



## UvA-DARE (Digital Academic Repository)

### The biocalcification of mollusk shells and coral skeletons: Integrating molecular, proteomics and bioinformatics methods

Sequeira dos Ramos Silva, P.

**Publication date**  
2013

[Link to publication](#)

#### **Citation for published version (APA):**

Sequeira dos Ramos Silva, P. (2013). *The biocalcification of mollusk shells and coral skeletons: Integrating molecular, proteomics and bioinformatics methods*.

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## **Appendix B**

The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling

- Supplementary data -



## Appendix B

		ATLDDKRYFVDYQTVSKRGNGTICAYSLPSGSR <b>RWVQCCYTDLPSSGGKVL</b> SSSSPPESGG PYLIALPGSPVIVSDADGHEFCSSSQSLYYRLRPP <b>RSFCGYTLRR</b> RGLIFGDPHFHTLDD NTTYTFNGLGEYTVIADDEAFEMQARTARTSGRGLGTAFSAAVAKERGTATVEARINQK AGDLEVL IAGKPFNISTITTTGNTIPDGNITLVRDNGSITALFPSPNIAFTFTDVEGTLA IAFEAPDDFKNRTKGLLGTWDDPSDDFVTPDGLTLPADAAPRRIHYEFGKWKQINASQS LFTYSDLESPTFVDSL YIPMFDNITWVNDSPRYEAVKACGNNTQCLFDAAVTEDTSYG INTKLEDNNNEINKELANFPKILGPKVINATIGQAEVKITAEHNSDDFFVTVNNLP DVII LANTSRVLLIRWTPSLQKVEPVFIVTDSHNSSELRLPILLCCPCANGSRCIDDEE VSNQRNKGSEFLLSCTCPAGLTGOYCOHKIDACVENNQPCFPGVKCTDVSSSSNGTRYQ CDPCPKYSGNGSICEDIDECSDANVSKCDHSCINLPGSYVDCNQGSLEGGDTSCCKDI NECLISNDQMONTNLPGGRTCSCLDGFQIDPKDQACVPI SRCDTFKVGCCQVCMVDRG QPKCACHKGYSLNADGRTCDDINECTTHRHKCSQICHNLGDSYTCSCQPGFNLSPDQTTTC EDIDECGLINEAHCEGSLEICINTMGSFRCECQDGFHRVNDTCQESLPSTNGPTGTGIV ASSVSIALTIKDADLHEWQARLSRMFMDAVAKVVVDYCKGNANGNCYGNAVI AKRYTRSI SGTSLVARVHILNDFPETRDANLLVAFYVMLSTNQGEVYVMNKDSLRLALQESQTELSWA IKKEISEIRALKVDDSEPTPYETREDGLEMIWLLVGVSVAVAVPLMIVIVILYREYRRIA KORRKTNNFDLRQWGSARETIYSGFTNSKSARL				
Amil-SAP1	B3EWZ1	>Secreted acidic protein 1; JT018094 MARDLLL VVFFACLLQSFWGLPLPL <b>KNENAIVDGDGTSVVTTKEDASTIFERD</b> ENPANQV SAMVTGVILDENGDPGESDESVENVDNDGEGGKDDDKNGEDNDLNDKHEEHEEKGGDDRG DDEEEDDAEGDNDSDNEGDDDDDDSDDDDDVDESGADEDDDDSDSG	59.2	0.71 - ASM 0.60 - AIM	341 343	5
Amil-SAP1	B3EWZ0	>Secreted acidic protein 1; JT006291 SDDESGDDENGDDKGGDDDDKEEDGEDVSEDE <b>KADVGGDDGDDDDNETGGSDTNDVDY</b> <b>GDGNDAREIGDHSIQDIRDLILDAIHNKDGEMDANPLQNLVPGPKLKELYLRS</b> GGG HFKGQLLNI TLGLGFCILFLLL*	16.7	0.43 - ASM	95	-
USOMP-8	B3EWZ2	>USOMP-8; JT014391 MVTPHGILLTTTAAASLLWITFAEITIPND <b>AKSFENFLKE</b> HGPGKPGPLGYPNISYMAF TREEAENFPNLVSVHTRMKRIKQNSTIPDKYVILGIQAPNDTQEONSTRNK <b>RDSSESYTA</b> <b>TTQSGTCTSTIGGLQRLCEVCPARTDLGPDITPRFINEVLCDVPLGDCGVGVGGKCRSA</b> <b>SVFQDFLRFSSSDSNLEVYSGEIRVCCCECALALS*</b>	47.7	0.23 - ASM 0.36 - AIM	333 674	-
Coadhesin	B3EWZ3	>Coadhesin; JT016638 QGNYSYGGTTPGTPIGCTNLITLSNVKFFASSSDGPDIPVLNSTDYWCSEFNWKNQSL TVDLGFVTFP <b>RLLVQGEFFTSRSV</b> SEYFVLTSIDGINYTYILGTNGQSMKVFVGLFNGD QTRDNLNLTAPVQARYVQFNPOEPMIAEDDSICMRVGVESCOLVPAAVNGAWSHSPYGPC THACLGTAKRT <b>RTCADPAPVFGGSPCEGVNEEIKI</b> CNDCVGTVNGGSPWGLWSRCSSTTC NPGQSRQRTCTNPSFKNGGDCSGPSTQSEPCQVQFCPVDGWSAWSGLSRCTRACGGG RQYQSRCTSNFPFHGGGRDCVGRSLSTPNTCQCPVHGGSPWGSFSSCTRTCGGGQKS RTRVCNSPAPSCNGITCPGGNQDIQPCNQQTCTPSTSPFINGNYSNWGTACSVTCG QGTRETRLCDNPAPAQGGGSCQGGPSSSELVGTETIPCPVNGNWSWGDWNSCSSSGCGPK SYRYRDCDNPANGLNCTGPQDES KDCNSTACPVGGWSAWSSTPCSATCGGQTLKRT REC�NFKPQYGGASCFGNETEQEVACNKGPCPTSPPTISPPTTGSPADSNIPEDLDLVFAV SATSSN <b>RLATYNSMRD</b> TIN <b>RFI</b> TTYGSNKVHYSIIVYGKAVQVVISFNHTFPPSVGELQE	19.4	0.18 - ASM 0.37 - AIM	271 733	

## Appendix B

		<p><b>AISRHPAISGPTVLKNALETQTFIQEIPSRPNAKVLVVFSDNSPSDGNLVQAVRPLE</b>  <b>NNKILVVSVGVDVNRTELLTISPNDLVLSVQPTAGPGALSKRIMDRILRRDIPLIDIG</b>  <b>FALSATSSDFQDIFVKMKNVIRTIVERYGVERVKFSLIVYQNVTTVLGDFNRNLQTADDL</b>  <b>VYVNNLQVFPQNKNDLSALLEAESLFRQRARPNPKVVFVVLTDGVSTLSNANSLINTA</b>  <b>ELRKSDDLILSVGFGSQTNQVGNQMNSVVFAPRDYIAVPNYPARDDVIAETIMFKALEV</b>  <b>NLPLIDLTFALSSSSILSQETFKLMKETVQSLVHTYGDRIHYGVIVFGSVATRSDFAT</b>  <b>NFPDQNELIRKVSQLTRSGGSPDLVAALKKARKVFLKKEVRPYARKVLVVMIDDESSANK</b>  <b>NDLNEVRALRNRSLVIVGVIGTQTLPKDLGIIITDDKRNFLKAGINKNRDELAREIISI</b>  <b>ILRPSGLSKWSSWSACSRTCRYLKGAGTQIRTRDCKIPELGCDDMRI DTVECNMDCGEC</b>  <b>GQRGPLNESAYTASSNSESAPFLAALNTSDPTAWCLINNENGGYVQLDLGELTRVYKVAT</b>  <b>KGEQQGDRWVTSY YLTSEDEGTFFFDYKAAQRLSGNTDSTSVAFNVVNTTRPYRYVRFHP</b>  <b>VNFKGEPQAAVFGCNEEKILPPPETIADQADAARGILIVLWILAGILTFLLLMACCYY</b>  <b>CCWHVCCGRGKRRKGLVYRERSIEDDGYLINDEKRWTLGSAPMTPVPRVREDEIQEVTIE</b>  <b>MKEDNEQPLGVIQFGIETDETEKHKVTAEDVKSEKPKYSEASSTIKSGSTMRRMKAND</b>  <b>GSDRRKRTKSEGDALDAVDGLDWSYLSDEQTAFTNEAFVKSQEQFLEPPGSAFRGNK</b>  <b>VDMRRSLSADELATLDYDLFEDRQGPLHTATLGRDGYMRMHKANQGSLLPSPDGGREMGTV</b>  <b>DVAIGGIRVPSPKDDPIYDTAGQEIHLAVEQAGRSVPLEDDGGYRGEWYSRWG*</b></p>				57
Amil-SAP2	B3EWZ4	<p>&gt;Secreted acidic protein 2; JR983041  <b>WSXSGDDDDDDGSDGDDDDDDGDDSDDDNDADDSDGAEDNDDSDGENEDTDDSDGD</b>  <b>WKMIKPTTVMTGWMMTIADSSDDDNERRDTSDDSVGDDAYNDDSDAGELNSDSTYDQL</b>  <b>RSQGDVQSQGFKNLQSYSNQFKVSSGLVATVVSTLACLFLTLNH*</b></p>	28.5	0.59 - ASM 0.08 - AIM	245 88	-
MAM and LDL-receptor domain-containing protein 1		<p>&gt;MAM and LDL-receptor domain-containing protein 1; JR994474  <b>K.VGLTYTR.L, K.VVFEGR.G, R.AEIALMSR.V</b>  <b>R.LGQVAVSSR.A, R.GDIAIDDLK.L</b> <b>R.SDDNFDWR.L,</b>  <b>R.SVGSNLVYIK.K, K.YQVFEGR.G, K.VYQVFEGR.G</b>  <b>R.VPFQVIESVR.G, R.SYTGDIADDDV.K, R.APFQVVFQGR.G</b>  <b>R.IESVTIPATQK.C, R.TPFTIEFEALR.G, R.FTSQFSPVSVR.G</b>  <b>K.AQLLSPSPSTSGK.C, R.SGSQFQVFEGR.G, R.SANVYQVFEGR.G</b>  <b>R.LMSEDFNPTSSGR.C, K.IMSGSCPAPGDCSFEK.G</b>  <b>K.VPVSNLNAYQVFEGR.G, R.QSGGSPSIGTPTSDHTTGLR.G,</b>  <b>R.QSGATSSSGTGPTFDHTLGTAR.G, R.FAQVNLNQPFFVFEGR.G</b></p>	6	0.07 - ASM 0.13 - AIM	198 331	-
MAM and LDL-receptor domain-containing protein 2	B3EWZ6	<p>&gt;MAM and LDL-receptor domain-containing protein 2; JT011118  <b>K.VGLTYTR.L, K.VVFEGR.G, R.VPIVSGNR.Y, R.LGQVAVSSR.A,</b>  <b>R.GDIAIDDLK.L, R.SVGSNLVYIK.K, K.YQVFEGR.G, K.VYQVFEGR.G</b>  <b>R.QFVSVFEAIR.G, K.FVDCALPPVAR.S, R.YYQIILEGR.G,</b>  <b>R.VPFQVIESVR.G, R.VPFQVIESVR.G, R.SYTGDIADDDV.K, I,</b>  <b>R.APFQVVFQGR.G, R.IESVTIPATQK.C, R.TPFTIEFEALR.G</b>  <b>R.SNAFQIIFLGR.G, K.AQLLSPSPSTSGK.C, R.SGSQFQVFEGR.G</b>  <b>R.SANVYQVFEGR.G, R.LMSEDFNPTSSGR.C, K.IMSGSCPAPGDCSFEK.G</b>  <b>R.QSGGSPSIGTPTSDHTTGLR.G, R.QSGATSSSGTGPTFDHTLGTAR.G</b>  <b>R.FAQVNLNQPFFVFEGR.G</b></p>	3.7	0.04 - ASM 0.02 - AIM	182 285	-
Thr-rich	B3EWZ7	>Threonine-rich protein; JT013896	17	0.18 - ASM	131	

## Appendix B

protein		MKAFLSLATLLACIVLTESAPHSADV <b>REEAFDALVRSYLQAVQRD</b> SHMENLTCAECQGV TERNCTLG <b>RQVCNPGVEVCTTLEAFNLDGTGTTVTRG</b> CFNI TGLNCGDNPGCGALNTTG NIQSCDQFCNNTSLCNAGTLT <del>TVTTPQTTDGN</del> TTEAPTSTEPPTNASTEAPTSTEPPTNA STEAPTSSTEPPTNASTEAPT <del>TTEAPT</del> TTEAPT <del>TTEAPT</del> TTEAPT <del>TTEPT</del> TTEPT <del>TTEPT</del> TTEPT <del>TTEPT</del> PTTAAPTTTPAPTTPAPTTPFCNATLAGLSGFTSPNFQLITQTG		0.18 - AIM	111	-
Ectin	B3EWZ8	>Ectin; JR978035 MMQASFSICLLSFYLLSFCHGAPLPAFLRSVLSGNGMKEESRVLKRSAPVMQDEIPVCAQ NQTDRYSSSSRLCRLVKDLGFCDFDLYQVLOSQPCIGCGFRVEDGNWSVWGAWSPCSA TCGDGQRRSRSCNPPSPGGGADCLGVSQEI EDCNRRSCEGIGGWSNWGWSACSSECN IGIQARTRCTNPPPTIPEGACEGFSFETQICSTSGCNVSASVSTAAATTSVSSSTAQTQ IGPTVVSLTAKQQAQLDAHNAKRAIHGSPPLEWDFTLAMNADEWANELAVT <b>RQLEHDPNI</b> <b>MNEGENLFKSAGALECVDVERWFLEKDYDYEDDNKLLDDTSNFTQLVWRN</b> TTRVGVAT VVEVSEGSVETIVARYTPPNIEGKFEENVIKPSAEAL	14.2	0.09 - ASM 0.14 - AIM	113 98	-
Hephaestin-like	B3EWZ9	>Hephaestin-like; JT019463 MMDRSNAAFVLTA <del>CFIFSOLICHVAAIT</del> <b>RTYYIAAVEKEWDYAPSGYNKIKGVKLEDDSD</b> <b>ATVFA</b> TGKAHRI <del>GRIYDKVL</del> <b>REYEDASFTKEK</b> PH <b>KYLGLGPI</b> L <b>KGE</b> IGDPTIVVHFKN NGSRVYSMHPHGVFYSKDEGALYEDNTKGGFKDDKVPVGGTHTYSWHLTQSHAPADQE DKCITWIYHSHVVPKDI <del>NTGLLGI</del> MLICRKGALNQGQSGVDKEFVALFTVLDENESWL LSKNIERCSDPTRVNP <del>DD</del> EDFKESNMKHAINGYFVGNLPGLD <del>MCY</del> GD <del>SVK</del> WHLAGIGNEV DIHTAYFHGQSFTIDGHR <b>KDVASLLPATFTV</b> AS <b>MKAL</b> NPGKMWLNLVNDHYNAGMYTLF NVTKCPGKVGVPAPSVSGGKRTYFIAANEVEWNYGPTGVNGMDGQSLIAPGSDSAVFFAQ NAQRIGGTYL <b>KAIYEQYTDARF</b> STKVPKPEHLGFLGPV <b>IAE</b> VND <b>IEVVFKN</b> NARFNFS IQPHGVFNKSNEGALYEDGTSRAQ <b>KADDNVPGQFTTYRWTVP</b> EEV <b>GP</b> TKSDAACITWV YHSSVDPVKDTYSGLFEP <del>LLTCKKGT</del> LNNDNTRKDTKEFVLLFTVTDESESWYHEKNKE MKANAILLINDDDEDYKESNMKMGINGFLYANLPGLEMCLGDTISWHVIGLGNVDMHTAY FYGNFTFHQGSVKDTVSL <del>LLPGV</del> GT <del>LTMT</del> PDNAGD <del>WALV</del> CR <del>TND</del> HYSAGM <del>QAKYK</del> VNTCN RNPEL <del>KTS</del> GKTRDYIIA <del>AFEM</del> EW <del>DYAPT</del> GLDALDGGKLDQSEAK <b>VFTVTS</b> <b>DKRI</b> GRKYV KAVY <b>REFTNDQFNQQL</b> RTPAEHLGILGPMLHAEVGD <del>TIK</del> VFKNNANRNYSVHPHGLY YSKAHEGSDYNDGTS <del>GADK</del> LDNAIQPG <b>KTYTYIWKV</b> PERAGPGKDGACATWAYYS <del>DVNP</del> <b>IKD</b> NS <b>GLIGPLI</b> <b>ICK</b> GKLGTEERSD <del>VDF</del> FLMFTVLDENESWYLDENIKKYCNP GDKETL <del>KADDD</del> FMESNMKGINGFVFNGLKGLKMYQDEKVDWLLLGINEVDMHTVHFHG QSFLRKQVSYHREDVYDLFPGVFATVEMVPDSTGDWLLHCHVNDHMVAGMETLYSVLDKS LKTTPKPI <del>TAA</del> SS <del>FVT</del> SSIFLYLSPFVLAMLLKA*	15.2	0.11 - ASM 1.08 - AIM	101 180	75
USOMP-1	B3EX00	>Uncharacterized skeletal organic matrix protein-1; JT021412 KSNGMVSEGHAYFSQQLNFETPIRTENGTETISMIKMTVKSRVLLXGTVALLIYSPESIDF QGLFVKLFLSKPSPVLSLNETDAGQSLNDTNE <del>DP</del> APLSR <del>SRRA</del> VNSXNANASLVS EILERIGPVCLFPDRQPOLYSLNVNSVNL <del>TL</del> SASVSQIDGPH <del>TSR</del> IDVSLVSVGQNL <del>T</del> SVVIQKFVRMVSLELS <del>SVN</del> L <del>FPPI</del> FRFLRG <del>ST</del> SFLESNTDVRGR <del>LV</del> ARFRLSLPLQ NNSVDPPRLNL <b>KIEPYAVIVVRL</b> LIVAMS <del>VBXIQ</del> XV <del>XARV</del> VX <del>XS</del> GPKV <del>TL</del> SFND <del>QLC</del> VTVSD <b>RVIGDPDVPVTFRR</b> LR <del>CR</del> IPRVGR <del>LWVR</del> RRGR <del>LR</del> RI <del>FT</del> SR <del>RC</del> FWIISGF <b>RGRLSPTVTQEGFVRV</b> CNITKAANPSILLPTPTSQIAQ <del>S</del> ISTAQMV <del>S</del> TSASIFATPVLA LQSSSLRISPASTAPTSATVSSPVASIS	8.9	0.11 - ASM 0.17 - AIM	96 111	-

## Appendix B

CUB domain-containing protein	B3EX01	>CUB domain-containing protein; JR989025 MFLESLTVLSALVLTESIPSVATDFPPFFETKKFPDDIETYNNDYGILKFEQEPMENLT CASCEAPSERECTLNQTAVVCDDPNIACLTFEAFNNMTMTTFR <b>RGCFLSGLICENACR</b> SFNASQDGNLTSVCVQCCNSSLCNAGSLPTEVTEASTAQETTATSTTTKQSTGASTTA EPSTTAAPSTTTKQTTVASTTATTTKPTAPQTRATTLPTTAPTAPAPIACGGVLR <b>RG</b> <b>TF</b> TS <b>PGFP</b> GNYPNNVRC <b>EW</b> RVFLPRRQAI <b>IV</b> RV <b>LSLDLADPGDSLEFFDSGRV</b> IRTFRGL SRRKRSPSHRQTTNEKVLGEGEDGYDDQEVVDYVYDGRKREPFYQRRKRRQDRI VIQGRNOVAGAI <b>FQSDAAGNAAGFSTQ</b> PVQGAADSESEASASSESD*	13.7	0.12 - ASM 0.12 - AIM	84 116	-
MAM and fibronectin-containing protein	B3EX02	>MAM and fibronectin-containing protein; JT013217 KPHYHMYGATINRLNVFNGNCTVFTKLGHQGNMWMYAEVTVFVQNNITFEGIRGYSYTG IAIDVSLMEGICAGCKENLTDSPFGLHITYSAKFSPDCTWTIRNSSISEPVAIIIEEV QFAYCRGYI <b>KVFDGSGAQIFTRR</b> GCNENHTSNTFLEITFQESQNVTIQVSLENNQSYAR GYGILEGGLESALLLPGWNASLENKTSLSLQLRWMDISSWLRDGLRFFVVTAKSSYSNLT VKGLFSSNTTFAEISGLDPYMAVDVSVVAVDGDGSGQFKSTVLQ <b>ARTEWVPSRA</b> PSVFVT SVTSTSVTQWNPLPQQYHNGRLLYRVFIRKTANSPPFLDESNAVYNTSWVTLNLLKP GQPYEVNVSAFTSKGDGPRSTHYIVTAVCGKRPHTSLNCRHSSTHQRILASLNATDA RW	5.7	-	81 99	-
MAM and fibronectin containing protein 2	B7T7N1	>MAM and fibronectin-containing protein (isoform); JT016410 <b>FSPDCTWTIRNSGISQPVAIVSIEEVQPGYCRGYIKVFDGSGAQIFTR</b> EGCHENHSSNAF LEIAFQESQNVTIQVSLQNNQSYARVGVILEDDLESASLLPAWNVAIENKT	42	1.27 - AIM	90	-
Glu-rich protein	B7W112	>Glutamic acid-rich protein; JR983175.1 MKVFVYLLVTFSLTNA <b>SPLRNR</b> FNEHDHDEFKDDMARESFDTTEEMYNFLNRRDSSSEQ EDHLLSHAKPLYDDFFPKDTSPPDDEDSYWLESRDDGYDLAKRKRGYDDEAYDDFDEV DDRADDEGARDVDESDFEEDDKLPAEESKNDMEETFDEPEDEEKREAREEFAEDERAD EREDDDADFENDEEDEVNDKAESDIFTEPFAGVSDAEMDNFRDNEEEVADESDE AEDSEETADDFEDDPEDESDETFRDEVEDESEENYQDDTEEGSEI <b>KQNDTEEQPEKFF</b> <b>DADKE</b> HEDAPEPLKEKLSDESKARAEDSDKSEDAAKEIKEPEDAVEDFEDGAKVSEDEA ELLDDAEALSDDEAELSDEAEQSSDEAEKSEDKAE <b>KSEDEAELSEDEAKQSEDEAKAE</b> <b>DAAGKE</b> SNDEGKKREDEAVKSGIARDESEFAKAKSNLALKRDENRPLAKGLRESAAHL RDFPSEKKSDAQAQNIENELDYFKRNFAFADSKDAEPYEFDFK*	9.4	0.13 - ASM	77	-
Protein similar to cephalotoxin	B7W114	>Protein similar to cephalotoxin; JR986059 RWLGWQKFCWISCLFSSISSGLDPGEQ <b>KVTALD</b> TAQ <b>FAINAINEEYIAQAKA</b> IEEALK VSTQARSADLLRRQTE <b>LAKFGSKV</b> GALKAVQAASAIASFVTFMPSELVDVITSLINER FNEVNAKLDRIDEKLEMEKSIKADTAPNVFLSAWIKWEYKVRNGAKKLSIRKAMGTKT QRIDQVKLAEEYVKYYETNNLDGNVLSLY <b>RMAALPESITQRN</b> IFD <b>RFAIQPGCDITK</b> LSE LMILVQNIMTSAGQKLTYYYFKGDQSRANSS <b>KDIQMYFFKI</b> RQGFDRDVRHCRNRSLD YAKRDANKILKNMRGSSRESIV <b>RAIFNELK</b> VYKYPWYTWAVAAVKSDRPRI <b>RGLELR</b> GSTY FRLEDRSDAKKVGYPVYEDTRSSASCSDITQAKTLVFKKCDGCNSDYIYAADNLSK KRCGESTLERLVDFQ <b>QCPVCHR</b> WYSITCYCANRVKQDSQNMGLYCISSQH	15	0.04 - ASM 0.02 - AIM	76 75	-
USOMP-2	B7WFQ1	>Uncharacterized skeletal organic matrix protein-2; JR982706 MILFTAIIILVASVVHVSSPQQCYCYVEDDCETMSLWINQTCATSQRSLSGTSHCGTAAV	10.9	0.08 - ASM	70	-

## Appendix B

		RYHEGYLGGVPLETTVKGCFDCTDKSAACFALAGLLKSSLGWVVQQDCINCCNDTNCNTN VTILSQNATNVLRRDAFGTTSYCEESDNYTCILKQQSQTCRTSRALGITHCSSAKVK <b>TRNVLTGTVDVSFIRGC</b> ISCEDK <b>KSACALLAGSFKFR</b> KHATMLECDIECCNGSYCNDGAA SLSKCFHCEMEDDGLSCSARQQRICSLDPELSGTHCGSAVGRKRNRONGAIQNYFYRGC DCSKKKEACFTLGGYWKGDVNPAGATLLECELCQCCDPNVINGSYCNVETPILKPAAITV FTPTVTGPAQCNCVLEKDETSCESENQQTQVCGIDPYSLGTHCGSAVGRY <b>RQSGNDMVYG</b> <b>FYRGC</b> INCAD <b>KMAACA</b> <b>AVGGFRK</b> NVQKWTQLQCEIECCTEDNCNHTPRLVEVEQPSAP RGEIHQLFRCTFFVAVFIVFACFIVC*		0.12 – AIM	57	-
USOMP-3	B8RJMO	>Uncharacterized skeletal organic matrix protein-3; JR997000 MKICGLEKFRVFLSLISMVSLLCNGVNGFTTIV <b>RSMAVNGESV</b> <b>PDRF</b> SNPSCRPSDCALKR ASTTNGCSTTRDCCSCQCSRT <b>RATYLTSPFNR</b> CTTSEYIDEDCSSFFVLDDSPPPVADI TKPGHINFFSETRCHKGLRTRSWSHSDATSWTTGKPNGFSELVEGSSSSWKWRLSWQN GMDAKFSGLI <b>I KLEFSCQNT</b> RGCPMLKSKGNITIPNSEQWPSIIPTDVSFNLTGENANP TANSGETSARSNRNEQNKMEEPARNQAELEPKKTGVVAVGTVLSAAGFVLALATLLMKK KQTS LAVNAKARPNSYLGYEPEVDSAGRPEQTATESPSFDNEFYTTDCVLSLGSNGVGGK VTRMGFLPPLPGEES IYAEPMIKRSVAYQGLAEKNKQDAGTACNVQPQPECKVIEKTSN ENSHDKGTDEDKG	8.5	0.07 – ASM 0.07 – AIM	69 75	9
Galaxin 2	B8UU51	>Galaxin 2; JR976690 MTRFTSIGLCAVLLFNVSCATLQKDTIASMLKKGNSPRVTRQRQLPSPCGSLQPQGLC CDSYKYNPVTHLCCNDNPAVKPASPATAIPGCCDQASAYDRNTHLCCDATLSPHPPATLLPA CCGPVVYDSSVNSTQLCCAGAVLNKFPVGV <b>PALCCGTATYNPATQVCCMGFPVPKAGGPN</b> ATSLLCCGPFYSYDISTQMCNNGNIALKSATHTHCCMGFSFNPATHLCCNGYPP <b>KLGFISIP</b> <b>SCCGSLVYD</b> <b>TLTMRC</b> CDGSHVVLITPNQDPCANLA*	17.5	0.12 – ASM 0.49 – AIM	68 307	-
PKD1-related protein	B8UU59	>Polycystic kidney disease 1-related protein; JR991141 <b>K.VASQVLYNVIK.N, R.SSTAFQILVVR.E, K.GGQTYLATFDVR.D</b> <b>K.SGLASGSDGCTGNEIK.Y</b>	1.9	0.04 – ASM 0.03 – AIM	65 90	131
Zona pellucida domain-containing protein	G8HTB6	>Zona pellucida domain-containing protein; JN631095 MFLYSFVFLMLLGLSSAQTESATSPDEVETEPTMSTDQPETSPSMSTETEPTTETPPVT PPPPDSLSVICTNEKMEVFLDHAKHDNLDLKVTLKDANCKASGTLNATHLWMDVFPDSC MTNHSTDGDITTYQNSLVAET <b>RASAGSSLSISRE</b> FQAEFPFKCTY <b>PSAVLSVAVFSPRE</b> IVYTKTAEFGNFTTMDMYKTDKYETPYDS <b>FPRLDLDMPFLEVKV</b> SSNDSKLVILPLK CWATPSSDLQDDKYTYTFIENGCKADDPSPVFNYESNVQRFKIGAFRFIGESLNSNVYL HCDVEACRKGSDSRCAKGCETSRRRRRSSLASSAGTEQTVTLGPMKISEKAEVGAQEA VSSLTIFAAVAGVLGVIVLFLAVALVMLYKRYRSPQSATRVVYTKTANEGLKLV*	9.7	0.03 – ASM 0.07 – AIM	65 79	18
USOMP-4	B8UU74	>Uncharacterized skeletal organic matrix protein-4; JT004498 SYHGAAATRAKQLLVQAAQPPPAARKHPAAAMIPTGPVTAPK <b>GRHTVEAEAQALPQAKM</b> <b>QATVAAGPLSTGGVLLRLIK</b> TMDTKMKEFNEIIFIIISRCQLTRNCRMNSVDAILKILP SIRGKLFGLKARIPMXXXHGVMDDDEYTAAMPVGF <b>PLEIEDETRKF</b> VFRIPKFSKRAL VDPSVTPGERTPKLGNKCWNMAAA*	24	0.1 – AIM	64	-
Galaxin	D9IQ16	>Galaxin; HM163215 MKPSGAFSLSCVLLSLATHCFSPSDSLRRDAHSDTNALKSRDRRQAPAPQLSCGGVLY NPAEMCCHGNEVPRVGSMPCCCESSYDPSTQMCCEGTVSNKPPGIAMCCGSEAYDANS	14.8	0.19 – AIM	152	



## Appendix B

		QICCNNGINTKATGPTAQPCCGCEFSYDAASQLCCDSDHPVLVMSGLPSCCRNGYDANTS LCCGDNVAVFVSGPQAACCGDMYNNRTHLCCDSDNVLPMAMPAGCCGSWTYSQTHLCCCE GVQLYKGMNTGCCGAVGYNQVNSLCCCEGTVPV <b>KSPSKPVCCGTTSYNPLTELCCDGIAFF</b> <b>KTGFIPTCCGGAIYDATVARC</b> CDGVPTYNVASAGLA*				-
EGF and laminin G domain-containing protein	B8U78	>EGF and laminin G domain-containing protein; JR980881 RTFVKKYSAS <b>RQFTGEGYLEVRT</b> TSGNIIDSDKDEL <b>RVEFSTVQPSGLLFYARNSSGGPFA</b> <b>DYVALELVGGRLR</b> FSIRYGRSSHSTENLHETLLGKNLNDAKSHSVEILHDKDVTIYLDK TSDQEKAEHSFKTKYTKLIDVAMYVGGADPFKALLSVKSNALFMCGIFQAEFKKILPGP EKVIDFLKDDKVTYPRTMNQKVAQTYEPTFFSSDDSFVCSVGLSSANSLSGSFVFR TYKPSGVLLKQVDGGNGFELSYMEMDVQLKVIIRNSETLLNINYQNELTKINKGNWHYVT FNISQTSFELSVGSKRET <b>RTPAVTLPSNFFKD</b> GLTAGGFVGCNMNELINKQKCPNAGSR IKNVEWGCNITDFCIFSPCLHGGECTQTGKTFSCGCSGTGYDKGPNLSVCFSESEST CESLK <b>KNNPSLSLSDRSYALDFFDSDGPIRT</b> YKAFCNFSADPPTRVESRDFKIKLTPSKQ PISQ <b>RISYEPSLDAAKA</b> LARRSEWCYQVDFGCKKAKLHTGNSNEKLGFWVSSNGVYQSY WGAQKQSRSCACGETNPNSCIDSSKKNCDAGLDKWHNDEGLNSTLLPVVEVMFKGV TSGTEANFTVGHLYCAGEISNTATFVNEDGFIELEKWSPPSNGVISLFFKTPYEKGVLLY NGMPEKDFVQVEINETSVLGSLYINIGVRRKIELSLGDKQVNDRSWHHVMYHNMKVFGF RLDNQEGKHENPLFLKRELNLNNELVAGYYPYDVSKGPFVGCIRGLDVGNEVQDLSKLAG AVFVKSGCCAACENNSCKNHAKCLDNYNVPFCDSKTPYYGYFCHENGASFKDPGSQLV YEYPSASDVFRFDIVVGFKLGEKPCIGDIIRLGSDDKSQFYRLSLNTRKLPQDFKGRPG QGSITIDPPSVGDFCRDVHTFALSRRYKVVNYTIDGVKKPKKEIERLDGLFTSMKKVTIG KEGDGFGKCIITGVKVTREAVGQKPEVPEIKEYLYDDKNKLDVTSKHVSRATCGPEPKV PEIPTPRPVGQRADVSTPQGITTPNKLOAEDDDKTAIVVVVLLLVLLLVLLVLLVYIYWY ARHKGEYHTHEDEELKATDPYIEPAAPRKLKGEPEKKEKWEI	9	0.07 -AIM	110	66
Carbonic anhydrase	B8V7P3	>Carbonic anhydrase; JR998014 CLKRLQPGEMSLQLLSGCLRLEQETGVLGRFADL <b>TKIIPQDSDETVRFSDGIFIRGL</b> IPQRNTRF <b>SRLAILNCYTYTKGSLTTPICSENVTLIVKPRLPATNNMMRKFR</b> RLLETPA <b>GKNPPLMCDNFRPVQPLNGRTVFEVHRI*</b>	56.8	0.43 - AIM	108	-
Protocadherin-like	B8V7Q1	>Protocadherin-like; JT011093 <b>R.FEGIAANGR.V, K.AELEALSLK.I, K.FAVIDASGR.F, R.TVYTFEVR.E,</b> <b>R.AETGVIVTAR.V, K.FSADSIVTK.V, R.ITFMEAQPK.N, R.EDITINTQVK.L</b> <b>R.LLSYCILDKV.V, R.QSQYDLIVEAR.D, R.DTFVTVIHATR.D,</b> <b>R.GTAVSYIASAAVKG.F, R.VIATDPDTGAAAIAK.Y, R.ISGLVTTVETMEK.E</b> <b>R.AYDGANSATTGITVK.I, K.IDNLLCIAAYGVR.G, K.NAPYSVTVPENLKG.I</b> <b>R.VSDGNDQAPVFNPR.E, R.FFPGGTSLSIIFPQK.A, K.NIAIEDFSPPGSPVIR.V</b> <b>R.MKILKIPQLNVTDDK.Y</b>	6.3	0.04 - AIM	103	207
Collagen	B8V7R6	>Collagen, type I, alpha 1, JR991083 APGPDGLTGTKGSMEPEPTDGEPSGPGQAKGETGLAGRRGLTGIPGKQGRQGERGEPG TAGSQQQQPGTQGPPLGPKQGETGEPGESGEDSTPGPRGERGAQGERGATGMMGPPSG DPGEAGIPGADKAGERGVGAPGVPVGTPLGMPMPQQGMPGIGAKGSKDVGPTGERG YDKGDGEPGRDGSPPGIGQPGIPEKEGDGVPGSDDTTPGSRGSDGPRPLGNPGLPPGPRFG ALGPSGPPGQPRGRGEPGMKGPAGPPRGATGALGQLGKTGLKGEPEGNQGRGRRPFG	1.9	0.08 - AIM	101	

## Appendix B

		LQGDPGKPGQSGPPGPPGSPGSGRDSGQKGSSEGPGRPGKDGIPGQPSNGKDGEPG TPGSDRAGEIIGPSGPIGPKGERGTPGATGPMGNSGPPGVQSGKKEKGGPPTNGRNGSPG ISGSRGAQQPPGAPGSSGQNGVDGGTGENGTNGRPLKGESGAPGDPGASGAGPAGPPG PKGDTGPPGIQGEKRRRAGDGI PKGTGEPGPGDQKQKQKGEVGVGVEKDKGWTGTTPG DPGPGDRGEFPGPPGRDGVDPGPPGRGAPGEMGAVGDPGLNGSMGEPGNKGPDDGLGESG AKGPDGIKGPPGPPGPPGPPGQPMSEIASYLSVGNLEKPGFRLYSSSGEEMPKQKIKAKA ENVLKDLEKDKEMDSL IAPDGRKFPKACTCYDLFLDHGNFESGEYWIDPNGTGVKDAIK VYCDKKNSSCVYTPNKIISDLVLKSGFESKEDKWLKAFKKESEVEYDAHYTQINFLRT LSNYANQNVTYACRNSKAWEDGQHSIKLMGSDMEYHASSKISLRPTVIMNECANGKLD KWGKTVLEIDTR <b>RSRLPIVDVSADFVGRE</b> QDFKLEIGPACFHHIKY*					-
CUB and Ser protease domain-containing protein 1	B8V7S0	>CUB and serine protease domain-containing protein 1; JR970990 SGFHLSFSFFRRAVCGIRPTLSGFIVGGTVAPINSWPWQAKLRIAGNFLCGSLIQPEWV LTAACHVEGESPSIIKVTLGAHYLSTAQVVGTEQYFVQIIQHENYKMPK <b>RFSNDVALL</b> <b>KL</b> SRPAALRNGVGLVCLSDDDQFQRPFNGTSCWTTGWG <b>RLSWGPPVAKE</b> LMQVLDLPLVSPQ NCLSSYPNGYDPNTMICAGRSQGGTGAC <b>RGDSGGPLVCEFKG</b> KWYLEGVTSWGOLPCDLP NKPTVYADVRLKLSWITGKISRSALKVATNCSSVLNNTLKSYPYDPSYPIINMFCVYRVP IPCDTELVIHFNSFHLENHVFVWYDRLRITDGSN <b>RVIGTYCGQQTGRS</b> VLVNMTVAVLTF KTRDSLNSGFLHSFSFFPRGNATLPPFTTPTQTTRPTTTPGCGVQNNLTRSPGY PSNYPRNTHCVYRVF	11.5	0.16 - AIM	98		-
CUB and Ser protease domain-containing protein 2	B8VIV4	>CUB and Serine protease domain-containing protein 2; JT008002 QPKELMQVDLPLVSTQNCSSLYANYDPSTMICAGTRQGGTGACNGDSGGPLVCEFKGKMY LEGVTSWAGVPCASPSKPTVYADVRLKLSWIAAKITGVPVLRVATNCNSVINNTLKSYPY PNSYPIINMFCVYRVPICDTELVIHFNSFHLENHVFVWYDRLRITDGSN <b>RVIGTYCGQQT</b> <b>GRS</b> VLVNDTVAVLTFKTRDSLNSGFLHSFSFFRRAVCGIRPTLSGFIVGGTVAPINSWP WQAKLRIAGNFLCGSLIQPEWVLTAAACHVEGESPSIIKVTLGAHYLSTAQVVGTEQYFD VQIIQHENY <b>KMPKPFNSNDVALL</b> KLSRPAALRNGVGLVCLSDDEQFQRPFNKTTKSCWTTG WGTLYRGSQPKELMQVDLPLVSTQNCSL	7.5	0.18 - AIM	82		
USOMP-5	B8VIU6	>Uncharacterized skeletal organic matrix protein-5; JR973117 MGAARFLVQVAIFLLVKPARSAPAPMWKGNSTARKSCSQASINNCSCRELSPASTTANA VSALEDKIDQVIALANRTRPRHSAPVASISSCKEQFDKNSSPSQVYELTFGQVVPVYC HMGNFGCGNGGWTLAMKMDGKTTHFYHDSLWSAQSSYNPAAG <b>KTFGDMLETKL</b> PTYWST PFDKVLGMRLLGQQLNFVVLNMTANSLSLFIADGLYRATSLGRNT <b>WKSLLGAQASLQRNS</b> IEKGSTPLGVVIGMPG*	9.4	0.08 - AIM	72		-
Neuroglian-like	B8VIW9	>Neuroglian-like; JR993827 <b>MWQILLASISFSLSKLSNAQQQPKVAPPQITNPLAEDKVAPPEVKFRDTRDQVWQLVLP</b> CRATGSNPLKVVVHNNAEINKNKFIFDRDWELLSDGLTRARGLNISDRGTYQCFVEDVTVKV STFSRKLREVTAVGDFKSHKDFTS <b>SVKLGEP</b> LV <b>VECP</b> PRG <b>PS</b> FGVTFAWTSKKARSIQF PISNRVAIDPSTGNLHIMYITTEEDVSTFNDELIGRCTISAANTFYSSGALTLQIIPGKEI KLSSPSTSSSTSPNENAVEGRKLDLYCEATARPPPKIVLVKNGVEL <b>KSGIDFIEIPEAF</b> <b>EGRL</b> LSITSVKESLHETTFTEASNNQTIASGPAQONFVLNVEVAPRWASKPPDLSKEIP ISSNGNLSCDVYAQPEPEIKWYRDGREITQSSKVEVSGSKLLFKDPTLDEAGIYQCSAE NVHGMIVSSTYV <b>KVLA</b> I <b>APS</b> PKNGFGFPYLFQDSEGRLLKCDPEAAPRSTFKWFDENGAE	4.2	0.07 - AIM	69		

## Appendix B

		IKSGNGYTI EEDGTLVITKVERSQHAGKFCYAKNFLGNATAEGTATVYDRTRIVRGPSD LSVNEGTRVDLRCEAVADSSLELHYTWKRDDATI EYNRRVQWLKDNQVLT IADLTVEDAG IYTCVAYTPQPKYSEAKASAI VNIAGAPFPPTNLMSSECCNRNTT LSWVTGESNNAS I L YFLIERKKSQYADDFWQVI ANVTNP NATSHPLV <b>KL</b> AGNAD <b>LAFRI</b> RAVNRF GPSRSPSEPTG SFCRTIRAVPEKWPDPNFRGVPGKAEELTIAWTAMRRVEWNGPGLYKLVYRRVNSGDALV EVRRREASSDSFVVPDAGYRQWEFQIQAINVEVGEKPSPLVKQFSQDPP T GKPEDVTVG TITARSVELSWKKVTFTRGSVDGYRIYFWGESRVSAKRRRRAIPGYASVTNVTGVNTERY TVTGLKPYTNYKFVITAYNSGGNGPESDQVAADTDEAEPGPPSDQVQVFAKYILVWQOP PSEPNGVITNVRVGTET Y TGSQPTDVTVMMEETGVEARRKLLRDLVPE TNVYVEMQAATS KGWGTSPFRKTEKTVAWAAPAKPEKPIVEGTAVDEVVRVDYKFGGGGYTHDFLVMFRKKIE GQEFQNTSWVDHFQQS I I I GNLDPELYQFKTVARNDYPSQENPQESPASDI TEARPRPG ISNVGKRVSTPIYQSAWF IALLVLI <b>IA</b> LLLLLVLLTFVLYTRHOGAKYLVGKREKKRAALI DREHFD EEEGFSNNGRADHPPYPYSGSLPRGADSDRDSLDDYGE GPQFNEDGSFIEEY GDEKKAPPEEKDPSSLATFV				74
USOMP-6	B8VIX3	>Uncharacterized skeletal organic matrix protein-6; JR971508 MKCAVAILLVCLTQQAA YGFLYNEEVKTEFORRKQSL EEEAGSLKMGQNLQDNMQRSL AEGQEALQKHIKNLQOSMLSQKEALRNRGEALRET VGERLES MONGKDMKKMQBEGRET LQKKLGEQVETFNQTFQAGRLAIAKKVLEGS E TMRKT I QNTTQSLQDKAEKVQETAGKNV EALKLIARKNALSLKESLDTRENSVEENMQAL <b>RNPLPSQSEAMDLPKE</b> KLQELMASIQN NTGLFQESWGQEKEMKEMLRGLKRVGERTEDMKRKM KARKEELEAEFQS <b>RGDEAVQTV</b> <b>MEIRN</b> VTIKHLREAGKKIKEIEEKIASLLPNSCLDFLRSKALKMGVKIVVQDLKSVFRMG WLRVPETFEKEE E IAPSTEEDGSEEL EADSYDSKVGESPI SQRT EERQGA EERSLR LRRR RAAVLRRMFGQWSRKS*	6.9	0.1 - AIM	66	-
USOMP-7	B8WI85	> Uncharacterized skeletal organic matrix protein-7; JR998260 MLS L I P F T V C A F L A L I T S K G G S A T P S T I S L E C S E N D V C A A E T L T R R Q D R L Q K T L N L C T D D E S Q P T L T A V V K C T S V I Q V P F P N R H F K M A A L D L S S G I C R A L A Q P V V A S Q A T V S A E I L N E A G N K G V N S G H P G L L F N A I D E N N F D F V Y L R P H S V S G C Y Q T G Y M S A G V N K F V E S K R C P N G P P K G G E W F P F S V T V N G Q Y A T V Y <b>RS</b> GVLV <b>TFK</b> THF ASSRARGGVIFNGYKNVILFRKFKTA PKHFFSKRCKEVVEFFPAGYVKMDAGIGSWPKDAFCQVEFGSDGRIASYELKVDLYNFIGR DKANLGHGPGVFFNAEDEDNYDFVYFRPHSVGGCCFTGYLLKKGKPRFDGAKSASC PKGPPK GKTWFNV <b>KLTVSNATPAGEVRV</b> LDLDTLVTSFNPRYP I KRRGGVLVANGYKNV I YLRNFK IL	4.2	0.07 - AIM	65	-

## Appendix B

Table 2: MASCOT hits identified solely with one peptide. These 7 sequences were not included in the list of biomineralization proteins.

GenBank Accession Number	Unique Peptides	emPAI	Total Mascot Score	BlastX Hit	E value
JT000026 (JR989881)	R.QVQCNPGEVCTTLEAFNLDLTTTVTR.G	0.23	88	Q9TU53.1 RecName: Full=Cubilin; [Canis lupus familiaris]	3.00E-04
JR981801 (JT020142) (GQ228826) (EZ012961)	R.VGLSDAFVILQR.D	0.12	71	Q9DC11.1 RecName: Full=Plexin domain- containing protein 2; [Mus musculus]	1.00E-48
JT004105	R.SAMVSQDVIR.A	0.07	70	-	-
JR977100 (EZ012413)	R.ASVTDLTDAENR.L	0.04	69	P49614.2 RecName: Full=Beta- hexosaminidase subunit beta; [Felis catus]	6.00E-161
JT021931	K.IISQLCALCQGTSR.S	0.26	62	P27425.1 RecName: Full=Serotransferrin; [Equus caballus]	4.00E-22
JT002294 (EZ012364)	K.TCLQIEPGSLEEEIEK.C	0.08	61	Q96RW7.2 RecName: Full=Hemacentin-1 [Homo sapiens]	3.00E-51
EZ012413 (JT002295)	K.IVTESITSEAQK.T	0.05	59	Q4R4T8.1 RecName: Full=Legumain; [Macaca fascicularis ]	5.00E-130

## Appendix B

Table 3: Results from the similarity and homology comparisons between the 36 SOMPs and the proteome of *Acropora digitifera*, *Nematostella vectensis* and *Hydra magnipapillata*. Grey scale: dark - Similarity (with no conclusive evidence to infer homology); medium - Homologues in *N. vectensis*; light - No homologues in *N. vectensis* and *H. Magnipapillata*; white - Homologues in *N. vectensis* and *H. magnipapillata*.

<i>Acropora millepora</i>	<i>Acropora digitifera</i> Source: <a href="http://marinegenomics.oist.jp/acropora_digitifera">http://marinegenomics.oist.jp/acropora_digitifera</a>						
Protein name	BLASTP (SEG)		BLASTP		BLASTN		InterPro domains
	Hit (Protein No.)	E value	Hit (Protein No.)	E value	Hit (Transcript No.)	E value	from NT to CT
SAARP1	adi_v1.11068	2.00E-153	adi_v1.11068	5.00E-174	adi_EST_assem_12928	0	SP
Acidic SOMP	adi_v1.06327	5.00E-64	adi_v1.06327	5.00E-67	adi_EST_assem_995	0	SP
SAARP2	adi_v1.01441	8.00E-69	adi_v1.01441	7.00E-81	adi_EST_assem_6252	0	SP
Mucin-like	adi_v1.09809	0	adi_v1.09809	0	adi_EST_assem_5353	0	SP, Thrombospondin, type 1 repeat, Nidogen, AMOP, vWD, EGF
SAP1	sap1	5.00E-42	sap1	1.00E-43	adi_EST_assem_34783	0	SP
SAP1	sap1	7.00E-28	adi_v1.06593	2.00E-34	adi_EST_assem_31408	0	SP
Uncharacterized SOMP-8	adi_v1.01189	7.00E-78	adi_v1.01189	8.00E-97	adi_EST_assem_8846	0	SP
Coadhesin	adi_v1.05945	0	adi_v1.05945	0	adi_EST_assem_1538	0	SP, Coagulation factor 5/8 CT type, Thrombospondin type 1 repeat, vWA
SAP2	sap2	1.00E-31	sap2	2.00E-42	adi_EST_assem_16174	0	SP
MAM and LDL-receptor domain- containing protein 1	adi_v1.09968	0	adi_v1.09968	0	adi_EST_assem_1163	0	SP, MAM, Fibronectin type II collagen binding, Ricin Blectin domain, P-type trefoil, Low density lipoprotein receptor

## Appendix B

MAM and LDL-receptor domain- containing protein 2	adi_v1.09969	0	adi_v1.09969	0	adi_EST_assem_4944	0	MAM, Low density lipoprotein receptor, EGF-like
Thr-rich protein	adi_v1.04566	3.00E-68	adi_v1.10941	4.00E-74	adi_EST_assem_9510	0	CUB
Ectin	adi_v1.13233	9.00E-154	adi_v1.13233	7.00E-166	adi_EST_assem_19083	0	SP, Thrombospondin type 1 repeat, CAP, Zinc finger, RING-type
Hephaestin-like	adi_v1.16742	0	adi_v1.24015	0	adi_EST_assem_13507	0	SP, Cupredoxin
Uncharacterized SOMP-1	adi_v1.21723	6.00E-126	adi_v1.21723	2.00E-138	adi_EST_assem_114	0	SP
CUB domain-containing protein	adi_v1.14283		adi_v1.14283	3.00E-173	adi_EST_assem_21039	0	SP, CUB
MAM and fibronectin-containing protein	adi_v1.01383	9.00E-150	adi_v1.01383	6.00E-162	adi_EST_assem_14016	0	MAM, Fibronectin type III, Petidase cysteine/serine trypsin-like, Metridin-like SHK toxin
Glu-rich protein	adi_v1.04188	5.00E-113	adi_v1.04188	6.00E-142	adi_EST_assem_1759	0	SP
Cephalotoxin-like protein	adi_v1.09855	0	adi_v1.09855	0	adi_EST_assem_33327	4.00E-136	SP, EGF-like, Thrombospondin type 1 repeat, Low density lipoprotein receptor
Uncharacterized SOMP-2	adi_v1.15064	0	adi_v1.15064	0	adi_EST_assem_1253	0	SP
Uncharacterized SOMP-3	adi_v1.14490	5.00E-98	adi_v1.14490	8.00E-114	adi_EST_assem_6836	5.00E-170	No
Galaxin 2	adi_v1.15065	2.00E-135	adi_v1.15065	2.00E-135	adi_EST_assem_8935	0	SP
PKD1-related protein	adi_v1.02830	0	adi_v1.02830	0	adi_EST_assem_6849	0	SP, Carbohydrate-binding, PKD/Chitinase domain, PKD/REJ-like, GPS, Lipoygenase LH2, Polycystin cation channel PKD1/PKD2
Zona pellucida domain-containing protein	adi_v1.07627	0	adi_v1.07627	0	adi_EST_assem_2269	0	SP, ZP
Uncharacterized SOMP-4	adi_v1.01440	7.00E-27	adi_v1.01440	6.00E-27	adi_EST_assem_13773	5.00E-112	No

## Appendix B

Galaxin	adi_v1.18631	6.00E-103	adi_v1.18631	6.00E-103	adi_EST_assem_14006	0	SP
EGF and laminin G domain-containing protein	adi_v1.06122	0	adi_v1.06122	0	adi_EST_assem_51	0	SP, LamG, EGF-like
Putative carbonic anhydrase	adi_v1.22702	8.00E-39	adi_v1.22702	1.00E-39	No hit	No hit	Alpha-CA
Protocadherin-like	adi_v1.19518	0	adi_v1.19518	0	adi_EST_assem_2804	0	SP, Cadherin, EGF, LamG, Cadherin cytoplasmatic
Collagen alpha-1 chain	adi_v1.00434	5.00E-62	adi_v1.09052	4.00E-64	adi_EST_assem_818	0	Collagen triple helix repeat, Fibrillar collagen, C-terminal
CUB and peptidase domain-containing protein 1	adi_v1.08323	0	adi_v1.08323	0	adi_EST_assem_9461	0	MAM, Fibronectin type III, CUB, Petidase cysteine/serine trypsin-like
CUB and peptidase domain-containing protein 2	adi_v1.16372	6.00E-115	adi_v1.16372	6.00E-115	adi_EST_assem_9127	0	Fibronectin type III, Petidase cysteine/serine trypsin-like, CUB
Uncharacterized SOMP-5	adi_v1.22918	1.00E-116	adi_v1.22918	1.00E-116	adi_EST_assem_8248	0	SP
Neuroglian-like	adi_v1.16442	0	adi_v1.16442	0	adi_EST_assem_1371	0	SP, Immunoglobulin-like, Fibronectin type III, Fibronectin type III C-terminal domain
Uncharacterized SOMP-6	adi_v1.05151	0	adi_v1.05151	0	adi_EST_assem_360	0	SP
Uncharacterized SOMP-7	adi_v1.08326	8.00E-153	adi_v1.08326	8.00E-153	adi_EST_assem_2396	0	No

## Appendix B

Table 3 (cont.): Results from the similarity and homology comparisons between the 36 SOMPs and the proteome of *Nematostella vectensis* .

Acropora millepora	<i>Nematostella vectensis</i> Source: <a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>										
	Protein name	BLASTP (SEG)		BLASTP		TBLASTX		InterPro domains	Global sequence alignment*		Neighborhood Correlation Coefficient
		Uniprot Ac. No.	E value	Uniprot Ac. No.	E value	Uniprot Ac. No.	E value	Annotated from NT to CT	% Identity	% Similarity	
SAARP1	A7RRP3	7.00E-15	A7RRP3	3.00E-14	A7RRP3	2.00E-11	SP	20	33.3	0.647675963	
Acidic SOMP	A7RRP3	2.00E-13	A7SQ27	2.00E-15	A7RRP3	5.00E-07	SP	22	39.2	0.634254745	
SAARP2	A7SQ27	2.00E-18	A7SQ27	2.00E-19	A7SQ27	5.00E-13	SP	18.8	30	< 0.6	
Mucin-like	A7S664	2.00E-48	A7S664	2.00E-48	A7S664	1.00E-60	SP, Thrombospondin type 1 repeat	13.5	22.4	0.735211996	
SAP1	No hit	No hit	No hit	No hit	No hit	No hit	-	-		No	
Uncharacterized SOMP-8	A7RLD3	9.00E-07	A7RLD3	2.00E-08	A7RLD3	1.30E-02	SP	26.2	39.2	< 0.6	
Coadhesin	A7RL2 (fragment)	3.00E-78	A7RL2 (fragment)	1.00E-78	A7S9H7	1.00E-103	Thrombospondin type 1 repeat	9.4	11.9	0.778148606	
SAP2	No hit	No hit	No hit	No hit	No hit	No hit	-	-		No	
MAM and LDL-receptor domain-containing protein 1	A7RL30	0	A7RL30	0	A7RL30	0	P-type trefoil, MAM, Low density lipoprotein receptor class A repeat, EGF	37.2	46.2	0.948425179	
MAM and LDL-receptor domain-containing protein 2	A7RL31	0	A7RL31	0	A7RL31	0	Fibronectin type II collagen binding, Kringle like-fold Carbohydrate-binding WSC, MAM, Ricin B lectin domain	-	-	-	



## Appendix B

Thr-rich protein	No hit	No hit	A7RGF1	2.00E-07	No hit	No hit	EGF-like, Zona pellucida	-	-	-
Ectin	A7S664	3.00E-26	A7S664	8.00E-27	A7S664	8.00E-35	SP, Thrombospondin, type 1 repeat	14	22	0.849871768
Hephaestin-like	No hit	No hit	No hit	No hit	A7SVQ9 (fragment)	0.044	Cupredoxin	-	-	-
Uncharacterized SOMP-1	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	-
CUB domain-containing protein	A7S3J5	2.00E-07	A7S3J5	3.00E-07	A7S3J5	3.00E-06	Peptidase M12A astacin, CUB, EGF-like	7.4	10.6	< 0.6
MAM and fibronectin-containing protein	A7RL30	2.00E-17	A7RL30	2.00E-17	A7RL30	2.00E-13	P-type trefoil, MAM, Low density lipoprotein receptor class A repeat, EGF	6.6	9.9	0.83411391
Glu-rich protein	No hit	No hit	A7S5Q6 (fragment)	2.00E-04	No hit	No hit	No	-	-	-
Cephalotoxin-like protein	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	-
Uncharacterized SOMP-2	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	-
Uncharacterized SOMP-3	A7SZS2	5.00E-05	A7SZS2	6.00E-05	No hit	No hit	SP	8.1	13	No
Galaxin 2	A7SES0	4.00E-21	A7SES0	4.00E-21	A7SES0	1.00E-20	SP	26.2	37.4	0.76502779
PKD1-related protein	A7RGF9	1.00E-97	A7RGF9	7.00E-111	A7RGF9	1.00E-179	SP, Carbohydrate-binding, PKD/Chitinase domain, PKD/REJ-like, GPS, Lipoxigenase LH2, Polycystin cation channel PKD1/PKD2	24.3	39	0.898049115
Zona pellucida domain-containing protein	A7SIJ9	5.00E-57	A7SOC8	1.00E-69	A7SIJ9	6.00E-53	SP, ZP	36.5	51.6	0.975282219
Uncharacterized SOMP-4	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	-

## Appendix B

Galaxin	A7SES0	7.00E-23	A7SES0	7.00E-23	A7SES0	3.00E-24	SP	22	33.3	0.764475978
EGF and laminin G domain-containing protein	A7RRW5	2.00E-174	A7RRW5	0	A7RRW5	9.00E-134	SP, LamG, EGF-like	34.9	52.9	0.917005544
Putative carbonic anhydrase	A7SHS9	5.00E-14	A7SHS9	5.00E-14	A7SHS9	3.00E-16	Alpha-CA	-	-	-
Protocadherin-like	A7SAP5	0	A7SAP5	0	A7SAP5	0	SP, Cadherin, EGF, LamG, Cadherin cytoplasmatic	42.6	58.1	0.973760893
Collagen alpha-1 chain	A7S046	1.00E-87	A7S046	3.00E-179	A7S046	2.00E-122	SP, Whey acidic protein, Collagen triple helix repeat, fibrillar collagen CT	51.7	60.8	0.976769774
CUB and peptidase domain-containing protein 1	A7RGS8	7.00E-67	A7RGS8	4.00E-67	A7RGS8	4.00E-71	SP, Petidase cysteine/serine trypsin-like	6	7.9	0.761518309
CUB and peptidase domain-containing protein 2	A7RGS8	6.00E-47	A7RGS8	6.00E-47	A7RGS8	5.00E-53	SP, Petidase cysteine/serine trypsin-like	-	-	-
Uncharacterized SOMP-5	A7RHX3	2.00E-37	A7RHX3	2.00E-37	A7RHX3	2.00E-40	SP, PAN-1 domain, EGF-like	22	33.9	-
Neuroglian-like	A7RT95	0	A7RT95	0	A7RT95	4.00E-171	Immunoglobulin-like, Fibronectin type III, Fibronectin type III C-terminal domain	37	52.5	0.915141387
Uncharacterized SOMP-6	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	-
Uncharacterized SOMP-7	A7SQ30	2.00E-79	A7SQ30	2.00E-79	A7SQ30	2.00E-78	SP, Concavalin A-like lectin/glucanase, subgroup	23	31.4	0.896210923

## Appendix B

Table 3 (cont.): Results from the similarity and homology comparisons between the 36 SOMPs and the proteome of *Hydra magnipapillata*.

<i>Acropora millepora</i>	<i>Hydra magnipapillata</i> Source: <a href="http://compagen.zoologie.uni-kiel.de/segret.html">http://compagen.zoologie.uni-kiel.de/segret.html</a>									
Protein name	BLASTP (SEG)		BLASTP		TBLASTX		InterPro domains	Global sequence alignment*		Neighborhood Correlation Coefficient
	Compagen No.	E value	Compagen No.	E value	Compagen No.	E value	Annotated from NT to CT	% Identity	% Similarity	
SAARP1	Hma2.232959	1.00E-13	Hma2.232959	9.00E-16	Hma2.232959	2.00E-07	SP	15.3	26.1	Hma2.232959
Acidic SOMP	Hma2.232959	2.00E-08	Hma2.232959	2.00E-10	Hma2.232959	0.006	SP	15.1	24.7	Hma2.232959
SAARP2	Hma2.232959	2.00E-16	Hma2.232959	6.00E-17	Hma2.232959	2.00E-13	SP	15.1	28.8	Hma2.232959
Mucin-like	Hma2.205838	1.00E-41	Hma2.205838	1.00E-41	Hma2.205838	3.00E-60	Thrombospondin, type 1 repeat	16.5	23.6	Hma2.205838
SAP1	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
Uncharacterized SOMP-8	Hma2.228261	4.00E-03	Hma2.228261	8.00E-04	Hma2.228261	9.00E-05	SP	19.1	29.8	Hma2.228261
Coadhesin	Hma2.220156	1.00E-89	Hma2.220156	1.00E-89	Hma2.205838	2.00E-124	SP, Thrombospondin, type 1 repeat	19.1	27.5	Hma2.220156
SAP2	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
MAM and LDL-receptor domain-containing protein 1	Hma2.217613	0	Hma2.217613	0	Hma2.217614	0	MAM, P-type trefoil	19.9	27	Hma2.217613
MAM and LDL-receptor domain-containing protein 2	Hma2.217613	0	Hma2.217613	0	Hma2.217614	0	MAM, P-type trefoil	-	-	Hma2.217613
Thr-rich protein	No hit	No hit	Hma2.229982	2.00E-05	No hit	No hit	-	-	-	No hit
Ectin	Hma2.214763	2.00E-22	Hma2.214763	3.00E-23	Hma2.220156	1.00E-31	Thrombospondin, type 1 repeat, vWA	10.4	15.4	Hma2.214763
Hephaestin-like	Hma2.212999	8.00E-03	Hma2.212999	8.00E-03	Hma2.212999	7.00E-04	SP, Cupredoxin	11.9	20.8	Hma2.212999
Uncharacterized SOMP-1	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit

## Appendix B

CUB domain-containing protein	Hma2.231497	2.00E-06	Hma2.231497	8.00E-07	Hma2.231497	1.00E-06	SP	10.8	19.8	Hma2.231497
MAM and fibronectin-containing protein	Hma2.217613	2.00E-12	Hma2.217613	1.00E-12	Hma2.233869	2.00E-06	MAM, P-type trefoil	13.7	22.8	Hma2.217613
Glu-rich protein	No hit	No hit	Hma2.222848 (fragment)	4.00E-07	Hma2.230913	1.00E-04	-	-	-	No hit
Cephalotoxin-like protein	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
Uncharacterized SOMP-2	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
Uncharacterized SOMP-3	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
Galaxin 2	Hma2.228867	2.00E-12	Hma2.228867	2.00E-12	Hma2.228867	1.00E-17	SP	19.9	26.5	Hma2.228867
PKD1-related protein	Hma2.221316	2.00E-40	Hma2.221316	6.00E-44	Hma2.221316	4.00E-28	GPS, Lipoxigenase LH2, Polycystin cation channel PKD1/PKD2	12.3	19.8	Hma2.221316
Zona pellucida domain-containing protein	Hma2.216869	6.00E-13	Hma2.216869	3.00E-13	Hma2.216869	2.00E-05	SP, ZP	12.2	22.2	Hma2.216869
Uncharacterized SOMP-4	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
Galaxin	Hma2.228867	5.00E-21	Hma2.228867	5.00E-21	Hma2.228867	3.00E-23	SP	22.4	35.1	Hma2.228867
EGF and laminin G domain-containing protein	Hma2.230285	3.00E-63	Hma2.230285	3.00E-69	Hma2.230276	1.00E-31	SP, LamG, EGF-like	24.3	40.7	Hma2.230285
Putative carbonic anhydrase	Hma2.218404	3.00E-11	Hma2.218404	2.00E-11	Hma2.218404	5.00E-11	alpha-CA	-	-	Hma2.218404
Protocadherin-like	Hma2.217969	1.00E-179	Hma2.217969	0	Hma2.217969	0	SP, Cadherin	19	30.2	Hma2.217969
Collagen alpha-1 chain	Hma2.232959	1.00E-13	Hma2.232959	9.00E-16	Hma2.232959	2.00E-07	SP	15.3	26.1	Hma2.232959
CUB and peptidase domain-containing protein 1	Hma2.232959	2.00E-08	Hma2.232959	2.00E-10	Hma2.232959	0.006	SP	15.1	24.7	Hma2.232959

## Appendix B

CUB and peptidase domain-containing protein 2	Hma2.232959	2.00E-16	Hma2.232959	6.00E-17	Hma2.232959	2.00E-13	SP	15.1	28.8	Hma2.232959
Uncharacterized SOMP-5	Hma2.205838	1.00E-41	Hma2.205838	1.00E-41	Hma2.205838	3.00E-60	Thrombospondin, type 1 repeat	16.5	23.6	Hma2.205838
Neuroglian-like	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
Uncharacterized SOMP-6	No hit	No hit	No hit	No hit	No hit	No hit	-	-	-	No hit
Uncharacterized SOMP-7	Hma2.228261	4.00E-03	Hma2.228261	8.00E-04	Hma2.228261	9.00E-05	SP	19.1	29.8	Hma2.228261

Table 4: Results from the comparison of the domains from *Acropora millepora* SOMPs versus those identified in other skeletal proteomes from *Strongylocentrotus purpuratus* (tooth, spicules, test and spine) [169,171], *Gallus gallus* (eggshell) [172,173,183], *Lottia gigantea* (shell) [174], *Pinctada margaritifera* and *P. maxima* (shell) [54], *Stylophora pistillata* [156] and *Crassostrea gigas* (shell) [175]. + indicates domains from proteins that were identified through proteomics and are expressed in skeleton secreting-tissues, or have further experimental evidence of involvement in biomineralization, (+) indicates domains from proteins identified in the organic matrix only by proteomics but for which no other evidence related to biomineralization is currently available. \* Domains corresponding to more than one InterPro entry (*i.e.* with parent/child relationship), <sup>a</sup> Domains identified only in corals and <sup>b</sup> Databases containing intracellular proteins.

<i>Acropora millepora</i>		Versus species:	<i>S. purpuratus</i> <sup>b</sup>	<i>G. gallus</i> <sup>b</sup>	<i>L. gigantea</i>	<i>P. margaritifera</i> <i>P. maxima</i>	<i>S. pistillata</i> <sup>b</sup>	<i>C. gigas</i> <sup>b</sup>
Key domains (as in Figure 4)	InterPro entries identified in the SOMPs	Structure:	Tooth, spicules, test and spine	Eggshell	Shell	Shell	Skeleton	Shell
		Interpro no:						
Thrombospondin	Thrombospondin, type 1 repeat	IPR000884	(+)	+	-	-	(+)	(+)
Nidogen	Nidogen, extracellular domain	IPR003886	(+)	(+)	-	-	-	(+)
AMOP	AMOP	IPR005533	(+)	-	-	-	(+)	(+)
von Willebrand factor, type D	von Willebrand factor, type D domain	IPR001846	(+)	(+)	-	-	(+)	(+)
von Willebrand factor, type A	von Willebrand factor, type A	IPR002035	(+)	(+)	+	+	(+)	+
Epidermal growth factor*	Epidermal growth factor-like domain	IPR000742	(+)	+	+	+	(+)	+
	EGF-like calcium-binding	IPR001881	(+)	+	-	-	(+)	(+)

## Appendix B

Coagulation factor 5/8 CT type domain*	Coagulation factor 5/8 C-terminal type domain	IPR000421	+	(+)	-	-	(+)	(+)
CAP	Galactose-binding domain-like	IPR008979	+	(+)	-	-	(+)	(+)
MAM domain*	CAP domain	IPR014044	(+)	-	+	+	-	(+)
Ricin B lectin domain	MAM domain	IPR000998	(+)	(+)	-	-	(+)	(+)
Fibronectin type III*	Concanavalin A-like lectin/glucanase	IPR008985	(+)	(+)	-	+	(+)	+
	Ricin B lectin domain	IPR000772	-	(+)	-	-	-	-
	Fibronectin, type III	IPR003961	+	+	-	+	-	+
ZP sperm-binding	Fibronectin type III C-terminal domain <sup>a</sup>	IPR026966	-	-	-	-	-	-
CUB	Zona pellucida sperm-binding protein	IPR001507	-	(+)	+	+	(+)	+
Laminin G*	CUB	IPR000859	(+)	+	+	-	-	(+)
Carbohydrate-binding WSC*	Laminin G domain	IPR001791	(+)	(+)	-	-	(+)	(+)
PKD/Chitinase domain*	Concanavalin A-like lectin/glucanase, subgroup	IPR013320	(+)	+	-	+	(+)	+
PKD/REJ-like protein*	Carbohydrate-binding WSC	IPR002889	(+)	-	-	-	-	-
GPS	Carbohydrate-binding WSC, subgroup <sup>a</sup>	IPR013994	-	-	-	-	-	-
Cadherin*	PKD domain	IPR000601	-	(+)	-	-	-	(+)
P-type trefoil <sup>a</sup>	PKD/Chitinase domain <sup>a</sup>	IPR022409	-	-	-	-	-	-
Fibrillar collagen, CT	PKD/REJ-like protein	IPR002859	-	-	-	-	-	(+)
Collagen triple helix repeat	Egg jelly receptor, REJ-like <sup>a</sup>	IPR014010	-	-	-	-	-	-
Immunoglobulin-like*	GPS domain	IPR000203	(+)	-	-	-	-	-
Low-density lipoprotein receptor	Cadherin	IPR002126	+	+	-	-	(+)	(+)
Cupredoxin*	Cadherin-like	IPR015919	(+)	+	-	-	(+)	(+)
	P-type trefoil <sup>a</sup>	IPR000519	-	-	-	-	(+)	-
	Fibrillar collagen, C-terminal	IPR000885	+	+	-	-	-	-
	Collagen triple helix repeat	IPR008160	+	+	-	-	-	(+)
	Immunoglobulin subtype 2	IPR003598	+	(+)	-	-	-	+
	Immunoglobulin subtype	IPR003599	+	(+)	-	-	-	+
	Immunoglobulin-like	IPR007110	+	+	-	-	-	+
	Immunoglobulin I-set	IPR013098	+	(+)	-	-	-	+
	Immunoglobulin-like fold	IPR013783	+	+	-	+	-	+
	Low-density lipoprotein (LDL) receptor class A repeat	IPR002172	(+)	(+)	-	-	(+)	-
	Lipoxygenase, LH2 <sup>a</sup>	IPR001024	-	-	-	-	-	-
	Lipase/lipoxygenase, PLAT/LH2 <sup>a</sup>	IPR008976	-	-	-	-	-	-
	Cupredoxin	IPR008972	(+)	-	-	-	-	+

## Appendix B

	Multicopper oxidase, type 2	IPR011706	-	-	-	-	-	+
	Multicopper oxidase, type 3	IPR011707	(+)	-	-	-	-	+
Alpha carbonic anhydrase	Alpha carbonic anhydrase	IPR001148	+	+	+	+	+	+
Peptidase cysteine/serine, trypsin-like*	Peptidase S1/S6, chymotrypsin/Hap	IPR001254	(+)	+	-	-	-	+
Polycystin cation channel, PKD1/PKD2	Peptidase cysteine/serine, trypsin-like	IPR009003	(+)	(+)	-	-	-	+
Neurexin/syndecan/glycophorin C	Polycystin cation channel, PKD1/PKD2	IPR013122	-	-	-	-	-	(+)
Cadherin, cytoplasmic domain	Neurexin/syndecan/glycophorin C	IPR003585	(+)	-	-	-	-	-
	Cadherin, cytoplasmic domain	IPR000233	-	(+)	-	-	-	(+)

## Appendix B

**Table 5: Comparison between *Acropora millepora* (AM) SOMPs and the proteins identified in the skeletal organic matrix from *Stylophora pistillata* (SP) [156]. Pairs of related proteins are indicated by x – for more than 35% of identity (min. 100 aa) and by X – for homologous pairs. Homology could not be determined for protein fragments (\*).**

AM \ SP	Protocadherin fat-like (P1)	CARP4 (P2)	Thrombospondin (P3)*	Viral inclusion protein (P4)	Hemicentin (P5)*	Actin (P6)	Actin (P7)*	Major yolk protein (P8)*	Protocadherin fat-like (P9)*	Cadherin (P10)*	Actin (P11)*	Unknown protein (P12)	Sushi domain-containing (P13)*	Collagen-alpha (P14)*	CARP5 (P15)*	Unknown protein (P16)*	Glyceraldehyde 3-phosphatase dehydrogenase (P17)*	Collagen alpha (P18)*	Contactin-associated protein (P19)*	MAM domain anchor protein (P20)*	Zona pellucida (P21)*	Unknown protein (P22)	Protocadherin (P23)*	Vitellogenin (P24)*	Ubiquitin (P25)*	Vitellogenin (P26)*	Integrin-alpha (P27)*	Late embryogenesis protein (P28)*	Tubulin-beta (P29)*	Myosin regulatory light chain (P30)*	Neurexin (P31)*	Kielin/Chordin like (P32)*	Flagellar associated protein (P33)*	MAM/LDL receptor domain containing protein (P34)*	Carbonic anhydrase (STPCA2) (P35)*	Zonadhesin-like precursor (P36)*					
SAARP 1		X													x																										
Acidic SOMP		X													x																										
SAARP2*		X													x																										
SAP1*																																									
SAP2*																																									
Glu-rich protein																																									
Mucin-like*						x							x																												
Coadhesin*			x		x									x																											
MAM and LDL-receptor domain-containing protein 1*																				x																					
MAM and																				x																					





