Science
AAAS

Comment on "Tequila, a Neurotrypsin Ortholog, Regulates Long-Term Memory Formation in *Drosophila*"

Peter Sonderegger, *et al.*

Science **316**, 1698 (2007);
DOI: 10.1126/science.1138410

Comment on "Tequila, a Neurotrypsin Ortholog, Regulates Long-Term Memory Formation in *Drosophila*"

Peter Sonderegger[1]* and Laszlo Patthy[2]

Didelot *et al*. (Reports, 11 August 2006, p. 851) claimed that *Drosophila* Tequila (Teq) and human neurotrypsin are orthologs and concluded that deficient long-term memory after Teq inactivation indicates that neurotrypsin plays its essential role for human cognitive functions through a similar mechanism. Our analyses suggest that Teq and neurotrypsin are not orthologous, leading us to question their equivalent roles in higher brain function.

Motivated by the assumption that Tequila (Teq) is the *Drosophila* ortholog of human neurotrypsin, Didelot *et al*. (*1*) studied the involvement of Teq in associative memory by behavioral testing of genetically modified flies. Their finding that Teq inactivation results in specific defects of long-term memory formation prompted their conclusion that the neurotrypsin pathway in humans, by analogy with the Teq pathway in *Drosophila*, is essential for information processing and functional plasticity through a similar mechanism. However, our analyses indicate that neurotrypsin is not a 1:1 ortholog of Teq, which leads us to question the assertion that Teq and human neurotrypsin play equivalent roles in higher brain functions.

The contention that Teq and neurotrypsin are 1:1 orthologs means that they evolved from a single ancestral gene that was present in the last common ancestor of flies and humans and that this 1:1 relationship has not been complicated by subsequent duplications of these genes in the two lineages. The sequence comparison provided as supporting evidence in figure S1 in (*1*) is not in agreement with the author's statement that the two proteins "show a high degree of amino acid conservation, particularly in the region of the functional domains." The indicated kringle domain and two of the four indicated SRCR (scavenger receptor cysteine-rich repeat) domains are not present in the Teq sequence. Most obviously, the sequence indicated as a kringle domain of Teq does not contain a single cysteine, whereas six cysteines are highly conserved in all known invertebrate and vertebrate kringle domains . Therefore, this sequence lacks an essential structural hallmark of kringle domains (*2*). Thus, the domain organization of Teq shows no kringle and only two SRCR domains. Instead, Teq contains two low-density lipoprotein receptor class A domains, alternating with the SRCR domains, and several chitin-binding peritrophin-A domains [see (*3*) for illustration]. Overall, only two domain-types are shared by Teq and neurotrypsin: Both proteins have SRCR domains and a trypsin-like protease domain. However, these are not distinguishing features of the two proteases. A survey of multidomain trypsin protease architectures in the Pfam database (*4*) identified several other human serine proteases that share these features, including complement factor I, enterokinase, and several transmembrane proteases. Therefore, the presence of these two domains does not prove an orthology relationship. Conversely, the fact that there are major differences in the domain organization of Teq and neurotrypsin does not necessarily disprove orthology, because changes in domain organization could have occurred following divergence from a common ancestor.

For complex multidomain proteins such as Teq and neurotrypsin, the question of orthology versus paralogy may only be settled by studying the constituent domains separately (*5*–*8*). In view of the high degree of functional divergence of the members of the serine protease family, for which substrate selectivity is an important aspect, the orthology of a pair of serine proteases is best supported when based on conserved features of their catalytic domain (*9*). A standard and simple method for determining 1:1 orthology is the reciprocal best-hit approach, based on the principle that in any two completely annotated genomes, orthologous pairs of genes/proteins should give best reciprocal hits. A BLAST (basic local alignment search tool) search with the sequence of the serine protease domain of Teq against human proteins of the National Center for Biotechnology Information (NCBI) database yielded TMPS3 as the best hit ($E = 2 \times 10^{-43}$). TMPS3 is a member of a large family of transmembrane serine proteases that contain a low-density lipoprotein receptor class A domain and an SRCR domain. Human neurotrypsin gave a match with $E = 9 \times 10^{-36}$, which argues against a 1:1 orthology relationship between the protease-domains of Teq and neurotrypsin; this would be the only sound basis for assuming functional equivalence (*10*).

We also analyzed the phylogenetic relationship of the protease domains of representative members of human and fly protease families with PAUP (phylogenetic analysis using parsimony) maximum parsimony and bootstrap analyses. These analyses provided no statistical support for a simple 1:1 orthology relationship between Teq and neurotrypsin.

Although both loss of neurotrypsin function in humans and inactivation of Teq in *Drosophila* result in neurobiological abnormalities related to higher brain functions, there is no reason to assume that the observed deficits are the result of the loss of orthologous, functionally equivalent molecules. As such, the observation that deficiency of two serine proteases is associated with compromised higher brain functions does not allow for conclusions about their functional similarity. Given the structural differences between neurotrypsin and Teq, detailed above, it seems unlikely that these two genes have equivalent functions in flies and humans.

[1]Department of Biochemistry, University of Zurich, CH-8057 Zürich, Switzerland. [2]Institute of Enzymology, Biological Research Center, H-1113 Budapest, Hungary.

*To whom correspondence should be addressed. E-mail: Peter.Sonderegger@bioc.uzh.ch

References and Notes

1. G. Didelot *et al.*, *Science* **313**, 851 (2006).
2. L. Patthy, M. Trexler, Z. Váli, L. Bányai, A. Váradi, *FEBS Lett.* **171**, 131 (1984).
3. Wellcome Trust Sanger Institute, www.sanger.ac.uk/cgi-bin/Pfam/swisspfamget.pl?name=Q9V5U2_DROME
4. Wellcome Trust Sanger Institute Pfam software, www.sanger.ac.uk/Software/Pfam
5. M. A. Huynen, P. Bork, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5849 (1998).
6. E. V. Koonin, L. Aravind, A. S. Kondrashov, *Cell* **101**, 573 (2000).
7. R. R. Copley, C. P. Ponting, J. Schultz, P. Bork, *Adv. Protein Chem.* **61**, 75 (2002).
8. C. P. Ponting, R. R. Russell, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 45 (2002).
9. L. Patthy, *Cell* **41**, 657 (1985).
10. Teq also emerges as best hit ($E = 4 \times 10^{-35}$) if the serine protease domain of human neurotrypsin is used as a query in a BLAST search of *Drosophila* proteins of the NCBI database, but because neurotrypsin is far from being the best match in the reciprocal test, this cannot be used as evidence for 1:1 orthology of Tequila and neurotrypsin.

5 December 2006; accepted 23 May 2007
10.1126/science.1138410