COMMENTARY

Unusual phyletic distribution of peptidases as a tool for identifying potential drug targets

Neil D. RAWLINGS¹

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, U.K.

Eukaryote homologues of carboxypeptidases Taq have been discovered by Niemirowicz et al. in the protozoan *Trypanosoma cruzi*, the causative agent of Chagas' disease. This is surprising, because the peptidase family was thought to be restricted to bacteria and archaea. In this issue of the *Biochemical Journal*, the authors propose that the *Trypanosoma* carboxypeptidases are potential drug targets for treatment of the disease. The authors also propose that the presence of the genes in the zooflagellates can be explained by a horizontal transfer of an ancestral gene from a prokaryote. Because peptidases are popular drug targets, identi-

fying parasite or pathogen peptidases that have no homologues in their hosts would be a method to select the most promising targets. To understand how unusual this phyletic distribution is among the 183 families of peptidases, several other examples of horizontal transfers are presented, as well as some unusual losses of peptidase genes.

Key words: carboxypeptidase, Chagas' disease, drug target, horizontal gene transfer, *Trypanosoma*.

In this issue of *Biochemical Journal*, Niemirowicz et al. [1] report the discovery of novel peptidases from the protozoan *Trypanosoma cruzi*, the causative agent of Chagas' disease in Latin America. Niemirowicz et al. [1] have been able to show that two genes are present in *T. cruzi*, encoding metal-dependent carboxypeptidases with different specificities and distributions during the protozoan life cycle. Homologues have also been found in the closely related protozoan, *Leishmania major*.

What makes this discovery interesting is that the carboxypeptidases in question are from a family of proteins that are otherwise unknown in eukaryotes. An enzyme that is present in a parasite which is not only absent in its hosts, but its hosts have no homologues, makes the enzyme an ideal drug target, and Niemirowicz et al. [1] suggest that the carboxypeptidases might be potential drug targets for the prevention of Chagas' disease.

The *Trypanosoma* and *Leishmania* peptidases are members of family M32. All characterized members of the family are carboxypeptidases, and the best known member of the family is carboxypeptidase Taq from *Thermus aquaticus*. Apart from the zooflagellate protozoa, the family is restricted to bacteria and archaea. The surprisingly high similarity between the *Trypanosoma* and *Leishmania* carboxypeptidases and those from bacteria has led Niemirowicz et al. [1] to suggest that an ancient horizontal transfer of a gene took place, either after the separation of zooflagellates from other protozoa or the gene was lost in the ancestors of other protozoa.

Peptidases are increasingly being seen as drug targets, and a recent estimate was that 14 % of human peptidases are under investigation by pharmaceutical companies [2]. This follows on from the success with synthetic inhibitors of retropesin, the polyprotein-processing peptidase essential to the replication of HIV [3]. Chagas' disease is estimated to affect 16 to 18 million people, and although this heart disease is frequently chronic, there are 21 000 deaths a year. The most promising drug targets so far had seemed to be cruzipain, the cysteine endopeptidase that may be responsible for host tissue invasion by the protozoan [4], and its

endogenous inhibitor chagasin [5], although attention has also turned to a thrombin inhibitor from the insect vector [6].

Although Niemirowicz et al. [1] may have identified an alternative point of attack, characterization of the carboxypeptidases is a long way from establishing a promising drug target. First of all, it is not known what the physiological functions of the carboxypeptidases are, and whether the protozoan can survive without these gene products. Even if deleting one or both genes proves to be lethal, then finding an inhibitor that can be used as a drug is a major task. There are few known natural inhibitors of carboxypeptidases and synthetic inhibitors have been developed against the unrelated human enzymes such as carboxypeptidase A. Most protein inhibitors either block binding of the C-terminus of the substrate protein, or chelate the catalytic zinc ion. So although the *Trypanosoma* carboxypeptidases are unrelated to the carboxypeptidase A family, inhibitors would have to be very carefully designed not to affect the host peptidases because the reaction mechanisms are so similar. Other than metal chelators, no inhibitors of carboxypeptidases from family M32 are known.

However, the principle that a pathogen's 'Achilles heel' might be peptidases that have been acquired by horizontal gene transfer is a valid one. An endopeptidase from a family not present in the host might be an easier target to work with. With this thought in mind, I looked through the MEROPS database for peptidases [7] with the intention of finding peptidases with unusual distributions amongst organisms, to find other examples of horizontal gene transfers.

Most surprisingly, only one peptidase family has representatives from every completed genome: M24. This family includes methionyl aminopeptidase, which has a fundamental role to play during synthesis of cytoplasmic proteins: namely, to remove the initiating methionine from the N-terminus. There are several peptidase families containing proteins from only one superkingdom of organisms, which presumably evolved after the divergence of the superkingdoms. Family A8 is one of the 35 families with only bacterial members; this family includes signal peptidase II,

¹ email ndr@sanger.ac.uk

which removes the signal peptide from premurein, the lipoprotein component of bacterial cell walls. The only family restricted to archaea is A5, which includes thermopain, an endopeptidase active at very low pH and high temperatures. Family A1, which includes the enzymes pepsin from the stomach and cathepsin D from the lysosome, is one of the sixteen families found only in eukaryotes. There are 45 peptidase families restricted to viruses. Although the majority of peptidase families have members from all three superkingdoms, there are many families in which the distribution is discontinuous, so that only a few examples are known from one of the superkingdoms (such as the *Trypanosoma* and *Leishmania* members of M32), or where there are surprising absences.

There is another example of a horizontal gene transfer in the zooflagellates. Within chromosome 13 of *L. major* there is a gene for a member of peptidase family S15 (UniProt accession no. Q4Q871), another family from which all other known members are from bacteria, and the family is typified by Xaa-Pro dipeptidyl peptidase from *Lactococcus lactis* [8]. Again, a homologue is also known in *T. cruzi* (UniProt accession no. Q4DBP7).

Family M12 is virtually confined to eukaryotes, except for flavastacin from *Chryseobacterium meningosepticum* [9]. Family S12 includes serine carboxypeptidases that process the precursors of the peptides which help form the network of cross-links in the glycoprotein murein, the major component of the bacterial cell wall. The cross-linking peptides are synthesized as precursors with an extra D-alanine residue at the C-terminus, which is released by the carboxypeptidases. D-Alanine is not found in eukaryotes, so it was something of a surprise to discover a D-Ala-D-Ala carboxypeptidase homologue in the genome of *Caenorhabditis elegans* (UniProt accession no. Q09621). This nematode feeds on soil-borne bacteria, so perhaps the presence of this peptidase helps it digest its food. But a similar gene also exists in the human genome, and although it has been localized to the mitochondrion, at the moment no function has been ascribed to it [10]. Family S1 contains both eukaryotic and bacterial members, but the differences between them are so great that the members from each superkingdom were consigned to separate subfamilies. Subfamily S1B contained only bacterial endopeptidases such as glutamyl endopeptidase and the membrane-bound protease Do; at least, that was the situation until human homologues of protease Do were discovered. One of these has been characterized, and is found in fibroblasts and osteoblasts and is known to degrade the C99 peptide from amyloid β -A4 protein [11]. A homologue of the 'eukaryote' subfamily, S1A, which includes trypsin and chymotrypsin, has been found in the *Vibrio cholerae* genome (UniProt accession no. Q9KLE3), and subsequently other *Vibrio* genomes.

It had been thought that horizontal gene transfers did not occur in archaea, until the genome of the bacterium *Thermotoga maritima* was sequenced and was shown to have acquired genes from an archaeon [12]. Horizontal transfer in the opposite direction appears to have occurred in the huge genome of *Methanosarcina acetivorans* [13], including the only archaean genes for homologues of thermolysin (peptidase family M4) and clostripain (peptidase family C11): both peptidase families are otherwise only known from bacteria. The closely related *M. barkeri* is equally a 'kleptomaniac' for foreign genes [14], but this genome contains a real surprise. Multicellular eukaryotes have to kill their own cells during development and to recycle tissues. This is done by a highly controlled process known as apoptosis or programmed cell death. There is a proteolytic cascade involving peptidases known as caspases (family C14) that selectively cleave after aspartate residues, either to activate other caspases in the cascade or to degrade cellular proteins [15]. The peptidase family is restricted

to eukaryotes, except for a homologue in *M. barkeri* (hypothetical protein Mbar_A1574). If the gene had been acquired by horizontal transfer, then its origin certainly was not from a thermophilic bacterium.

Horizontal transfers of genes are not restricted to cellular organisms. Viruses, particularly DNA viruses, have acquired peptidase genes from their hosts. One of the first to be discovered was of a papain-like endopeptidase (family C1) within the *Autographa californica* baculovirus genome [16], and others include a subtilisin homologue in the ictalurid herpesvirus 1 (UniProt accession no. Q00139) and a caspase (family C14) homologue in the *Spodoptera frugiperda* ascovirus 1a (UniProt accession no. Q5K5C1).

One has to be aware that any unexpected appearance of a homologue of a bacterial peptidase gene in a eukaryote genome may be the result of experimental error. The presence of LexA repressor homologue (peptidase family S24) in rat (UniProt accession no. P97845) and an aminopeptidase T (peptidase family M29) homologue in tomato (UniProt accession no. O65828) are most likely the results of bacterial contamination. Careful experimental work to confirm the presence of any unusual peptidases is necessary, such as that done by Niemirowicz et al. [1], which establishes beyond doubt that peptidase family M32 is found in prokaryotes and eukaryotes.

Absences of peptidase genes can be as difficult to explain as unexpected presence. Family S26 includes the signal peptidases that remove the N-terminal targeting signals from secreted proteins, and because this is such a fundamental process the family might be expected to be universally distributed. However, the family is absent from *Chlamydia* and *Mycoplasma* species, although these bacteria synthesize proteins with predicted signal peptides. Family S51 is typified by dipeptidase E from *Escherichia coli*, and almost all known members of the family are from bacteria. However, homologues are present in some animals: *Anopheles*, *Drosophila*, *Xenopus* and chicken. But there are no known homologues in any of the mammalian, fish or nematode genomes so far sequenced. That makes the appearance of homologues in insects, frogs and birds very difficult to explain without invoking multiple gene transfers or losses.

At the moment how and why these laterally transferred peptidase genes occur is open to speculation. Presumably, an evolutionary advantage is conveyed on the recipient. In eukaryotes the horizontally transferred genes appeared to have diverged at similar times, between 1.5 and 2 billion years ago, and we have tentatively suggested that this coincided with the origin of the mitochondria and chloroplasts within eukaryotes [17], both organelles known to be derived from bacterial endosymbionts. It was also known that genes had been transferred from the organelle genomes to the nuclear genome. Could it also be that, besides genes encoding enzymes involved in energy metabolism, other genes encoding enzymes such as peptidases were also transferred in a similar way? Whatever the mechanism, if horizontal gene transfer is common, then establishing a tree of life from sequence homologies becomes even more difficult. If the products of horizontally transferred genes prove to be useful drug targets, then the phylogenist's nightmare becomes the pharmacologist's dream.

REFERENCES

- 1 Niemirowicz, G., Parussini, F., Agüero, F. and Cazzulo, J. J. (2006) Two metallo-carboxypeptidases from the protozoan *Trypanosoma cruzi* belong to the M32 family, found so far only in prokaryotes. *Biochem. J.* **401**, 399–410
- 2 Southan, C. (2001) A genomic perspective on human proteases as drug targets. *Drug Discov. Today* **6**, 681–688

- 3 Kempf, D. J., Sham, H. L., Marsh, K. C., Flentge, C. A., Betebenner, D., Green, B. E., McDonald, E., Vasavanonda, S., Saldivar, A., Wideburg, N. E. et al. (1998) Discovery of ritonavir, a potent inhibitor of HIV protease with high oral bioavailability and clinical efficacy. *J. Med. Chem.* **41**, 602–617
- 4 Cazzulo, J. J., Stoka, V. and Turk, V. (2001) The major cysteine proteinase of *Trypanosoma cruzi*: a valid target for chemotherapy of Chagas disease. *Curr. Pharm. Des.* **7**, 1143–1156
- 5 Santos, C. C., Sant'Anna, C., Terres, A., Cunha-e-Silva Scharfstein, J. and de A Lima, A. P. (2005) Chagasin, the endogenous cysteine-protease inhibitor of *Trypanosoma cruzi*, modulates parasite differentiation and invasion of mammalian cells. *J. Cell Sci.* **118**, 901–915
- 6 Campos, I., Amino, R., Sampaio, C., Auerswald, E., Friedrich, T., Lemaire, H., Schenkman, S. and Tanaka, A. (2002) Infestin, a thrombin inhibitor presents in *Triatoma infestans* midgut, a Chagas' disease vector: gene cloning, expression and characterization of the inhibitor. *Insect Biochem. Mol. Biol.* **32**, 991
- 7 Rawlings, N. D., Morton, F. R. and Barrett, A. J. (2006) MEROPS: the peptidase database. *Nucleic Acids Res.* **34**, D270–D272
- 8 Rigolet, P., Mechin, I., Delage, M. and Chich, J. (2002) The structural basis for catalysis and specificity of the X-prolyl dipeptidyl aminopeptidase from *Lactococcus lactis*. *Structure* **10**, 1383–1394
- 9 Tarentino, A. L., Quinones, G., Grimwood, B. G., Hauer, C. R. and Plummer, Jr, T. H. (1995) Molecular cloning and sequence analysis of flavastacin: an O-glycosylated prokaryotic zinc metalloendopeptidase. *Arch. Biochem. Biophys.* **319**, 281–285
- 10 Liobikas, J., Polianskyte, Z., Speer, O., Thompson, J., Alakoskela, J. M., Peitsaro, N., Franck, M., Whitehead, M. A., Kinnunen, P. J. and Eriksson, O. (2005) Expression and purification of the mitochondrial serine protease LACTB as an N-terminal GST fusion protein in *Escherichia coli*. *Protein Expression Purif.* **45**, 335–342
- 11 Grau, S., Baldi, A., Bussani, R., Tian, X., Stefanescu, R., Przybylski, M., Richards, P., Jones, S. A., Shridhar, V., Clausen, T. and Ehrmann, M. (2005) Implications of the serine protease HtrA1 in amyloid precursor protein processing. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6021–6026
- 12 Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A. et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329
- 13 Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., MacDonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D. et al. (2002) The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* **12**, 532–542
- 14 Maeder, D. L., Anderson, I., Bretin, T. S., Bruce, D. C., Gilna, P., Han, C. S., Lapidus, A., Metcalf, W. W., Saunders, E., Tapia, R. and Sowers, K. R. (2006) The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. *J. Bacteriol.* **188**, 7922–7931
- 15 Fuentes-Prior, P. and Salvesen, G. S. (2004) The protein structures that shape caspase-activity, specificity, activation and inhibition. *Biochem. J.* **384**, 201–232
- 16 Rawlings, N. D., Pearl, L. H. and Buttle, D. J. (1992) The baculovirus *Autographa californica* nuclear polyhedrosis virus genome includes a papain-like sequence. *Biol. Chem. Hoppe Seyler* **373**, 1211–1215
- 17 Barrett, A. J. and Rawlings, N. D. (1993) Evolutionary origins of the families of peptidases in eukaryotes. In *Proteolysis and Protein Turnover* (Bond, J. S. and Barrett, A. J., eds.), pp. 78–79, Portland Press, London

Received 27 October 2006/27 November 2006; accepted 1 December 2006

Published on the Internet 21 December 2006, doi:10.1042/BJ20061613