T-coffee. Similar residues are written in red bold characters and boxed in yellow, whereas conserved residues are in white bold characters and boxed in red. The sequence numbering on the top refers to the alignment. Abbreviations, species, and accession numbers are listed in table 1.

mapped in NvTIMPb, AmTIMPa, and AmTIMPd, they are unlikely to be exposed to the N-glycosylation machinery.

A proline-rich region, located between the signal peptide and the NTR domain (residues 25–72), was identified in the AvTIMP protein sequence. Similarly, NvTIMPa and AmTIMPf show a region separating the signal peptide from the NTR. Not surprisingly, TIMPs from several species including sea urchin *Strongylocentrotus purpuratus* and Pacific oyster *Crassostrea gigas* (Montagnani et al. 2005), present insertions in the N-terminal domain. Thus, both the processing and activity of TIMPs could not be affected by spacers of variable length.

Finally, the TIMPs from Hydra appear to be double-domain proteins as exclusively HmTIMPa lacks the C-domain. HmTIMP3a, 3b, and b show the canonical features of double-domain TIMPs in which the cleavage sites for proteolytic activation is located between signal peptides and the N-domain. The smaller C-terminal region is also present.

As previously described, several members from *A. millepora* (AmTIMPa-d-f and-g), AvTIMP, and TIMPs from *N. vectensis* lack the usual Cys-X-Cys motif downstream the cleavage/activation site. Because the absence of the Cys[1] in the corresponding mature proteins, which is responsible for the bidentate coordination of Zn^{2+} and MMPs inhibition, as well as the absence of N-Linked glycosylation sites or signal peptides, they were not considered for further analysis.

Structural and Phylogenetic Analysis of Cnidarian TIMPs

Because the high variability of the N- and C-terminal ends, a multiple sequence alignment was constructed for the selected TIMP proteins from *A. millepora* and other homologs of *H. magnipapillata* (fig. 2). Such alignment resulted in 181 variable residues, 69 of which were parsimony informative over a total of 226 amino acidic residues.

Therefore multiple sequence alignment for TIMPs from the stony coral *A. millepora* with *Hydra* ones retrieved identity comprised between 20.12% and 31.69%. In supplementary table S2, Supplementary Material online, are reported number of amino acid substitutions per site from each comparisons calculated using Poisson correction model, whereas supplementary tables S3 and S4, Supplementary Material online, reported pairwise percentage of identity.

Despite limited protein sequence identity, MSA analysis (fig. 2) provided evidences that the accepted amino acids changes showed similar physical chemical features. Therefore, we argue that amino acid substitutions may be able to fulfil the same structural and functional roles. To confirm such assumption, we computed the secondary elements and derived 3D structures of cnidarian TIMPs.

Different templates were selected to model cnidarian TIMPs based on heuristics to maximize confidence, percentage identity, and alignment coverage, while the N-term and the insertions, if required, were modeled ab initio. The generated models were validated by assessing Ramachandran plot

analysis and the percentage of residues in the favored/allowed region ranged from 90% to 95%. Modeling analysis predicted both and α-helical structures and β-strand configurations for the mature TIMPs similar to those described in previous studies. Hence, amino acid residues organizing helices and strands were substituted in ways that they maintained the overall stabilities of the secondary structural elements required for function (fig. 3).

In TIMPs from *H. magnipapillata*, the N-inhibitory domains are folded in a closed β-barrel usually composed of six β-strands and the C-terminal domains are organized in the canonical two parallel and two antiparallel strands followed by α-helix. A region organized in α-helices is located at the interface of these two domains. Additionally in HmTIMPa, in absence of the C-domain, several loops and long α-helices structured the C terminus of the protein.

Furthermore, an analysis of these structures indicates the presence of a helix located in the N-domain for HmTIMP3a, HmTIMP3b, and HmTIMPa. Interestingly, the occurrence of these helices is unusual if compared with well-defined template structures lacking such element.

Additionally, it appears that HmTIMPa shares an amazing structural similarity with HmTIMP3b in absence of the corresponding C domain. Thus we argued that it may likely represent a truncated form by loss of the C-domain.

Similar to *hydra* homologs, TIMPs from the coral *A. millepora* are variable in 3D structure than since this family consists of both members showing the two domains organization (AmTIMPb and AmTIMPc) and single-domain protein (AmTIMPe). AmTIMPb and AmTIMPc showed the canonical pattern of secondary elements as their N domains encompass a five- or six-stranded β-barrel with the Greek key topology. In the two domain-containing proteins, a central helical region is located at the interface of the domains, and the C-terminal structure is composed of two parallel and two antiparallel strands. Additionally, the single-domain protein AmTIMPe possesses an organization characterized by the insertion of an additional β-strand and β-3 appeared to be shifted downstream. To better understand the relationship between protein structure and diversity of TIMPs, prediction of SSEs was combined with evaluation of RSA of the corresponding residues in the folded protein. In particular, the RSA was computed for each amino acid residues structuring the unusual α-helices, previously described in some TIMPs, and the relative values for the corresponding residues are reported in table 2.

Despite sequence diversity, all cnidarian homologs presenting the additional helix in their N-domain showed a similar pattern of solvent accessibility in the corresponding residues.

Based on RSA values, α-helices were found to mainly consist of exposed and highly exposed residues (with $0.4 < RSA \leq 1$ and $RSA > 1$, respectively) or intermediate ones (with $0.1 \leq RSA \leq 0.4$). Conversely, buried residues ($RSA < 0.1$) were found to be strongly underrepresented.
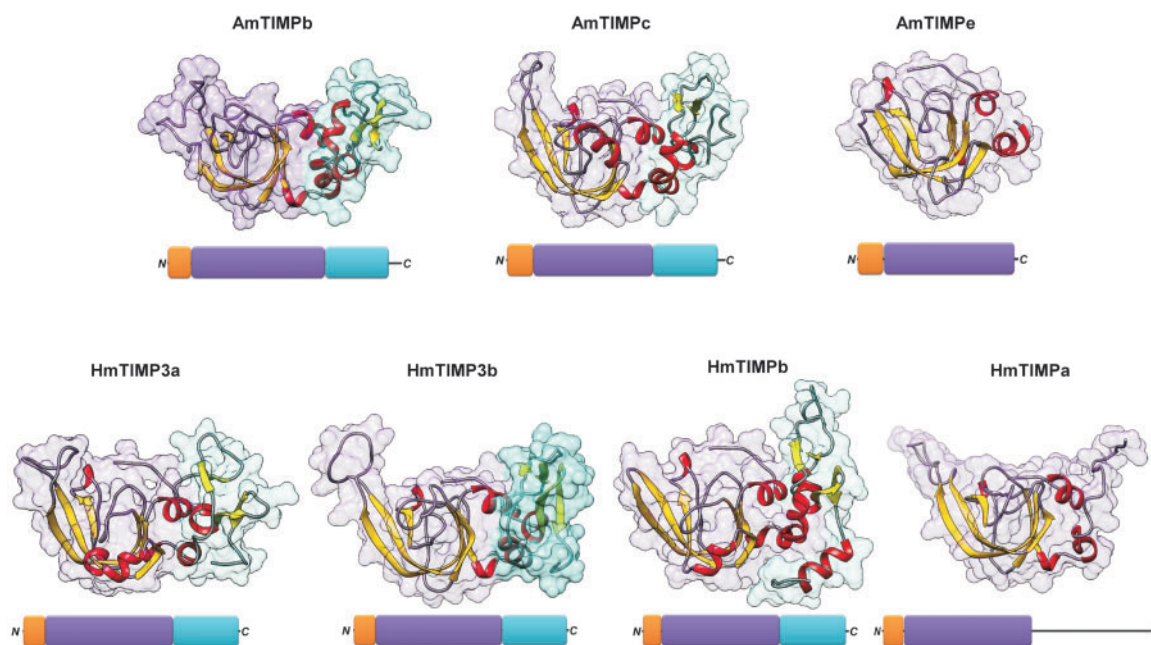
FIG. 3.—Ribbon diagrams an surface representations of the selected cnidarian TIMP structures studied in this work. General overview of the 15 TIMPs generated by homology modeling. The proteins were colored according to secondary structure with beta-sheets in yellow and helices in magenta. The protein surfaces were according to the single- or double-domain organization with the N-domain shown in purple and C-domains shown in pale blue.

Table 2

TIMPs Showing the Unusual Helices in the N-Domain and RSA Analysis

	Species Name	Accession Number	TIMP Isoform	Helix Residues[a]
Mammals	*Pan troglodytes*	XP_516284.1	PtTIMP-4	FEKV
				EBEE
		XP_515097.2	PtTIMP-3	KMPKV
				IEIBB
	Rattus norvegicus	NP_446271.1	RnTIMP-1	FDA
				EEI
		NP_001102863.1	RnTIMP-4	FEKAK
				IBEEE
Birds	*Columba livia*	EMC77392.1	ClTIMP-4	FEKL
				IBEE
Reptile	*Chelonia mydas*	XP_007056544.1	CmTIMP-4	FEKV
				EBEE
Echinoderm	*Strongylocentrotus purpuratus*	XP_781027.1	SpTIMP-3	EKLKH
				EEBEE
Insect	*Stegodyphus mimosarum*	KFM62985.1	SmTIMP	EKARRA
				EEEBEE
Molluscs	*Crassostrea gigas*	AAT73610.1	CgTIMP-1	SLLGS
				EEBIE
		NP_001292265.1	CgTIMP-2	KGSSLL
				IBBEEI
	Tegillarca granosa	AFB81539.1	TgTIMP	PAFEEL
				EEEBEE
Cnidarians	*Hydra magnipapillata*	449680372	HmTIMP3a	NPSYRFNLQQIH
				EIIIEEEIEEBB
		221128951	HmTIMP3b	YQFNL
				EEIIEE
		449683625	HmTIMPa	NLQQIH
				EBEEBI

[a]Amino acid residues structuring the helices and corresponding relative solvent accessibility: exposed (*E*), intermediate (*I*), and buried (*B*) residues.
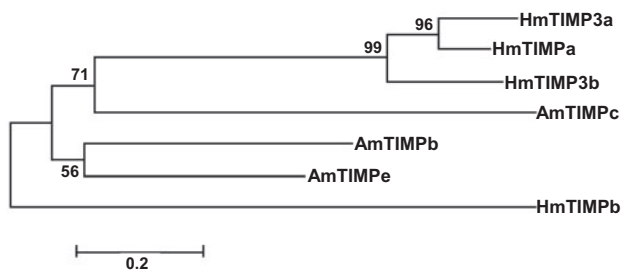
FIG. 4.—NJ phylogenetic tree based on the selected TIMPs from *Acropora millepora* and *Hydra magnipapillata*. The sequences used were obtained from GenBank at National Centre for Biotechnology Information (NCBI) and listed in table 1. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Internal branches were assessed using 1,000 bootstrap replications. Bootstrap values greater than 50% are indicated at the nodes.

To estimate the diversity among cnidarians TIMPs, their evolutionary relationships were analyzed using the NJ method with Poisson correction and are displayed in the figure 4. The phylogenetic analysis revealed a tree topology with a noncorresponding species distinction indicating the intraspecific variability of TIMP proteins among the analyzed cnidarians. Additionally, single- and double-domain TIMPs appear to be randomly distributed across the cnidarians herein analyzed, while HmTIMPb represents the most ancestral one.

Phylogenetic and Structural Analysis of TIMPs during the Evolution

To gain further insights and establish homology relationships between Cnidarians and Bilaterians, sequence similarity analyses were also performed considering TIMPs from Eumetazoa as listed in supplementary table S1, Supplementary Material online. Toward this end, TIMP homologs from different organisms representative of the Animalia evolutionary history were collected to infer phylogenetic relationship. The amino acid p alignment showed a high degree of variability with 277 variable sites over 253 total amino acids; 176 of which were parsimony informative.

The trees generated with NJ and MP methods were congruent as their overall topology is similar. In figure 5, the resulting phylogenetic tree is shown and the low support values (below 50) in some nodes suggest alternative branching patterns.

Cluster and topological analysis indicates that TIMP homologs from *C. elegans* and HmTIMPb rooted the tree in a paraphyletic grouping. After this, the cnidarian and sea urchin TIMPs based the branch of invertebrates in an unresolved manner as suggested by the lower bootstrap values supporting some nodes. Among them, phylogenetic grouping of TIMPs

was according to taxa distinctions; while, after these, the TIMPs from vertebrates branch-off on the basis of the four TIMP forms. In particular, the TIMP-1 resulted the basal one, being located on successively earlier branches consisting in invertebrates, as arthropods and molluscs, and mammals. Conversely, the other TIMPs (−2 to −4) arose later in the evolution.

It is also evident that the most recent branches consist of TIMP 4 forms and homologs are shared in birds, reptiles, and mammals.

An effort was also made to identify similarities of protein folds throughout the metazoan kingdom. Toward this end, the 3D structure of these TIMP homologs was predicted by homology modeling.

Despite considerable sequence differences, the proteins display a common structure. Only a few changes occurred in the 3D structure over 600 Myr and striking similarity emerged from Cnidaria to Vertebrata.

Similar to cnidarian TIMPs, homologs from arthropods molluscs, echinoderms as well as higher metazoans present the N-domain organized in the Greek-key motif, composed of β-strands arranged in a twisted and partially open β-barrel. Interestingly, a few members of this protein family presents an α-helix, located in the corresponding N-domain. This resembles the domain architecture previously described for some cnidarian TIMPs.

In particular, the occurrence of this SSE appears to be randomly distributed among metazoans as arthropods, molluscs, echinoderms, and vertebrates including apes were found to present such signature.

Moreover, to characterize the pattern of solvent accessibility for residues organizing these helices, RSA analysis was also performed for such TIMP homologs as carried out for cnidarians (table 2).

Despite difference in length and sequences, α-helices again were found to be predominantly determined by solvent-exposed residues, rather than buried ones.

To confirm the maintenance of a protein fold during the evolution, the members of cnidarian TIMPs were compared with the 3D structures of evolutionarily diverse TIMP proteins. Hence, algorithms to superimpose protein structures were applied, whereas structural divergences were evaluated and measured in terms of RMSD (supplementary table S5, Supplementary Material online).

TIMPs from cnidarians exhibit notable similarity in their SSEs as well as in the 3D organization when superimposed with lower Eumetazoan TIMPs (e.g., from Echinoderms, Artropods, and Molluscs) or higher metazoan as Fishes, Birds, Reptiles, and Amphibians.

Moreover, no striking differences emerged when the cnidarian structures were compared with the highest metazoans.

Finally, the structures of the mature cnidarian TIMPs can be also superimposed onto the resolved crystal structures of human sequences. Figures 6 and 7 illustrate representative
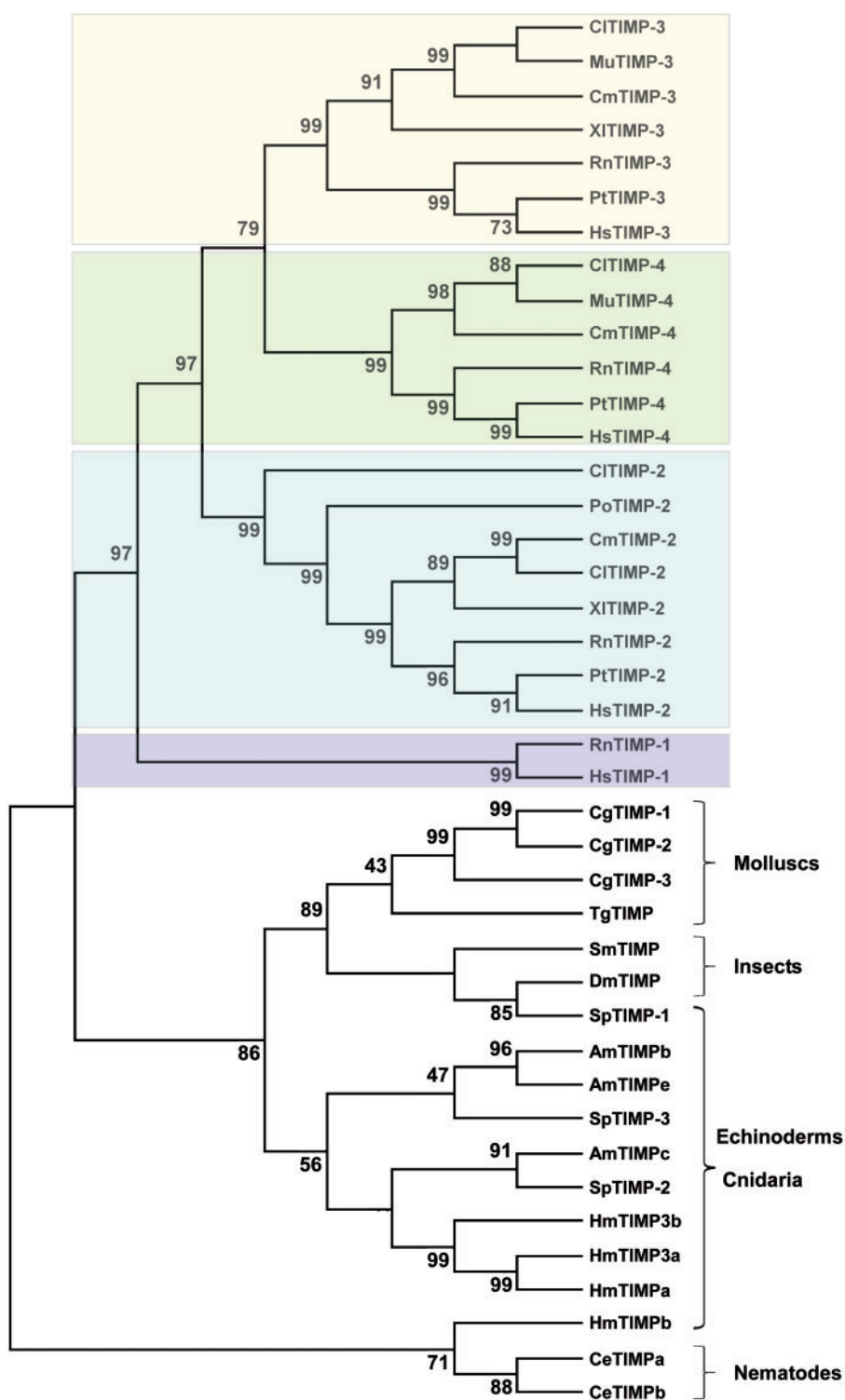
FIG. 5.—NJ phylogenetic tree based on the TIMPs of cnidarians and Bilateria. The tree was generated using MEGA 6.0 including TIMPs from 17 different species ranging from cnidarians to mammals. All the sequences used were obtained from GenBank at National Centre for Biotechnology Information (NCBI) and listed in supplementary table S3, Supplementary Material online. The p-distance model was used to construct the phylogenetic tree. Internal branches were assessed using 1,000 bootstrap replications. Bootstrap values greater than 40% are indicated at the nodes.
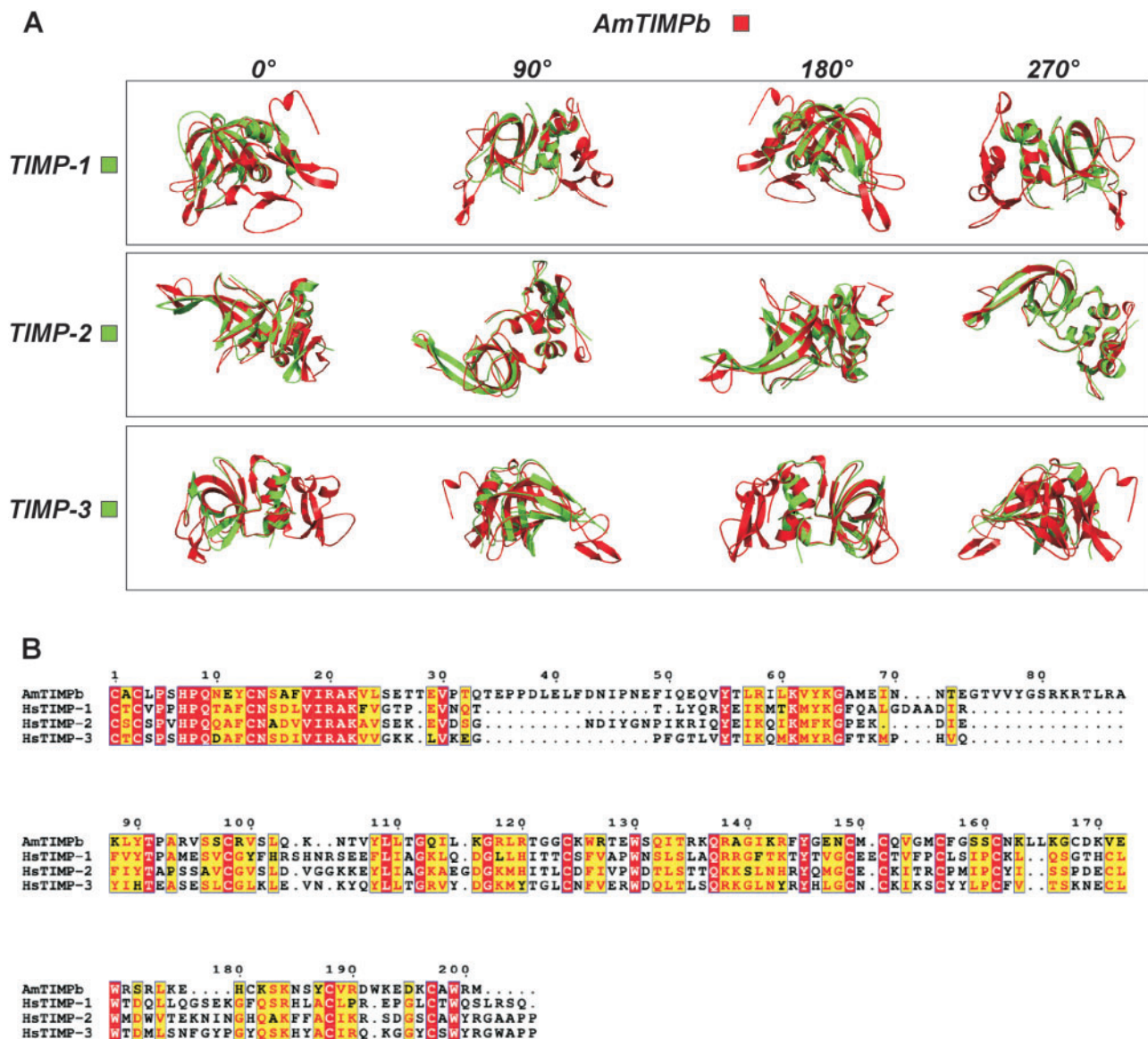
Fig. 6.—Structural similarities between TIMPb from the stony coral *Acropora millepora* and the human TIMP-1, -2, and -3. (*A*) 360° view of superposition of the 3D structure of AmTIMPb and TIMP-1(PDB:1OO9), TIMP-2 (PDB:1BR9), and TIMP-3 (PDB:3CKI). Proteins are in ribbon representation. TIMPb is shown in magenta and human TIMPs are colored in lime. Superposition was created and rendered by using Chimera package. (*B*) T-coffee structural-based sequence alignment of the mature forms of AmTIMPb and human TIMP-1, -2, and -3. Similar residues are written in red bold characters and boxed in yellow, whereas conserved residues are in white bold characters and boxed in red. The sequence numbering on the top refers to the alignment.

superpositions of HmTIMP3b and AmTIMPb with the humans HsTIMP-1, HsTIMP-2, and HsTIMP-3.

These proteins have highly similar structures because the pairwise computed RMSDs were <1 Å (supplementary table S5, Supplementary Material online). In particular, AmTIMPb showed RMSDs from 0.68 Å (TIMP-2) to 0.92 Å (TIMP-1), whereas RMSDs for HmTIMP3b ranged from 0.49 Å (TIMP-2) to 0.91 Å (TIMP-1). On the basis of these RMSD values, comparable with proteins sharing a much higher sequence identity (fig. 8), evidence of a structural conservation emerged.

Discussions

Different studies reported the identification of TIMP proteins in defined species and recently it has been shown that homologs of the TIMPs are distributed widely among both invertebrate and vertebrate animals. However, there is limited information on the properties of nonmammalian TIMPs. The availability of a growing number of sequenced genomes and transcriptomes allowed us to perform, for the first time, a comprehensive phylogenetic and structural analyses of
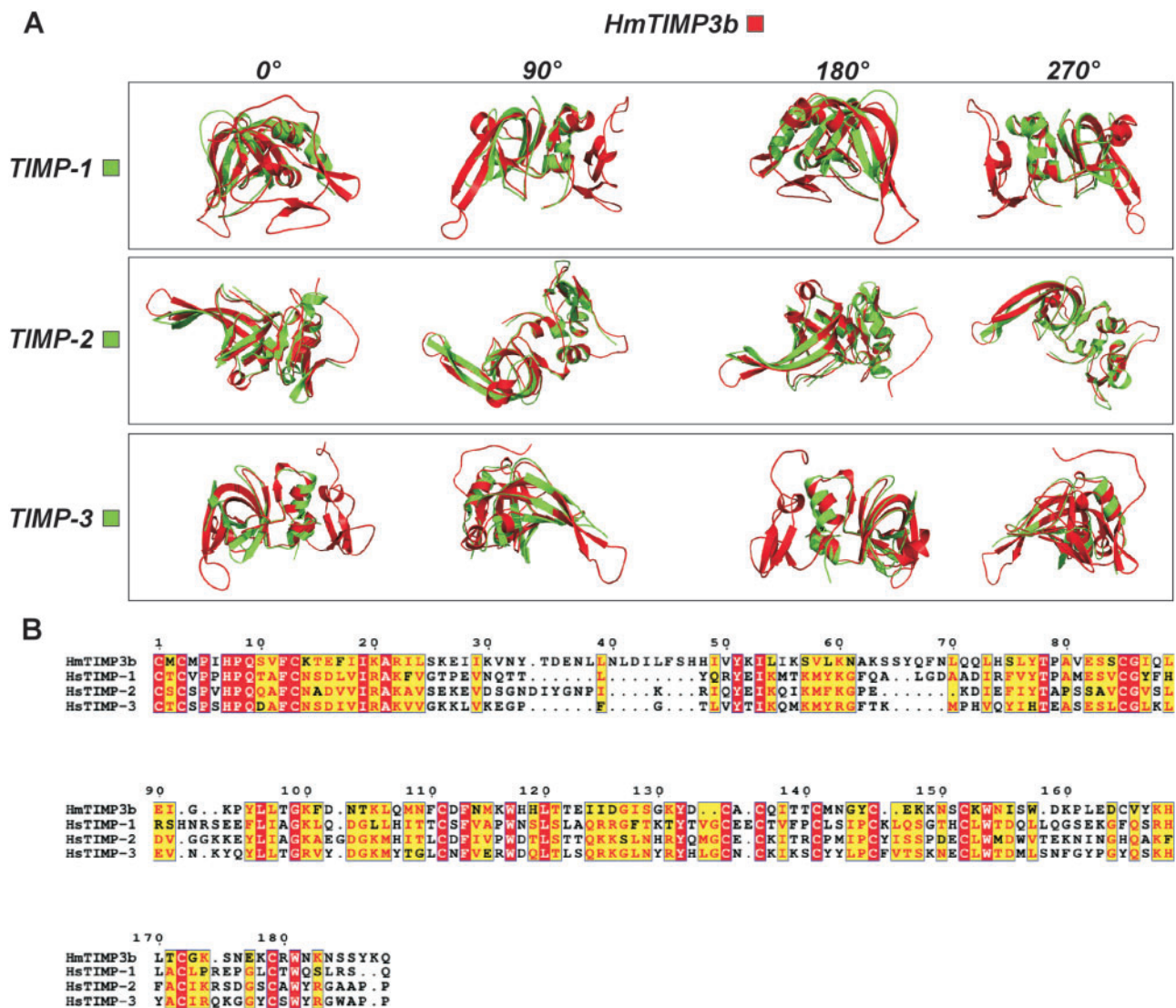
Fig. 7.—Structural similarities between TIMP3b from *Hydra magnipapillata* and the human TIMP-1, -2, and -3. (A) 360° view of superposition of the 3D structure of AmTIMPb and TIMP-1(PDB:1009), TIMP-2 (PDB:1BR9), and TIMP-3 (PDB:3CKI). Proteins are in ribbon representation. HmTIMP3b is shown in magenta, and human TIMPs are colored in lime. Superposition was created and rendered by using Chimera package. (B) T-coffee structural-based sequence alignment of the mature forms of HmTIMP3b and human TIMP-1, -2, and -3. Similar residues are written in red bold characters and boxed in yellow, whereas conserved residues are in white bold characters and boxed in red. The sequence numbering on the top refers to the alignment.

TIMPs. TIMP members were identified in organisms from different taxonomic divisions, ranging from earliest Bilaterians, as the cnidarians, to the highest metazoans, as apes and humans. This enabled us to trace the evolutionary origin of TIMPs during 600 Ma when it has been hypothesized that cnidarians diverged from the lineage leading to Bilateria (Ball et al. 2004).

Genomic and transcriptional surveys on cnidarians have shown that ECM-encoding genes originated before the eumetazoan radiation >600 Ma. This is consistent with the presence of the basement membranes, which separate the two epithelial layers in cnidarians. Additionally, mechanisms leading to ECM turnover and modification have been described

in hydra, which has been used as a model for studies on cell–ECM interactions during events of morphogenesis, regeneration, and transdifferentiation (Bode and Bode 1984; Bosch 1998). Similar to those occurring in vertebrate species, *H. magnipapillata* expresses metalloproteinases which regulates head and foot morphogenesis, remodeling of mesoglea via digestion of ECM components (Leontovich et al. 2000; Yan et al. 2000; Fujisawa 2003; Sarras 2012). Moreover, it has also been reported that hydra MMP is inhibited by recombinant human TIMP-1.

Additionally, several other predicted metalloproteinases were detected in cnidarians transcriptional databases, and in *A. millepora*, it has been hypothesized that Zn^{2+}
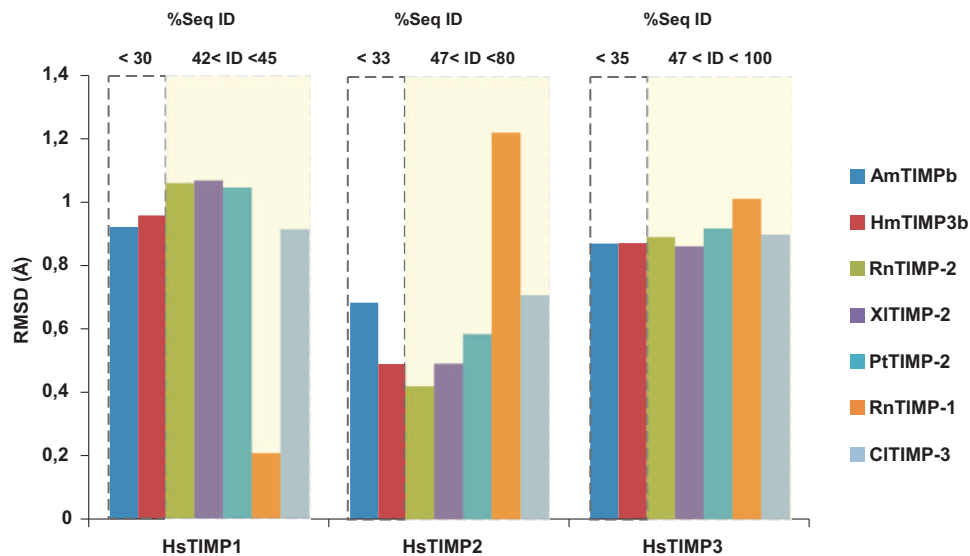
FIG. 8.—The change in RMSD between homologous protein pairs as they are superimposed onto human TIMPs. Cnidarian and vertebrate TIMPs were used. The percent sequence identities between each superimposed protein are shown. Similar RMSDs were calculated despite different sequence identities.

metalloproteases also affect the remodeling of the ECM (Császár et al. 2009).

Although only little information is available about the turnover of ECM components controlling mesoglea integrity in these basal organisms (Moya et al. 2012), TIMPs are widely known to control ECM catabolism in higher metazoans. Thus, it remains plausible to hypothesize that genes encoding TIMPs may be able to regulate the remodeling of cnidarian ECM components.

To gain insight into the TIMPs evolution, BLAST searches, supplemented by analyses of domain composition, retrieved several expressed members of such class in the earliest bilaterians as *A. millepora* and *H. magnipapillata*.

This was not unexpected, as the presence of at least one gene encoding TIMP appears to be a characteristic of most eumetazoans (Murphy 2011). Furthermore, neither the absence of sequences encoding TIMP homologs in *An. viridis*, showing the canonical Cys-X-Cys pattern at their N-terminal ends was surprising, as the corresponding EST database is considered to be only partial (Ganot et al. 2011), and fully sequenced genomes for the Mediterranean sea anemone are not available. However, it remains plausible to suppose that *An. viridis* may have fewer members of TIMPs than other cnidarians.

Although cnidarians are known to have diverged from Bilateria earlier than Protostomes and Deuterostomes (Martindale et al. 2002; Ball et al. 2004), these basal metazoans encode several TIMPs including both single- and double-domain homologs. Thus, it appears that an ancestral two-domain TIMP is already extant before the eumetazoan radiation >600 Ma.

Phylogenetic relationships, as herein reported, revealed an unusual status for the cnidarian TIMPs; in particular, *H. magnipapillata* concurrently expresses the ancestral and the derived form of TIMPs. It showed paraphyletic grouping in the tree, being clustered once in the ancestral clade with *C. elegans* and once with the other invertebrates including the cnidarians.

Similar to those already hypothesized by Brew et al. (2010), the scattered distribution of single- and double-domains homologs supports the hypothesis that single-domain protein cannot be considered as a model for the ancestral protein.

In this scenario, it should be noted that a single-domain TIMP (HmTIMPa) is included in the most recent cnidarian branch (fig. 5). Such evidence, in association with the position in the ancestral clade of the two-domain homologs from cnidarians, seems to corroborate the above-mentioned hypothesis.

The topology of the tree and the phylogenetic relationships, as herein described, provides additional data on cnidarians evolutionary status confirming their ancestral molecular complexity as previously demonstrated (Technau et al. 2005; Nicosia et al. 2014). This is consistent with the presence of genes responsible for the development of bilaterian traits, such as mesoderm and bilaterality, in their genomes (Finnerty et al. 2004; Fritzenwanker et al. 2004; Martindale et al. 2004).

Because biophysical and structural considerations have to be taken into account to probe the evolution of protein families (Fitch and Markowitz 1970; Fletcher and Yang 2009), computational approaches were herein used to integrate structural consideration with phylogenetic analysis.

As described for several protein families, including insulins (Blundell and Wood 1975) and aspartic proteinases (Tang et al. 1978), amino acid substitutions were accepted during evolution in a way that satisfied restraints arising from structure and function. In cnidarian TIMPs, amino acid residues were substituted, so that both SSEs and 3D structures were usually maintained. Thus, the 3D organizations of cnidarian TIMPs appeared grossly similar as the overall structures fully resembled the OB-fold category.

However, some changes were observed especially for specific positions located in the N-domain of proteins. In particular, short stretches of amino acid residues organizing α-helices, herein located, were identified. Analysis of complex data set of proteins have revealed significant differences in the occurrence of SSEs among conserved and random regions (Han and Baker 1996; Mizuguchi and Blundell 2000); as helices were found underrepresented in conserved regions (Sitbon and Pietrokovski 2007).

It is also well accepted that exposure to the solvent is anticorrelated with conservation (Overington et al. 1992; Goldman et al. 1998; Bloom et al. 2006; Franzosa and Xia 2009). Thus, models of linear relationship between evolutionary rate and RSA have been recently proposed (Ramsey et al. 2011).

In our study, all the identified α-helices were found to possess a state of solvent-exposed as we computed an average accessibility of more than 60% per element. Hence, these residues are likely to localize on the surface of the proteins.

In the light of relationship between the evolutionary rate of a region and the RSA, these solvent-exposed sites may likely provide tolerances to amino acid substitutions, which in turn allowed variability to accumulate without functional alteration in protein structures.

To evaluate the presence of these features in organisms, diverging at different time scales from basal metazoans to the Protostomes and Deuterostomes lineages, a phylogenetic-based vertical approach was herein used.

Despite considerable sequence differences, a canonical fold was maintained over 600 Myr in evolutionarily distant TIMPs as indicated by RMSD pairwise comparisons. Noteworthy similar RMSD values were shared by highly diverging sequences; conversely, highest sequence identities are not related with lowest RMSDs. Although the number and length of beta strands generating the OB-fold of the N-terminal domain were different among homologs, generally the N-domains were found to be spatially conserved retaining the ability to fold as a wedge.

These results suggest the hypothesis of an evolutionary conservation of a prototypical structure, which shows significant structural similarity to the human TIMPs. However, small changes, not surprisingly organizing exposed helices, occasionally appeared.

Both the occurrence of such SSE and the RSA of the corresponding residues in the 3D structures of TIMPs raise the possibility that these sites were likely prone to accept variations without affect the overall protein functions. Moreover, the scattered distribution of these elements across phyla and species results in the inability to consider this helix as a signature of protein evolution. This resembles the unreliability to consider the single-domain TIMP as an ancestor.

Thus, it could be hypothesized that the molecular evolution of TIMPs in invertebrates and vertebrates cannot be merely explained as a linear process from a common ancestor to the derived one. It is reasonable to suppose that single- and double-domain proteins may belong to distinct gene line. Evolutionary forces, through different pathways, may have acted to create the high degree of variability in amino acid sequences and domain organization.

However, a unique gene line could also be hypothesized. In this scenario, the diversity of TIMPs could arise from mechanisms of protein domain loss from a two-domain ancestor.

Clear example of such assumption can be found in the repertoire of expressed TIMPs from hydra. In the light of sequence similarity and protein architecture (namely single- and double-domain organization), it is reasonable to hypothesize that TIMP3a and 3b resulted from duplication events. Similarly, the single-domain TIMPa may have arisen from the same duplication followed by loss of the C-term domain, whereas TIMPb, probably belonging to a different gene line, may have undergone convergent evolution.

Furthermore, while the evolution of TIMPs across early diverged metazoans (including Protostomes and Echinoderms other than Cnidaria) seems to proceed according to taxa distinction, homologs from Vertebrata evolved differently. Indeed, these TIMPs have been shown to branch-off according to polyphyletic groups associated with signature of convergent evolution. Among the vertebrates, TIMP-1 appears the closest to the ancestral form, whereas TIMP-4 seems to represent the most modern one.

Similar evolutionary relationships were reported by Brew and Nagase (2010), which argued that mammalian TIMPs represent products of gene duplication. Mechanisms leading to gene duplication are considered to be essential for the development of novel genes (Lynch and Conery 2000; Demuth et al. 2006; Bergthorsson et al. 2007), and the presence of duplicated genes involved in physiological processes underlines its pivotal role (De Grassi et al. 2008; Dong et al. 2009; Pavlopoulou et al. 2010). We argued that successive rounds of gene duplications, followed by deletions, substitution, and domain loss have likely given rise to the TIMPs found in the extant genomes.

These mechanisms may likely have generated the diverse array of functions exhibit by TIMPs.

Additionally, it can be supposed that evolutionary relationships among Metazoa TIMPs are made more complex by the probable nested status of TIMP gene(s) within an intron of synapsin genes, as already demonstrated in *Homo sapiens* and *D. melanogaster* (Pohar et al. 1999).
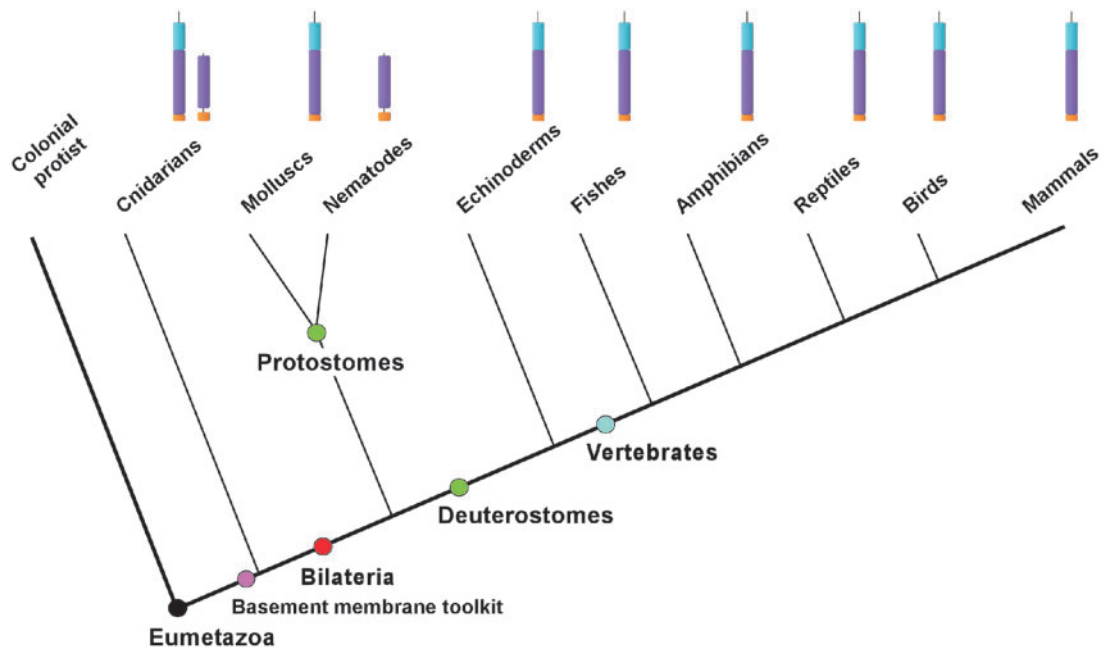
Fig. 9.—Eumetazoa phylogeny and elaboration of single- and double-domain TIMP proteins. The diagram represents relationships, and the branch lengths are not intended for evolutionary distances. The main taxonomic divisions are indicated as well as the appearance of basement membrane. Single- and double-domain TIMPs are showed beside the corresponding species. Number of TIMPs per species are not represented. The domains color key is according to figure 1.

More knowledge about the nesting status of TIMP genes in the other Metazoans could be useful to better understand the evolutionary process along the tree of life.

In conclusion, the evolution of TIMP proteins, as described and analyzed in this work, appears to be dynamically represented by a set of complex processes of amino acid residues rearrangement. However, evidence for the maintenance of a common 3D structure, from the cnidarians to humans, emerged.

The evolutionary conservation of the structure is coherent with the pivotal role of the TIMP functions in the cellular process from their metalloproteinase inhibitory activity, to the promotion of cell proliferation, antiangiogenic, pro- and antiapoptotic, and synaptic plasticity activities. Additionally, data herein presented strongly suggest that comparative analyses on protein family should be integrated with other studies including homology modeling analyses, to encompass structure, function, and evolution of these members.

Cnidarians occupy a crucial position in the evolution, and recently, it has been suggested that despite diploblastic organization, they are reduced Mesodermata representing an important step in the early evolution of mesoderm and in the tissue organization (Seipel and Schmid 2005). It has been also argued that the origin of the C-terminal domain of mammalian TIMPs and the moment at which double-domain proteins appeared in the evolution remains still uncertain (Brew et al. 2000).

Thus, it is particularly striking that the repertoire of the cnidarians TIMPs, including the ancestral stony coral, displayed homologs with the double-domain organization.

Intriguingly, it seems that in the Deuterostome lineage (fig. 9) the single-domain TIMPs disappeared, whereas proteins, including complement proteins as C3/C4/C5 (Bramham et al. 2005), sharing the same fold, are widely distributed. Thus, it could be hypothesized that mechanisms of neofunctionalization (Ohno 1970; Kimura and Ohta 1974), in which the duplication of a gene precedes the emergence of a new function, may have likely acted providing a diverse array of functions to such ancestral organization.

Supplementary Material

Supplementary tables S1–S5 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgment

Literature Cited

Altschul S et al. 1997. BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Ball EE, Hayward DC, Saint R, Miller DJ. 2004. A simple plan-cnidarians and the origins of developmental mechanisms. Nat Rev Genet. 5:567–577.

Bergthorsson U, Andersson DI, Roth JR. 2007. Ohno's dilemma: evolution of new genes under continuous selection. Proc Natl Acad Sci U S A. 104:17004–17009.

Bloom JD, Drummond DA, Arnold FH, Wilke CO, 2006. Structural determinants of the rate of protein evolution in yeast Mol Biol and Evol. 23:1751–1761.

Blundell TL, Wood SP. 1975. Is the evolution of insulin Darwinian or due to selectively neutral mutation? Nature 257:197–203.

Bode HR, Bode PM. 1984. Patterning in hydra. In: Malacinski GM, Bryant SV editors. Pattern formation: a primer in developmental biology. Vol. 1. New York: Macmillan. p. 213–241.

Bosch TCG. 1998. Hydra In: Ferretti P, Geraudie J editors. Cellular and molecular basis of regeneration: from invertebrates to humans. New York: Wiley. p. 111–134

Bramham J et al. 2005. Functional insights from the structure of the multi-functional C345C domain of C5 of complement. J Biol Chem. 280:10636–10645.

Brew K, Dinakarpandian D, Nagase H. 2000. Tissue inhibitors of metallo-proteinases: evolution, structure and function. Biochim Biophys Acta Prot Struct Mol Enzymol. 1477:267–283.

Brew K, Nagase H. 2010. The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. Biochim Biophys Acta. 1803:55–71.

Bridge D, Cunningham CW, Schierwater B, DeSalle R, Buss LW. 1992. Class level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. Proc Natl Acad Sci U S A. 89:8750–8753.

Chapman JA et al. 2010. The dynamic genome of Hydra. Nature 464(7288):592–596.

Collins AG. 2002. Phylogeny of Medusozoa and the evolution of cnidarian life cycles. J Evol Biol. 15:418–432.

Császár NBM, Seneca FO, van Oppen MJH. 2009. Variation in antioxidant gene expression in the scleractinian coral Acropora millepora under laboratory thermal stress. Mar Ecol Prog Ser. 392:93–102.

D'Alessio S, et al. 2008. Tissue inhibitor of metalloproteinases-2 binding to membrane-type 1 matrix metalloproteinase induces MAPK activation and cell growth by a nonproteolytic mechanism. J Biol Chem. 283(1):87e99.

De Grassi A, Lanave C, Saccone C. 2008. Genome duplication and gene family evolution: the case of three OXPHOS gene families. Gene 421:1–6.

Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. PLoS One 1:e85.

Docherty AJ et al. 1985. Sequence of human tissue inhibitor of metallo-proteinases and its identity to erythroid-potentiating activity. Nature 318:66–69.

Dong D, Jones G, Zhang S. 2009. Dynamic evolution of bitter taste receptor genes in vertebrates. BMC Evol Biol. 9:12.

Finnerty JR, Pang P, Burton K, Paulson D, Martindale MQ. 2004. Homology of bilateral symmetry in Cnidaria and Bilateria: axial expression of Hox genes and Dpp in the sea anemone Nematostella. Science 304:1335–1337.

Fitch WM, Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem Genet. 4:579–593.

Fletcher W, Yang Z. 2009. A Flexible Simulator of Biological Sequence Evolution. Mol Biol Evol. 26:1879–1888.

Flores TP, Orengo CA, Moss DS, Thornton JM. 1993. Comparison of conformational characteristics in structurally similar protein pairs. Protein Sci. 2:1811–1826.

Franzosa E, Xia Y. 2009. Structural perspectives on protein evolution. Ann Rep Comput Chem. 4:3–21.

Fritzenwanker JH, Saina M, Technau U. 2004. Analysis of forkhead and snail expression reveals epithelial–mesenchymal transitions during embryonic and larval development of Nematostella vectensis. Dev Biol. 275:389–402.

Fujisawa T. 2003. Hydra regeneration and epitheliopeptides. Dev Dyn. 226(2):182–189.

Ganot P et al. 2011. Adaptations to endosymbiosis in a cnidarian-dinoflagellate association: differential gene expression and specific gene duplications. PLoS Genet. 7(7):e1002187.

Gasson JC et al. 1995. Molecular characterization and expression of the gene encoding human erythroid potentiating activity. Nature 315:768–771.

Goldman N, Thorne JL, Jones DT. 1998. Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genetics 149:445–458.

Gong S et al. 2009. Structural and functional restraints in the evolution of protein families and superfamilies. Biochem Soc Trans. 4:727–733.

Greene J et al. 1996. Molecular cloning and characterization of human tissue inhibitor of metalloproteinase 4. J Biol Chem. 271:30375–30380.

Han KF, Baker D. 1996. Global properties of the mapping between local amino acid sequence and local structure in proteins. Proc Natl Acad Sci U S A. 93:5814–5818.

Jung KK, Liu XW, Chirco R, Fridman R, Kim HR. 2006. Identification of CD63 as a tissue inhibitor of metalloproteinase-1 interacting cell surface protein. Embo J25(17):3934e42.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 12:2577–637.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 10(6):845–858.

Kimura M, Ohta T. 1974. On some principles governing molecular evolution. Proc Natl Acad Sci U S A. 71:2848–2852.

Kortschak RD, Samuel G, Saint R, Miller DJ. 2003. EST analysis of the cnidarian Acropora millepora reveals extensive gene loss and rapid sequence divergence in the model invertebrates. Curr Biol. 13:2190–2195.

Krogh A, Brown M, Mian I, Sjolander K, Haussler D. 1994. Hidden Markov models in computational biology. Applications to protein modeling. J Mol Biol. 235:1501–1531.

Kubota S et al. 2003. Occurrence of two distinct types of tissue inhibitor of metalloproteinases-2 in teleost fish. Biochim Biophys Acta. 1629(1-3):102–108.

Leontovich AA, Zhang J, Shimokawa K, Nagase H, Sarras MP. Jr. 2000. A novel hydra matrix metalloproteinase (HMMP) functions in extracellular matrix degradation, morphogenesis and the maintenance of differentiated cells in the foot process. Development 127(4):907–920.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.

Maiti R, Van Domselaar GH, Zhang H, Wishart DS. 2004. SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res. 32:W590–W594.

Martindale MQ, Finnerty JR, Henry JQ. 2002. The Radiata and the evolutionary origins of the bilaterian body plan. Mol Phylogenet Evol. 24:358–365.

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.

Meyer E et al. 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. BMC Genom. 10:219.

Mizuguchi K, Blundell T. 2000. Analysis of conservation and substitutions of secondary structure elements within protein superfamilies. Bioinformatics 16:1111–1119.

Montagnani C, Avarre JC, de Lorgeril J, Quiquand M, Boulo V, Escoubas JM. 2007. First evidence of the activation of Cg-timp, an immune

response component of Pacific oysters, through a damage-associated molecular pattern pathway. Dev Comp Immunol. 31(1):1–11.

Montagnani C, Tirape A, Boulo V, Escoubas JM. 2005. The two Cg-timp mRNAs expressed in oyster hemocytes are generated by two gene families and differentially expressed during ontogenesis. Dev Comp Immunol. 29:831–839.

Morris SC. 2006. Darwin's dilemma: the realities of the Cambrian 'explosion'. Philos Trans R Soc B. 361:1069–1083.

Moya A, Ganot P, Furla P, Sabourault C. 2012. The transcriptomic response to thermal stress is immediate, transient and potentiated by ultraviolet radiation in the sea anemone Anemonia viridis. Mol Ecol. 21(5):1158–1174.

Murphy G. 2011. Tissue inhibitors of metalloproteinases. Genome Biol. 12:233.

Murphy G, Nagase H. 2008. Progress in matrix metalloproteinase research. Mol Aspects Med. 29:290–308.

Nancy CM et al. 1998. Glycosylation and NH2-terminal domain mutants of the tissue inhibitor of metalloproteinases-1 (TIMP-1). Biochim Biophys Acta: Protein Struct Mol Enzymol. 1388(1):21–34.

Nicosia A et al. 2014. Characterization of small HSPs from Anemonia viridis reveals insights into molecular evolution of alpha crystallin genes among cnidarians. PLoS One 9(9):e105908.

Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302(1):205–217.

Ohno S. 1970. Evolution by gene duplication. New York: Springer.

Okamoto T et al. 1977. Activation of human matrix metalloproteinases by various bacterial proteinases. J Biol Chem. 272(9):6059e66.

Overington J, Donnelly D, Johnson MS, Sali A, Blundell TL. 1992. Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. Protein Sci. 1:216–226.

Pavloff N, Staskus PW, Kishnani NS, Hawkes SP. 1992. A new inhibitor of metalloproteinases from chicken:ChIMP-3. A third member of the TIMP family. J Biol Chem. 267:17321–17326.

Pavlopoulou A, Pampalakis G, Michalopoulos I, Sotiropoulou G. 2010. Evolutionary history of tissue kallikreins. PLoS One 5:e13781.

Pettersen EF et al. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem. 5(13):1605–1612.

Pohar N, Godenschwege TA, Buchner E. 1999. Invertebrate tissue inhibitor of metalloproteinase: structure and nested gene organization within the synapsin locus is conserved from Drosophila to human. Genomics 57(2):293e6.

Putnam NH et al. 2007. Repertoire and genomic organization sea anemone genome reveals ancestral eumetazoan gene. Science 317:86–94.

Ramsey DC, Scherrer MP, Zhou T, Wilke CO. 2011. The relationship between relative solvent accessibility and evolutionary rate in protein evolution. Genetics 188:479–488.

Reitzel AM, Sullivan JC, Traylor-Knowles N, Finnerty JR. 2008. Genomic survey of candidate stress-response genes in the estuarine anemone Nematostella vectensis. Biol Bull. 214:233–254.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42:W320–W324.

Sabourault C, Ganot P, Deleury E, Allemand D, Furla P. 2009. Comprehensive EST analysis of the symbiotic sea anemone, Anemonia viridis. BMC Genomics 23(10):333.

Salamone M et al. 2015. Comprehensive analysis of a Vibrio parahaemolyticus strain extracellular serine protease VpSP37. PLoS One 10(7):e0126349.

Salemme FR, Miller MD, Seven RJ. 1977. Structural convergence during protein evolution. Proc Natl Acad Sci U S A. 74(7):2820–2824.

Sarras MP. Jr. 2012. Components, structure, biogenesis and function of the Hydra extracellular matrix in regeneration, pattern formation and cell differentiation. Int J Dev Biol. 56(6-8):567–576.

Sarras MP, Deutzmann R. 2001. Hydra and Niccolo Paganini (1782–1840)—two peas in a pod? The molecular basis of extracellular matrix structure in the invertebrate, Hydra. Bioessays 23: 716–724.

Seipel K, Schmid V. 2005. Evolution of striated muscle: jellyfish and the origin of triploblasty. Dev Biol. 282:14–26.

Seo DW, et al. 2003. TIMP-2 mediated inhibition of angiogenesis: an MMP-independent mechanism. Cell 114(2):171e80.

Sitbon E, Pietrokovski S. 2007. Occurrence of protein structure elements in conserved sequence regions. BMC Struct Biol. 7:3–18.

Shimizu H, Aufschnaiter R, Li L., Sarras MP Jr, Borza DB, Abrahamson DR, Sado Y, Zhang X. 2008. The extracellular matrix of hydra is a porous sheet and contains type IV collagen. Zoology (Jena). 111(5):410–418.

Steele RE, David CN, Technau U. 2011. A genomic view of 500 million years of cnidarian evolution. Trend Genet. 27:7–13.

Stetler-Stevenson WG, Brown PD, Onisto M, Levy AT, Liotta LA. 1990. Tissue inhibitor of metalloproteinases-2 (TIMP-2) mRNA expression in tumor cell lines and human tumor tissues. J Biol Chem. 265:13933–13938.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 30:2725–2729.

Tang J, James MN, Hsu IN, Jenkins JA, Blundell TL. 1978. Structural evidence for gene duplication in the evolution of the acid proteases. Nature 271:618–621.

Technau U, Steele RE. 2011. Evolutionary crossroads in developmental biology: Cnidaria. Development 138(8):1447–1458.

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

Wang Q, Bao Y, Huo L, Gu H, Lin Z. 2012. A novel tissue inhibitor of metalloproteinase in blood clam Tegillarca granosa: molecular cloning, tissue distribution and expression analysis. Fish Shellfish Immunol. 33(3):645–651.

Williamson RA et al. 1994. Solution structure of the active domain of tissue inhibitor ofrnetalloproteinases-2. A new member of the OB fold protein family. Biochemistry 33:11745–11759.

Xu XY, Shen YB, Yang XM, Li JL. 2011. Cloning and characterization of TIMP-2b gene in grass carp. Comp Biochem Physiol B Biochem Mol Biol. 159(2):115–121.

Yan L, Leontovich A, Fei K, Sarras MP Jr. 2000. Hydra metalloproteinase 1: a secreted astacin metalloproteinase whose apical axis expression is differentially regulated during head regeneration. Dev Biol. 219(1):115.

Yang Y, et al. 2003. EST analysis of gene expression in the tentacle of Cyanea capillata. FEBS Lett. 538:183–191.

Zhang J, Bai S, Tanase C, Nagase H, Sarras MP. 2003. The expression of tissue inhibitor of metalloproteinase 2 (TIMP-2) is required for normal development of zebrafish embryos. Dev Genes Evol. 213:382–389.

Associate editor: Maria Costantini