

Letter to the editor

Regarding the putative identity of a moth (*Thaumetopoea* spp.) allergen

While pursuing a literature search on allergies related to the processionary moths (*Thaumetopoea* spp.) we came across an earlier publication in *Allergy* (1). In this study, identification of one of the allergens was pursued, resulting in a 14 kDa protein, of which 18 amino acids (aa) were sequenced. In the following section, this 18 aa fragment was compared to other aa sequences using the BLAST algorithm. The retrieved hits all exhibited ‘high’ prediction values (‘*e*-values’) and it was in the paper concluded that the sequenced fragment thus represented a protein lacking homologues in the databases searched (‘BLAST-2 and SWISS-PROT’).

We would like to draw your attention to the fact that prediction scores in BLAST

searches are calculated comparing the target sequence with the entire entries in the databases. Hence, an 18 aa fragment compared to proteins of approximately 14 kDa or 120 aa, by default will produce high prediction scores. The prediction score evaluates the likelihood that the similarity between the compared sequences is due to random events. As the fragment to a large extent lacks comparable sequence positions, there is from the BLAST algorithm’s aspect a large risk that the similarity in the positions and composition of the 18 aa sequenced could be a random event, as the additional 100 aa *could* be very different – if they were there to be compared.

Performing a similar search in both NCBI and EMBL databases, the BLAST

algorithm produces a set of highly appealing hits, albeit with poor prediction scores (i.e. high values). In Table 1, we have aligned the Moneo et al. (1) 18 aa fragment to a random selection of retrieved sequences, demonstrating the apparent sequence homologies. The retrieved sequences are, with some minor exceptions, all representatives of chemosensory proteins.

S. Larsson¹, A. Backlund²
¹Jodrell Laboratory, Royal Botanic Gardens Kew, Richmond, UK; ²Division of Pharmacognosy, Department of Medicinal Chemistry, Uppsala University, Uppsala, Sweden

Table 1. Alignment of Moneo et al. (1) sequenced part of Thp1 allergen, compared to a random selection of five sequences retrieved from BLAST searches at NCBI databases

Thaumetopoea	-----getysdkydtidvnevld-----	allergen	
Bombyx mori	mk-----cltiaallfvaglsiaekytdkydnidvdeilenrkllvpyikcvldgrctpdgkelkahikdgmqtacaktcdkqkvsarkivkhikqheadyweqmakaydp--kdefkeiyegflagqn	chemosensoryprotein 6	0.011
Culex pipiens	mk-----llivfalvalvaagdstytnkydnidvdeilksdrllfknyfncldggpctpdatelkqslpdalennscskctpgkqevgykvvgwlinnrpeewnvlrakydenkfiekyrdeakaagin	chemosensoryprotein 1	0.048
Drosophila simulans	mksalalvfvcivglaaaapektytnkydsvnvdevltnnrvgnylkcldmgpcctpegrelkrllpdalhscskctevqrknsgkvinyllrankagewklllkkydpggiyrakkegh	(genomic clone)	0.21
Papilio xuthus	mkskaalvl---mcvlaaalaesysdkydnidlgeladndrllidayancillekgkcspegkelghmkdaletgcektcdakkgtnfmdhlirkkpeiwnqlankydpdgkwrkvyedrakehgviph	chemosensoryprotein	0.5
Spodoptera exigua	mk--vvflrvltavvyshphdshytdkydnidldelinnkiltayincldlgkctpdgkelkshirealenkcgktsaagkngtrkvmthlinfepdywnqlcakypdegkykamyekyktlvh	chemosensoryprotein CSP3	0.91
Heliothis virescens	m-----alarpdgaaytdkydnvldelienrrllvpyvkicldggkcapdakelkehiialeenecgkcteaqkgkgttrrvighlinneadywneltakfdpekkyyqgkyekelkevka	chemosensoryprotein 1	2.2

Similar results were obtained using BLAST facilities of EMBL.

Reference

1. Moneo I, Vega JM, Caballero ML, Vega J, Alday E. Isolation and characterization of Tha p 1, a major allergen from the pine processionary caterpillar *Thaumetopoea pityocampa*. *Allergy* 2003;**58**:34–37.