The Occurrence of Type S1A Serine Proteases in Sponge and Jellyfish

Ana Rojas and Russell F. Doolittle

Center for Molecular Genetics
University of California, San Diego
La Jolla, California 92093-0634

Correspondence to R. F. Doolittle
Center for Molecular Genetics
University of California, San Diego
La Jolla, California 92093-0634

Tel. (858) 534-4417
Fax (858) 534-4985
email rdoolittle@ucsd.edu

Abstract. Although serine proteases are found in all kinds of cellular organisms and many viruses, the classic "chymotrypsin family" (Group S1A by the 1998 Barrett nomenclature) has an unusual phylogenetic distribution, being especially common in animals, entirely absent from plants and protists, and rare among fungi. The distribution in Bacteria is largely restricted to the genus *Streptomyces*, although a few isolated occurrences in other bacteria have been reported. The family may be entirely absent from Archaea. Although more than a thousand sequences have been reported for enzymes of this type from animals, none of them have been from early diverging phyla like Porifera or Cnidaria. We now report the existence of Group S1A serine proteases in a sponge (phylum Porifera) and a jellyfish (phylum Cnidaria), making it safe to conclude that all animal groups possess these enzymes.

The origin of the "chymotrypsin family" of serine proteases (E.C. 3.4.21.-, or Group S1A, Barrett et al, 1998) in animals poses a perplexing problem in phylogenetics. There are literally hundreds of different enzymes from this group among invertebrate and vertebrate animals, but members of the family are not found at all among plants or protists. In fungi they are restricted to a single small group of ascomycetes, and in Bacteria, with only a few exceptions, they are restricted to the genus *Streptomyces*. These enzymes, which are mostly extracellular and contain three or more disulfide bonds, are extremely abundant in both vertebrate and invertebrate animals.

Although these enzymes have been reviewed extensively in the past (e.g., Lesk and Fordham, 1996), the issue of where and how the animal cohort of enzymes originated has not been specifically addressed in recent times. The current version of the MEROPS database contains about a thousand entries for such enzymes from 48 different animal species. Of these, about 200 are from two dozen different invertebrate species. None of these, however, are from early diverging phyla like Porifera and Cnidaria. Given the overall sporadic occurrence of these enzymes and the frequent invocation of horizontal gene transfer in explanation (e.g., Screen et al, 2000), we felt it important to find if sponges and jellyfish possess these enzymes, and if so, where their sequences fall in a general phylogeny.

We isolated total RNA (Chomczynski and Sacchi, 1987) from a sponge (*Verongia aurea*) and a jellyfish (*Aurelia aurita*) and prepared cDNA, first in single-stranded form and then double-stranded (Chenchick et al, 1996). We constructed sets of degenerate primers based on extensive alignments of genes coding for this kind of protease among various invertebrate animals. These primers were used in conjunction with sets of universal primers with adapters that facilitated the

isolation of the appropriate messages (Matz et al, 1999a; 1999b). Extensive use was made of kits from Clonetech, Invitrogen and Quiagen. PCR was conducted on a Perkin Elmer thermocycler. *Drosophila melanogaster* ds-DNA was used as a positive control to test the effectiveness of the primers.

Bands of DNA of the expected lengths (600-700 nt) were extracted from agarose gels after electrophoresis and subcloned into appropriate vectors. Multiple transformants were picked and submitted for DNA sequencing at the UCSD campus sequencing facility. In the end, sequences corresponding to serine proteases were isolated from each of the three kinds of starting DNA, including an already identified serine protease from fruitfly (GenBank Accession AAL 49280), a trypsin-like enzyme from sponge, and two serine proteases from jellyfish. In the cases of the sponge and jellyfish, sequences were confirmed by the use of different sets of primers that yielded overlapping regions. The single sequence identified in sponge encompassed 243 amino acid codons. Of the two isolated from jellyfish, one spanned 232 residues, but in the other the sequence did not extend beyond the 5' degenerate primers, and only 183 codons were obtained. The two jellyfish sequences differed at 53 of the 185 comparable positions (71 percent identical at the amino acid level). The sponge and one of the jellyfish sequences were confirmed by constructing exact primers from them, re-isolating bands from new cDNA from separate specimens, and determining the sequences again.

The three sequences have been submitted to GenBank; jellyfish serine protease 1, AF486486; jellyfish serine protease 2, AF486487; sponge trypsin, AF486488. An alignment of the sponge and jellyfish sequences with those from several other animals is shown in Fig. 1.

A phylogenetic tree was constructed with these sequences and a representative set from other animals, both vertebrate and invertebrate, as well as several from bacteria and fungi (Fig. 2). Both the sponge and jellyfish sequences appear relatively high on the tree; certainly neither is at the animal root. Interestingly, the two enzymes appear in different major groups, the sponge clustering among the digestive enzymes most often labeled trypsins, and the two jellyfish sequences with the chymotrypsin-elastase group. The sponge sequence also had an aspartic acid at "position 189," a diagnostic residue that occurs in all known trypsins (Hannenshalli and Russell, 2000). In contrast, the two jellyfish sequences have an asparagine at that position. The sponge and jellyfish enzymes also differ with regard to the codon employed for the serine at the active site, the sponge having a typical trypsin signature with a TCT codon, whereas one jellyfish sequence has the AGT codon and the other AGC. When the new sequences were searched against GenBank with Blast (Altschul et al, 1990), the highest scoring match for the sponge sequence was an arthropod trypsin (48 percent identity), whereas the best match for either of the jellyfish sequences was rat elastase (36 pecent identity). The jellyfish enzymes are the first S1A serine proteases from an invertebrate animal with five disulfide bonds, all the cysteines of which match vertebrate elastase and chymotrypsin exactly.

Although the S1A serine proteases from animals cover a wide variety of functional designations and are involved in physiological processes ranging from blood clotting to moulting, the largest number of reported enzymes have to do with feeding and digestion, including about 100 entries labeled "trypsin." Jellyfish (Cnidaria) have genuine organs and tissues, including a mouth and gastrovascular cavity. They have a well developed digestive apparatus that allows for the capture, swallowing and digestion of food, and we fully expected they would have digestive enzymes. Sponges, on the other hand, are loose colonies of cells that are filter

feeders. They lack a true body cavity. The surrounding water and small particulate matter pass through pores that perforate special cells called porocytes into an inner chamber, from whence the flow is guided to an exit called the osculum by flagellated cells called choanocytes that line the chamber. Presumably much of the digestion is extracellular, but little is known of the process. The existence of a typical trypsin-like enzyme supports that notion.

The origin of animal serine proteases remains enigmatic. If they are the result of vertical descent in a conventional manner, then it must be presumed that protists, plants and most fungi have lost these genes. With regard to the fungi, good cases have been made for the transfer of genes between ascomyctes and bacteria of the genus *Streptomyces* in both directions. Thus, the ascomycete *Fusarium oxysporum*, which has an animal-like trypsin (Rypniewski et al 1993), also has a respiratory system which looks to have been acquired from *Streptomyces* (Takaya et al 1998). But another ascomycete has a chymotrypsin which appears to have been transferred in the other direction, i.e., from the fungus to *Streptomyces* (Screen et al, 2000). We plan to address these apparently contradictory observations in a separate publication devoted strictly to phylogenetic analysis (Rojas and Doolittle, in preparation)

REFERENCES:

Altshul SF, Gish W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Barrett AJ, Rawlings ND, Woessner JF (1998) Handbook of Proteolytic Enzymes. Academic Press, San Diego.

Chenchick A, Diachenko L, Moqadam F, Tarabykin V, Lukyanov S, Siebert PD (1996) Full length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. Biotechniques 21:526-534.

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem 162, 156-159.

Feng DF, Doolittle RF (1996) Progressive alignment of protein sequences and construction of phylogenetic trees from them. Meth. Enzymol. 266, 368-382.

Hannenhalli SS, Russell RB (2000) Analysis and prediction of functional sub-types from protein sequence alignments. J. Mol. Biol. 303:61-76.

Lesk A. Fordham WD (1996) Conservation and variability in the structures of serine proteases of the chymotrypsin family. J. Mol. Biol. 258:501-537.

Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L Chenchick A (1999a) Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic Acids Res. 27, 1558-1560.

Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, Lukyanov SA (1999b). Fluorescent proteins from non-bioluminescent Anthozoa species. Nature Biotechnology 17:969-973.

Penalva MA, Moya A, Dopazo J, Ramon D (1990) Sequences of isopenicillin N synthetase genes suggest horizontal gene transfer from prokaryotes to eukaryotes. Proc Royal Soc Lond. [B] 241:161-169.

Rypniewski WR, Hastrup S, Betzel Ch, Dauter M, Deuter Z, et al. (1993) The sequence and X-ray structure of trypsin from *Fusarium oxysporum*. Prot. Eng. 6:341-348.

Screen SE, St. Leger RJ (2000) Cloning, expression and substrate specificity of a fungal chymotrypsin. Evidence for a horizontal transfer from an actinomycete to a bacterium. J. Biol. Chem. 275:6689-6694.

Takaya N, Kobayashi M, Shoun H (1998) Fungal denitrification, a respiratory system possibly acquired by horizontal gene transfer from prokaryotes. In <u>Horizontal Gene Transfer</u> (M. Syvanen and C. Kado, eds.) pp. 321-327. Chapman and Hall, London.

FIGURE LEGENDS

Figure 1. Alignment of serine protease sequences from the sulfur sponge (*Verongia aurea*) and moon jellyfish (*Aurelia aurita*) with a representative set of S1A sequences from vertebrate and invertebrate animals. Cysteine residues are colored yellow, except for two in thejellyfish enzyme that coincide exactly with vertebrate elastases and chymotrypsins but which have not been found elsewhere among the invertebrate animals.

Figure 2. Phylogenetic tree constructed from 34 S1A serine protease sequences including trypsin from the sulfur sponge (*Verongia aurea*) and a serine protease from a moon jellyfish (*Aurelia aurita*) and a representative set of serine protease sequences from vertebrate and invertebrate animals, as well as three from fungi and one from a bacteriu. A single type A1E enzyme (alkaline serine protease from *Streptomyces sp.*) was used as an outlier for rooting purposes. The tree was made by the method of Feng and Doolittle (1996).

Protein sequence alignment

```
Sponge trypsin        :IVGGDPVNKGYVPMQVSLQ-REGFFGRSH--FCGGSILDADTVLTAAHCTDGQ---VPSG-ITVVAGDHVLSTTDGDEQVGVASISE : 81
Shrimp trypsin        :IVGGTDATPGELPYQLSFQ-DISFGFAWH--FCGASIYNENWAICAGICVQGEDMNNPDY-LQVVAGELNQDVDEGTEQTVILSKIIQ : 84
Human trypsin         :IVGGYNCEENSVPYQVSL---NSGY---H--FCGGSLINEQWVVSAGHCYK------SR-IQVRLGEHNIEVLEGNEQFINAAKIIR : 72
Fruit fly trypsin     :IVGGSATTISSFPWQISLQ-RSG----SH--SCGGSIYSANIIVTAAHCLQSV--SASV-LQVRAGSTYWS-SGG-VVAKVSSFKN : 74
Human chymotrypsin    :IVNGEDAVPGSWPWQVSLQDKTGF----H--FCGGSLISEDWVTAAHC--GV--RTSD-V-VVAGEFDQGSDEENIQVLKIAKVFK : 75
Shrimp chymotrypsin   :IVGGVEATPHSWPHQAAL-----FIDDMY--FCGGSLISSEWVLTAAHCMDG-----AGF-VEVVLGAHNIRQNEASQVSITSTDFFT : 75
Rat elastase          :VVGGQEASPNSWPWQVSLQ-YLSSGKWHH--TCGGSLVANNWVLTAAHCIS-----NSRT-YRVLLGRHSLSTSESGSLAVQVSKLVV : 79
Moon jelly protease 1 :IISGTNARPGAWPWMASLY-----MLSRSH-ICGGSLLNSRWILTASHCVVGT-GATTKN-LVIKLGEHDEYDKDGFEQQFDVEKIIP : 80
Horseshoecrab factor G:IIGGIATPHSWPWMVGI------FKVNPHRFLCGGSIINKVSVTAAHCLVTQFGNRQNYSIFVRVGARDI---DNSGTNYQVDKVIV : 80
Mouse factor 7        :IVGGNVCPKGECPWQAVLK-INGLL------LCGAVLLDARWIVTAAHCFDNI---RYWGNITVVMGEHDFSEKDGDEQVRRVTQVIM : 78
                                 G           P                CG         A HC                          G
```

```
                                                     100
Sponge trypsin        :HPEYNSRTFY--NDICVLKLNSIIGGNVQPVGLPFPN-A-EVDEGV-MATVSGWG--TTSAGGSLSDVILAVNVPVISDAECRGA-- :161
Shrimp trypsin        :HEDYNGFTIS--NDISLLKLSQPLSFNDNVRAIDIP-AQ-G-HAASG--DCIVSGWG--TTSEGGSTPSVLQKVTVPIVSDDECRDA-- :162
Human trypsin         :HPQYDRKTLN--NDIMLIKLSSRAVINARVSTISLP--T-A-PPATGT-KCLISGWGNTASSGADYPDELQCLDAPVLSQAKCEAS-- :151
Fruit fly trypsin     :HEGYNA-TMV--NDIAVIRLSSSLSFSSSIKAISL--AT-Y-NPANGA-SAAVSWGCTQSSGSGSSIIPSQLQYVNVNIVSQSQCASS-- :152
Human chymotrypsin    :NPKFSILTVN--NDITLLKLATPARFSQTVSAVCLPSAD-D-DFPAGT-LCATTGWGKTKYNANKTPDKLQQAALPLLSNAECKKS-- :156
Shrimp chymotrypsin   :HENWNSWLLT--NDIALIRLPSPVSLNSNIKTVKLP--S-S-DVSVGT-TVTPTGWGRPSDSASGISDVLRQVNVPVMTNADC-DS-- :153
Rat elastase          :HERWNAQKLSNGNDIALVKLASPVALTSKIQTACLP-PA-G-TILPNNYPCYVTGWGRLQTNGA-TPDVLQQGRLLVVDYATCSSA-- :161
Moon jelly protease 1 :HPAYKRGPLK--NDIALIKLKTPARINKRVKTICLPKRG-S-APSVGSRECHLAGWGSIRHPGGSY-HTLQQAMLPVVSYTNCHNQ-- :161
Horseshoecrab factor G:HQGYKHHSHY--YDIGLILSKPVEYNDKIQPVCIPEFN-KPHVNLANNIKVVITGWG--VTGKATEKRNVIRELELPVVTNEQCNKSYQ :164
Mouse factor 7        :PDKYIRGKIN--EDIALLRLHRPVTFTDYVVPLCLPEKSFSENTLARIRFSRVSGQLLDRGA-TALEDMSIEVPRLMTQDCLEH-- :161
                               DI L                                    GWG              L
```

```
                           200
Sponge trypsin        :---YGETD--VADSMICAGDLANGGIDSCQGDSGGPLYM-G-ST----IIGIVSWGY--GCAYAGYPGVYTQVSYYVSFIKS----- :230
Shrimp trypsin        :---YGQSD--IEDSMICAG-VPEGGKDSCQGDSGGPLAC-S-DYASTYLAGIVSWGY--GCARPGYPGVYAEVSYHVDWIKANAV-- :237
Human trypsin         :---Y-PGK--ITSNMFCVGFL-EGGKDSCQGDSGGPVVC-N-GQ---LQGVVSWGD--GCAQKNKPGVYTKVYNYVKWIKNTIAANS :224
Fruit fly trypsin     :--TYGYGS--QINTMICAA---ASGKDACQGDSGGPLVS-G-GV---LVGVVSWGY--GCAYSNYPGVYADVAVLRSWVVSTANSI- :224
Human chymotrypsin    :--WGR-R--ITDVMICAG---ASGVSSCMGDSGGPLVC-Q-KDGAWTLVGIVSWGSDT-C-STSSPGVYARVTKLIPWVQKILAAN- :230
Shrimp chymotrypsin   :--VYG--I--VGDGVVCID--GTGGKSTCNGDSGGPLNL-N-GM----TYGITSFGSSAGC-EKGYPAAFTRVYYILDWIQQKTGVTP :226
Rat elastase          :--SWNGSS--VKTNMVCAG--GDGVTSSCNGDSGGPLNCQA-SNGQWQVHGIVSFGSTLGCNYPRKPSVFTRVSNYIDWINSVLAKN :241
Moon jelly protease 1 :-----------KNFVCAGFGKSSLTNACRGDSGGPLA-CRK-SDGSWEQHGIASFV-VEYC-K--YYTAFTPVANYIDWINQHINK-- :230
Horseshoecrab factor G:TLPFSKLNRGIITNDMICAGF-PEGGKDACQGDSGGPLMYQNPTTGRVKIVGVVSFGF-E-CARPNFPGVYTRLSSYVNWLQEITFGQ- :248
Mouse factor 7        :AKHSSNTPKITENMFCAGY-MDGTKDACKGDSGGPHATH--YHGTWYLLGVVSWG--EGCAAIGHIGVYTRVSQYIDWLVRHMDSK- :242
                                        C         C GDSGGP                G   S    C
```

Fig.

Human
Chicken
Fish
Ascidian
Mosquito
Shrimp
Crayfish
Scolopendra
SPONGE
Fungi 1
Fungi 2
Fungi 3
Fruit fly
Blow fly
Strep (bacteria)

TRYPSIN

Human
Fish
Fruit fly stubble
Shrimp
Crab
Mosquito
Rat elastase
JELLY FISH 1

CHYMOTRYPSIN

Horseshoecrab factor G
Horseshoecrab factor B
Fruit fly Nudel
Human thrombin
Mouse factor 7
Insect pro-phenol oxidase protein
Fruit fly Easter protein
Hornworm pro-phenol oxidase protein
Hornworm Snake protein
Fruit fly Snake protein
Bacterial alkaline protease

Fig -