Trypsinogen hL is not a New Member of the Human Trypsinogen Family, but a Known Mouse Ortholog

Jian-Min CHEN,^a Claude FÉREC,^a and
Miklós SAHIN-TÓTH^{*,b}

^aINSERM-EMI 0115, Génétique Moléculaire et Génétique Epidémiologique, Université de Bretagne Occidentale, Etablissement Français du Sang-Bretagne, Centre Hospitalier Universitaire; Brest 29275, France; and ^bDepartment of Molecular and Cell Biology, Boston University, Goldman School of Dental Medicine; Boston, MA 02118, U.S.A.

Received March 7, 2003; accepted April 14, 2003

An allegedly novel human trypsinogen cDNA clone termed trypsinogen hL was recently reported in the *Biological & Pharmaceutical Bulletin* (2003; 26: 361–364). This “new” trypsinogen sequence was isolated by PCR using human lung cDNAs as template and was presumed to be a new member of the human trypsinogen family. However, trypsinogen hL does not match any sequence in the human genome; whereas it is 100% identical to mouse trypsinogen 7. Thus, trypsinogen hL is not of human origin, and its cloning from a human cDNA library was clearly a result of contamination with mouse genetic material. Publication of this cloning artifact could have been avoided by a simple BLAST search of GenBank or other sequence data-banks.

Key words gene cloning; human genome; mouse; trypsinogen family

In the recently published issue of *Biological & Pharmaceutical Bulletin* (2003; 26: 361–364), Imamura and colleagues claimed that they have identified a new member of the human trypsinogen family. The new sequence was named “trypsinogen hL” because it was found to be specifically expressed in the human lung.¹⁾ This was a surprising finding since all of the human trypsinogen genes have been sequenced, and with the completion of the human genome project identification of so far undiscovered trypsinogen genes was unexpected.

The genomic structure of the human trypsinogen gene family was elucidated in 1996, owing its intercalation within the human T cell receptor (TCR) β locus, which was among the earliest targets of the human genome project.²⁾ There are a total of nine trypsinogen genes in the human genome that are separated physically into two groups. Group I (T4–T8) is located towards the 3' end of the TCR β locus and consists of five tandem 10 kb repeats within a 70 kb segment between TRBV29-1 and TRBD1 (GenBank accession number NG_001333). The sixth member (T9) of Group I trypsinogen has been translocated from chromosome 7q34 to chromosome 9p11.2 (GenBank accession number AF029308). Group II (T1–T3) is located towards the 5' end of the TCR β locus (GenBank accession number NG_001333). The two groups of trypsinogen genes tended not to exchange genetic information with each other and acted largely as separate gene families after duplication.³⁾ At the nucleotide level, there is no significant homology between these two groups, of which the group II genes represent highly diverged relic or pseudogenes.⁴⁾

Among the group I genes, T4, T8 and T9 were assigned to the three well characterized trypsinogen isoforms: cationic-, anionic-, and meso-trypsinogens. T5 and T7 are clearly pseudogenes due to the presence of frameshift or truncating mutations (GenBank accession number NG_001333). T6 was originally annotated as a functional gene,²⁾ but recently re-interpreted as an expressed pseudogene.⁵⁾ The cloned cDNAs for the three clearly functional trypsinogen isoforms are TRYI,⁶⁾ TRYII,⁶⁾ and mesotrypsinogen⁷⁾; all of which were isolated from pancreatic cDNA libraries. In addition, there is an alternatively spliced form of the mesotrypsinogen gene, trypsinogen IV,⁸⁾ which is exclusively expressed in the brain (for details see a review article⁹⁾). The originally reported TRYIII¹⁰⁾ cDNA is mesotrypsinogen with sequencing mistakes, and thus should not be referenced as a unique trypsinogen isoform.

We speculated that “trypsinogen hL” might have been inaccurately labeled as a new gene, and would represent an alternatively spliced form of one of the known trypsinogen genes. In this regard, the divergently evolved group II genes (T1–T3) and the two obvious pseudogenes T5 and T7 were excluded from consideration. As illustrated in Fig. 2,¹⁾ “trypsinogen hL” is different from the three known trypsinogen isoforms as well as the alternatively spliced trypsinogen IV. We also compared the sequence of “trypsinogen hL” with the T6 gene and found no match between them. Thus, the gene for “trypsinogen hL” is not located within the human trypsinogen locus.

Surprisingly, a search of the GenBank using the BLAST program (<http://www.ncbi.nlm.nih.gov/BLAST/>) revealed that the coding sequence of trypsinogen hL is 100% identical to that of a known mouse ortholog, trypsinogen 7 (GenBank accession number AE000663). This finding clearly indicates that cloning of the “trypsinogen hL” cDNA was a result of contamination with mouse genetic material.

In summary, “trypsinogen hL” is not a new member of the human trypsinogen family, but a known mouse ortholog. Publication of this cloning artifact could have been avoided by a simple BLAST search of the available sequence data-bases.

REFERENCES

- 1) Imamura Y., Katsu M., Sakai K., Okumura Y., Ariga H., Kido H., *Biol. Pharm. Bull.*, **26**, 361–364 (2003).
- 2) Rowen L., Koop B. F., Hood L., *Science*, **272**, 1755–1762 (1996).
- 3) Roach J. C., Wang K., Gan L., Hood L., *J. Mol. Evol.*, **45**, 640–652 (1997).
- 4) Chen J. M., Ferec C., “Nature Encyclopedia of the Human Genome,” ed. by Cooper D. N., Macmillan Publishing Ltd., Nature Publishing group, London, 2003, (<http://www.ehgonline.net>).
- 5) Chen J. M., Montier T., Ferec C., *Hum. Genet.*, **109**, 245–252 (2001).
- 6) Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T., Matsubara K., *Gene*, **41**, 305–310 (1986).
- 7) Nyaruhucha C. N. M., Kito M., Fukuoka S. I., *J. Biol. Chem.*, **272**, 10573–10578 (1997).
- 8) Wiegand U., Corbach S., Minn A., Kang J., Muller-Hill B., *Gene*, **136**, 167–175 (1993).
- 9) Chen J. M., Ferec C., *Pancreas*, **21**, 57–62 (2000).
- 10) Tani T., Kawashima I., Mita K., Takiguchi Y., *Nucleic Acids Res.*, **18**, 1631 (1990).

* To whom correspondence should be addressed. e-mail: miklos@bu.edu