Agri Gene xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Agri Gene

journal homepage: www.elsevier.com/locate/aggene

Structural components of chemosensory protein mutations in the silkworm moth, *Bombyx mori*

Ning Xuan [a], Balaji Rajashekar [b], Sergo Kasvandik [c], Jean-François Picimbon [a,d,*]

[a] Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Jinan, Shandong Province, China
[b] Institute of Computer Science, University of Tartu, Tartu, Estonia
[c] Proteomics Core Facility, Institute of Technology, University of Tartu, Tartu, Estonia
[d] School of Bioengineering, Qilu University of Technology, Jinan, Shandong Province, China

ARTICLE INFO

Article history:
Received 27 June 2016
Received in revised form 26 August 2016
Accepted 29 August 2016
Available online xxxx

Keywords:
Insects
Lepidoptera
Cysteine
Glycine
α-helix
Protein evolution

ABSTRACT

The letter addresses the contribution of RNA editing to tissue-specific diversification of insect chemosensory proteins (CSPs) with a particular focus on the moth *Bombyx mori*. This narrow focus on RNA editing of a single gene family in a single insect species might have a broad significance to an integrated view of protein structure evolution. Including protein structure models of new edited variants and mapping the position of specific cysteine and glycine residues believed to be inserted though or mediated via mutation, we suggest effects of RNA editing on CSP protein function.

RNA editing, i.e. base change, substitution or replacement at the RNA level, is not new. The examples are the human apolipoprotein B (APOB) and neurotransmittor receptor genes, which both can be recoded thanks to RNA editing (Maas and Rich, 2000). Glutamate and serotonin receptor genes are examples of A-to-I substitution, while APOB is an example of C-to-U change. In the brain, specific A-to-I conversion is mediated by protein enzyme called adenosine deaminase acting on RNA or ADAR (Bass, 2002; Maas et al., 2002; Schaub and Keller, 2002; Jantsch and Ohman, 2008; Nishikura, 2010; Samuel, 2012). In the lipid system, the mechanism known to perform C-to-U conversion refers to Apolipoprotein B editing catalytic subunit 1 (Apobec-1) enzyme functional in specific tissues such as the liver and adipose tissue from vertebrates (Chester et al., 2000; Blanc et al., 2014). These two types of base change and/or enzymatic mechanisms have for long questioned the dogmatic concept of one single gene encodes for one single protein.

This dogma is severely being questioned in insects such as the fruitfly *Drosophila* where genome-wide analysis of RNA transcripts has shown that RNA editing acts not only on ion channels, but also on many versatile gene families (St Laurent et al., 2013). Similarly, some *Papilio* genes involved, for instance, in environmental mimicry are subjected to a strong allelic polymorphism, possibly associated to novel butterfly traits (Thompson et al., 2014). In line with these two independent studies, our discovery in the silkworm moth *Bombyx mori* in regard to editing of chemosensory proteins (CSPs) argues strongly further to the concept of mutation to increase protein diversity and bring new function within a gene family (Xuan et al., 2014; Picimbon, 2014a, 2014b).

1. Discovery of RNA/DNA differences in chemosensory protein (CSP) family

1.1. RNA editing and protein variance, problem case study for CSP genes

Chemosensory proteins (CSPs) refer to a rather ancient family of small acidic soluble proteins (10–12 kDa, 110–120 amino acids) originally described in cockroaches and flies, but could date back to bacterial systems (Nomura et al., 1992; McKenna et al., 1994; Pikielny et al., 1994; Picimbon and Leal, 1999; Liu et al., 2016a). CSP genes show high expression levels throughout the whole insect body, particularly

Abbreviations: A, adenine; ADAR, adenosine deaminase acting on RNA; APOB, apolipoprotein B; APOBEC, apolipoprotein B editing catalytic subunit; Bmor, *Bombyx mori*; C, cytosine; cDNA, complementary DNA; CSP, chemosensory protein; Cys, Cysteine; DNA, deoxyribonucleic acid; EST, expressed sequence tag; G, guanine; gDNA, genomic DNA; Gly, Glycine; I, inosine; NMR, Nuclear Magnetic Resonance; RDD, DNA/RNA difference; RNA, ribonucleic acid; T, thymine; U, uracyl.

* Corresponding author at: QLUT, University road, Changqing district, Jinan, Shandong Province, PR China.
E-mail address: jfpicimbon@gmail.com.

http://dx.doi.org/10.1016/j.aggene.2016.08.006
2352-2151/

in sensory organs such as antennae and legs (Angeli et al., 1999; Picimbon et al., 2000a, 2000b, 2001; Picimbon, 2003). *CSPs* also express in many different tissues in the system development process and the general innate immune system, suggesting multi-function for this protein family (Picimbon et al., 2000a, 2000b, 2001; Picimbon, 2003, 2014c; Sabatier et al., 2003; Wanner et al., 2005; Maleszka et al., 2007; Xuan et al., 2015; Liu et al., 2016b). CSPs are α-helical proteins composed of four Cysteine residues at highly conserved positions (Cys29, 36, 55 and 58). These α-helices fold as a prism suitable for the transport of long lipid chains, cuticular hydrocarbons and/or small toxic xenobiotic compounds (Lartigue et al., 2002; Mosbah et al., 2003; Ozaki et al., 2005; Tomaselli et al., 2006; Jansen et al., 2006, 2007; Xuan et al., 2015; Liu et al., 2016b).

The problem comes out that most insect genomes contain only a limited number of CSP genes (Wanner et al., 2004; Ozaki et al., 2008). Then, how the number of CSPs can cope to a multifunctional task possibly including the transport of a million odorants, pheromones, plant volatiles, fatty acids, lipids, xenobiotic toxic insecticides and all many different other environmental chemicals of all sorts?.

CSPs might achieve multi-function via conformational flexibility and cooperativity upon ligand binding (Campanacci et al., 2003). However, the finding that the CSP gene family is subjected to tissue-specific RNA editing process strongly suggests that they might also achieve it via the production of a high diversity of protein variant isoforms (Xuan et al., 2014; Picimbon, 2014a, 2014b). This slopes to a new concept perhaps applicable for all various families of small soluble ligand-binding protein transporters: each protein transporter or binding protein does not exist as a single copy, but as a high number of variants and/or isoforms strictly depending on tissue function.

1.2. RNA editing and protein rearrangement, potential for new functional CSPs

Comparing the nucleotide sequences of genomic DNA clones encoding CSPs using five tissues in five individuals, we found no errors, artifacts or polymorphisms. However, analyzing CSP-cDNA clones in these five tissues in these five individuals, we found an extremely high number of base mutations (RNA/DNA differences or RDDs) in the coding region without necessarily altering the size of the protein product (Xuan et al., 2014). EST sequencing in *B. mori* has confirmed the existence of such numerous CSP-RNA variants (XM_012693789-XM_012693964). Interestingly, our analysis of transcripts and peptide fragments in the silkworm moth *B. mori* shows that most of all these CSP mutations or RDDs are not built hapharzardly. They are rather pinpointed in specific regions of the molecule such as α-helix 4, α5 and the interspace between α2 and α3 (Xuan et al., 2014), suggesting that specific mutations target key functional structures of the protein. RDD levels are also shown to be tissue-specific. They are found for all expressed CSPs, and some of the CSP-RDD sequences identified in *Bombyx* lead to CSP motifs described in other insect species (Xuan et al., 2014).

Additionally, there is the very intriguing point of finding a high number of stop mutations corresponding to truncated peptide sequences and correlatively, a high number of shortened CSPs specifically expressed in the moth pheromone gland where we find most remarkably high level of mutations or RDDs (Xuan et al., 2014).

Changes in protein sequence do not include only amino acid replacements, but also inversions and deletions that are difficult to imagine to have occurred through RNA editing alone (Picimbon, 2016). The mechanisms behind RNA/protein editing in *B. mori* are presently unknown. From RDD data, we found dozens of base replacements leading to amino acid substitution. From digested peptide data, we found about a dozen of locations where the amino acid composition of fragmented BmorCSP1, BmorCSP2 and other twelve BmorCSPs from the pheromone gland differ from the original copy, i.e. the original sequence of the full CSP protein produced by the gene without any single mistakes and/or mistmatches (Xuan et al., 2014).

2. Modelling suggests a functional impact of specific base changes

The structure of BmorCSP1 is known, providing a 100% match with template for modelling studies (Jansen et al., 2006, 2007, 2jnt.1.A). Modeling the protein structure encoded by CSP1-RDDs or mutations cloned in various tissues of the silkworm *B. mori* suggests that they can produce multiple kinds of variant CSP protein structures (Fig. 1). On BmorCSP1 models, any mutation (GCEKC-to-ECKGC, ECGKC or GCGEKC) between Cys55 and Cys58 leads to loss of the first link of α4 at the bottom of the prism (Jansen et al., 2006, 2007, 2jnt.1.A; Fig. 1A). The clone CSP1PG5C5 displays N-to-D-19, K-to-N-33 and substitution of C-terminus. The resulting protein is a CSP1 with a prominent α5-α6 loop and an α6-helical motif with only one link (Fig. 1B). The clone CSP1PG1C3 makes a reading frame switch after A-86; the α6-helical motif is maintained, but the free C-terminal end is shortened (Fig. 1B). The clone CSP1PG1C5 has G-to-W-40 mutation and results in a CSP protein with a shorter α1 (Fig. 1B). Clones such as CSP1A4C4 and CSP1L5C10 retain mutant sequence after TAHFD-89, providing a protein with no loop between α4 and α5. This rather gives a V-shaped CSP with two tunnels between α1/α2-α4 and α5-α6 helices, respectively (Fig. 1B). The clone CSP1PG5C7 has an E-to-G residue mutation at position 24 and another mutation after TAHFD-89, potentially leading to two very elongated N- and C-terminal motifs (Fig. 1B).

Hence, using BmorCSP1 for which 3D NMR structure is known as a study model, we suggest that a large repertoire of variant CSP1 proteins can be produced thanks to mutations that lead to truncated shortened proteins, and mutations that produce proteins of the original size, but bearing some key pinpointed amino acid residue replacements (Xuan et al., 2014; Fig. 1A-B).

2.1. Mutations and reorganization of N- and C-terminal α-helices (+Cys)

It is worth noting that we find many types of mutations, particularly Cys addition, not evenly distributed over the whole protein structure, but heavily concentrated on the N- and C-terminal regions, while amino acid inversions are abundantly found at the level of α5-helix (Xuan et al., 2014; see Figs. 1 & 2). The α5 helix is part of the ligand binding cavity of CSP (Campanacci et al., 2003; Liu et al., 2016b), suggesting that specific nucleotide base and/or amino acid mutations happen on crucial locations of the original sequence to produce a novel particular structure and/or binding capacity in the CSP protein family (Figs. 1–3 & S1–S3).

The roles of the free N- and C-termini separated from the true functional binding site of the protein are more obscure. However, it is known that the N-terminal profiling of α-helices (α1 and α2) identified in two CSP orthologs does not overlap (Picimbon and Regnault-Roger, 2008), suggesting specific function for the N-terminal arm of CSPs. This is a fact from crystallography and NMR studies (Lartigue et al., 2002; Jansen et al., 2006, 2007). Hence, adding Cysteine in the N-terminus could well help produce new structural features and change the functional mechanism of CSP (Fig. 2). In our survey of CSP-RDDs in the silkworm moth *B. mori*, we find multiple A-to-G mutations that can switch Tyrosine to Cysteine at multiple various positions, possibly converting CSP from lipid transporter to protein effector (Xuan et al., 2014; Fig. 2). Interestingly, we observe load of new Cys not only in the N-terminus but also at the level of α5 and α6, suggesting the production of new disulfide bridges enrolling the two terminal tails of CSPs through or mediated via RNA editing (Fig. 2). The mutations in the C-terminus of CSP may simply keep the protein in the endoplasmic reticulum aside the secretory pathway. However, there are a lot of examples where mutations in C-terminus

Please cite this article as: Xuan, N., et al., Structural components of chemosensory protein mutations in the silkworm moth, *Bombyx mori*, Agri Gene (2016), http://dx.doi.org/10.1016/j.aggene.2016.08.006