



IgE recognition of chimeric isoforms of the honeybee (*Apis mellifera*) venom allergen Api m 10 evaluated by protein array technology



Matthias Van Vaerenbergh^a, Lina De Smet^a, David Rafei-Shamsabadi^b, Simon Blank^c,
Edzard Spillner^d, Didier G. Ebo^e, Bart Devreese^f, Thilo Jakob^{b,1}, Dirk C. de Graaf^{a,*,1}

^a Laboratory of Molecular Entomology and Bee Pathology, Ghent University, Krijgslaan 281 S2, B-9000 Ghent, Belgium

^b Allergy Research Group, Department of Dermatology, University Medical Center Freiburg, Hauptstrasse 7, D-79104 Freiburg, Germany

^c Center of Allergy and Environment (ZAUM), Technical University and Helmholtz Center Munich, Ingolstädter Landstraße 1, D-85764 Munich, Germany

^d Department of Engineering, Aarhus University, Gustav Wieds Vej 10, DK-8000 Aarhus C, Denmark

^e Department of Immunology, Allergology, and Rheumatology, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

^f Laboratory of Protein Biochemistry and Biomolecular Engineering, Ghent University, K.L. Ledeganckstraat 35, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:

Received 25 September 2014

Accepted 28 September 2014

Available online 17 October 2014

Keywords:

Honeybee

Venom

Chimeric transcripts

Allergy

Protein array

ABSTRACT

Api m 10 has recently been established as novel major allergen that is recognized by more than 60% of honeybee venom (HBV) allergic patients. Previous studies suggest Api m 10 protein heterogeneity which may have implications for diagnosis and immunotherapy of HBV allergy. In the present study, RT-PCR revealed the expression of at least nine additional Api m 10 transcript isoforms by the venom glands. Two distinct mechanisms are responsible for the generation of these isoforms: while the previously known variant 2 is produced by an alternative splicing event, novel identified isoforms are intragenic chimeric transcripts. To the best of our knowledge, this is the first report of the identification of chimeric transcripts generated by the honeybee. By a retrospective proteomic analysis we found evidence for the presence of several of these isoforms in the venom proteome. Additionally, we analyzed IgE reactivity to different isoforms by protein array technology using sera from HBV allergic patients, which revealed that IgE recognition of Api m 10 is both isoform- and patient-specific. While it was previously demonstrated that the majority of HBV allergic patients display IgE reactivity to variant 2, our study also shows that some patients lacking IgE antibodies for variant 2 display IgE reactivity to two of the novel identified Api m 10 variants, i.e. variants 3 and 4.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Honeybee venom (HBV) allergy is the second most frequent form of Hymenoptera venom allergy, observed especially in children, beekeepers and their relatives and other individuals that have a higher risk of being stung (Muller, 2005). HBV allergy is mediated by IgE antibodies specific to protein allergen components present within the venom. The protein composition of HBV is highly complex, with at least 113 identified proteins and peptides (Van Vaerenbergh et al., 2014). The complexity is even increased by different glycosylation patterns and protein heterogeneity [phospholipase A₂ (Peiren et al., 2005; Blank et al., 2011a); Api m 6 (Peiren et al., 2006; Kettner et al., 2001)]. Currently,

12 different HBV proteins have been recognized as allergens (<http://www.allergen.org/>). One of these allergens is icarapin, designated as Api m 10 according to the International Union of Immunological Societies nomenclature. Api m 10 has recently been reported as novel major allergen, that is recognized by up to 62% of HBV allergic patients (Blank et al., 2011b; Köhler et al., 2014). Interestingly, while present in unprocessed HBV, Api m 10 seemed to be absent or underrepresented in therapeutic HBV preparations used for venom-specific immunotherapy (Blank et al., 2011b). Until now two alternatively spliced Api m 10 transcripts have been identified [variant 1 (Peiren et al., 2006) and variant 2 (Peiren et al., 2006; Blank et al., 2011b)], both of which display IgE reactivity that is independent of cross-reactive carbohydrate determinants (CCDs) (Peiren et al., 2006; Blank et al., 2011b).

Upon previous 2D-gel separation of pure HBV we identified four protein spots as Api m 10 (Peiren et al., 2005). Interestingly, the difference in theoretical molecular weight (MW) between the two known splice variants was known to be small (0.3 kDa) and thus

* Corresponding author. Tel.: +32 92648732; fax: +32 9 2645242.

E-mail address: Dirk.deGraaf@UGent.be (D.C. de Graaf).

¹ Shared senior authorship.

could not account for the observed MW differences between these protein spots. Two Api m 10 spots from the higher MW region of the gel differed by about 3 kDa, which may correspond to different glycoforms of variant 1 and/or 2. Curiously, the MW of the other two Api m 10 spots was about 30 kDa lower, which strongly suggested that additional Api m 10 isoforms existed. For other allergens, protein heterogeneity has been reported to be immunologically relevant. For instance, the birch genome contains at least 7 pollen-expressed genes that encode distinct isoforms of the major birch pollen allergen Bet v 1 that display varying IgE reactivity (Schenk et al., 2006; Wagner et al., 2008). Therefore, the present study focused on the identification of additional Api m 10 isoforms and the impact of Api m 10 protein heterogeneity on IgE reactivity.

2. Materials and methods

2.1. Venom gland RNA isolation and cDNA synthesis

Approximately 100 honeybee (*Apis mellifera carnica*) venom glands were dissected under anesthesia by chilling and were transferred to RNeasy lysis buffer (Qiagen, Crawley, UK). RNA was extracted using the RNeasy Mini Kit (Qiagen, Crawley, UK) following the protocol for purification of total RNA from animal cells using spin technology. Subsequently, cDNA was synthesized using the RevertAid™ H Minus First Strand cDNA synthesis kit (Fisher Scientific, Warrington, UK).

2.2. Screening of Api m 10 transcript heterogeneity

RT-PCRs were carried out as described before (de Graaf et al., 2010). Primers for amplification of the mature (without secretory signal sequence) Api m 10 sequence were developed, allowing ligation-independent cloning by a 5' incorporated sequence (in italics): forward primer 5'-GACGACGACAAGATGTTCCCTGGTGACACGATG-3' and reverse primer 5'-GAGGAGAAGCCCGTCAAGCAGTTAATACATCTCCT-3'. RT-PCR-amplification products were electrophoretically separated on a 2% agarose gel over a 25 cm distance. The complete length of the gel lane was divided in ten gel pieces and separate DNA extractions were performed by the Genejet gel extraction kit (Fisher Scientific, Warrington, UK). Subsequently, RT-PCR fragments were cloned in the pEx-7 Ek/LIC vector according to the instructions of the Ek/LIC cloning kit (Novagen, Madison, WI, USA). For each cloning reaction, plasmid DNA was extracted (Birnboim and Doly, 1979) from eight different colonies and inserts with different lengths were identified by agarose gel electrophoresis of restriction digests. Plasmid DNA was purified and DNA sequencing of both strands was performed as described previously (de Graaf et al., 2010) using vector sequencing primers. Sequences were analyzed by multiple sequence alignment (Clustal W2; Larkin et al., 2007). Putative N- or O-linked glycosylation sites were determined with NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and NetOGlyc 3.1 (<http://www.cbs.dtu.dk/services/NetOGlyc/>).

2.3. Proteomics

MS/MS data of HBV generated in prior studies (Peiren et al., 2005; Van Vaerenbergh et al., 2013, 2014) were reanalyzed by searching against the honeybee protein RefSeq database extended with all translated Api m 10 isoform sequences to identify isoform-specific tryptic peptides. Identical search parameters were used as those previously described (Peiren et al., 2005; Van Vaerenbergh et al., 2013, 2014).

2.4. Synthetic peptides and recombinant production

Isoforms smaller than 50 amino acids were synthetically produced (Genscript, Piscataway, NJ, USA) at HPLC purities higher than 70%. Mass spectrometry peak analysis was performed to analyze peptide quality.

Synthetic production of some of these isoforms was impossible due to technical limitations. These isoforms and isoforms larger than 50 amino acids were produced by bacterial expression. Api m 10 variant 2 was expressed and purified using the IMPACT protocol (New England Biolabs, Beverly, MA, USA) as previously described (Blank et al., 2011b). All other Api m 10 variants and uterocalin were cloned, sequenced and expressed using the pET system (Life technologies, Carlsbad, CA, USA) (de Graaf et al., 2010). They were overproduced in *Escherichia coli* as His-tagged fusion proteins and purified by immobilized metal chelate affinity chromatography on Ni-NTA beads (adapted protocol from (Van Vaerenbergh et al., 2013)). When native purification yielded insufficient amounts of soluble protein, denaturing purifications were applied. Purified proteins were dialyzed against phosphate buffered saline (PBS) and sample purity was evaluated by Coomassie staining of SDS-PAGE gels. Protein integrity was verified by Western blots using an anti-His antibody (Van Vaerenbergh et al., 2013).

2.5. Patients' sera

The diagnosis of HBV allergy was based on the history of systemic allergic reaction to a sting of a honeybee (as identified by the patient), positive skin test and sIgE to HBV (≥ 0.35 kU_A/L, FEIA ImmunoCAP i1, Thermo Fisher Scientific, Uppsala, Sweden), as recently described (Hofmann et al., 2011a). Api m 10-specific IgE titers were determined by ImmunoCAP FEIA tests (Thermo Fisher Scientific, Uppsala, Sweden) with rApi m 10 variant 2, recombinantly produced in Sf9 insect cells (Blank et al., 2011b). In addition, sIgE to Api m 1 (i208), and MUXF3 (CCD) as well as total serum IgE were determined (Table S1). Sera obtained from wasp-stung individuals without allergic sting reaction and/or lack of HBV sensitization were used as negative controls. All patients had given their informed written consent to draw a serum sample, and all experiments applying human sera were approved by the local ethics committee.

2.6. Protein array spotting and development

Synthetic peptides were dissolved at 0.1 mM in 0.01 M phosphate buffer pH 7.4 with 50 mM NaCl and recombinants were dissolved at 0.5 mg/ml in 0.01 M phosphate buffer pH 7.4 with the exception of variant 1 which had a final concentration of 0.2 mg/ml as it was insoluble at higher concentrations. Peptides and proteins were printed on nitrocellulose-coated glass slides (Sartorius, Göttingen, Germany) in duplicate, a procedure conducted by Innobiochips (IBL, Lüneburg, Germany). Besides the Api m 10 isoforms, an irrelevant control protein (uterocalin or BSA) was spotted. Additionally, two proteins to assess the quality of biochip development and spot normalization were spotted: protein G binds goat-IgG and Goat Anti-Mouse binds all antibodies from mouse.

A protocol for human serum IgE recognition was developed: arrays were saturated for 1 h at room temperature (RT) with 150 μ l of saturation buffer (PBS-BSA 1% – Tween20 0.05%). Incubation with 50 μ l of two-fold diluted (with saturation buffer) serum samples was performed overnight at 4 °C. Subsequently, arrays were incubated for 1 h at RT with 100 μ l of a secondary polyclonal horseradish peroxidase-linked goat-IgG anti-human IgE (2 μ g/ml in saturation buffer) (Acris Antibodies, Hiddenhausen, Germany). Finally, 50 μ l of 3,3',5,5'-tetramethylbenzidine solution (Calbiochem, La Jolla, CA, USA) was added which generates a stable precipitate at the reaction

site. Arrays were washed three times for 5 min with PBS-Tween20 0.05% between all consecutive steps. After development, arrays were air-dried and scanned using the Spotware colorimetric array scanner (obtained from European Biotech Network, Dolembreux, Belgium) at a scan gain of 0.9 and a scan resolution of 5 μm (Spotware 1.1 software). ImageJ was used to adjust image size to a 1000 pixel width. The resulting image is quantified using Mapix (Innopsis, Carbonne, France).

2.7. Data analysis

For each spot, the local median background intensity is subtracted from the median spot intensity. Only spots with signal to noise ratios greater than a minimum threshold of two were included in the analysis. Normalization between arrays was done using protein G and the average of spot duplicates was calculated. The average spot intensity per protein on control arrays developed with sera of honeybee non-allergic patients was subtracted from the average spot intensity per protein on the arrays developed with sera from honeybee allergic patients. Next, for each array a cut-off value was calculated as the average of the intensity of the irrelevant control protein plus three times the standard deviation (SD) (Table S2). SDs were calculated as follows: $((\text{SD of the control protein})^2 + (\text{SD of the control protein spots on the negative control arrays})^2)^{1/2}$.

3. Results and discussion

3.1. Api m 10 transcript heterogeneity

Mature Api m 10 transcripts were amplified from venom gland tissue by RT-PCR (Fig. 1). Sequence analysis of seventeen cloned RT-PCR fragments, identified eleven different variants, which were named variant 1 to variant 11. All cloned nucleotide sequences (and their GenBank numbers) and translated protein sequences are shown in Doc. S1.

The Api m 10 gene is positioned on chromosome LG1 (GeneID: 503505) and consists of 4 exons. Sequence alignment of all variants to the genome indicates the existence of a single genomic locus only (Doc. S2). Two distinct mechanisms seem to be responsible for the generation of these variants. First, the splice sites of variant 1 conform to the general canonical GT-AG splicing rule (Xiong et al.,

2009). Variant 1 and variant 2 differ by only 12 base pairs due to the presence of an alternative splice acceptor site in exon 3 and also this splicing is consistent with the GT-AG rule. In contrast, the nine newly detected variants are so-called chimeric transcripts, which are distinct from conventionally spliced mRNA isoforms as they are produced by joining exons from two or more different gene transcripts. Instead of splice sites, short homologous sequences (SHSs) are found at the junction sites of the source sequence. The Api m 10 gene contains multiple distinct SHSs in exon 2, which are joined to SHSs from exon 3 or exon 4 (Fig. 2 and Doc. S2). The SHSs are 5–9 base pairs in length. In four variants, this process generates frame shifts and premature stop codons, leading to C-terminally truncated variants. While variant 1, the longest mature variant, consists of 204 amino acids, variants 8 to 11 are severely truncated with lengths of 41, 25, 19 and 12 amino acids. Protein sequence alignments of all variants are shown in Fig. 3.

To the best of our knowledge, we here report the first identification of chimeric transcripts generated by the honeybee. Although thousands of chimeric transcripts have been reported in fruit fly, mouse and human (Li et al., 2009), evidence at the protein level is still scarce. In humans, the few characterized chimeric proteins are associated with cancer, although some also appear to be expressed at low levels under normal physiological conditions (Frenkel-Morgenstern et al., 2012; Casado-Vela et al., 2013). Their physiological role remains elusive but they certainly boost the complexity of the proteome (Casado-Vela et al., 2013). As until now neither function nor enzymatic activity of Api m 10 could be determined, the impact of this heterogeneity on its function/activity is impossible to establish. However, as the sequence length of some isoforms is severely reduced, we hypothesize that some of them are loss-of-function mutants.

3.2. Proteomics

The specific and unambiguous identification of the Api m 10 protein isoforms in HBV requires the detection of the isoform-specific junction peptides by bottom-up proteomics (Casado-Vela et al., 2013). Searching HBV MS/MS data from preceding research (Peiren et al., 2005; Van Vaerenbergh et al., 2013, 2014) against a database containing all translated Api m 10 isoform sequences identified multiple tryptic peptides (Fig. 3), but no isoform-specific peptides were found. The differentiation between variants can only be based on a limited number of peptides. For instance, there is only a single difference between the theoretical set of tryptic peptides of variants 1 and 2. Moreover this peptide is larger than 50 amino acids, which typically gives poor signal intensity in MALDI peptide mass fingerprinting.

In a preceding study (Peiren et al., 2005), venom proteins were separated using 2D-gel electrophoresis and Api m 10 tryptic peptides were identified in four spots. Two tryptic peptides (KNVDTVLVLP-SIER and VREQMAGILSR) were detected in two high molecular weight spots (spots 8 and 9), which we suggest to be different glycosylation forms of variant 1 and/or variant 2. In contrast, only one Api m 10 tryptic peptide (KNVDTVLVLP-SIER) was found in two low molecular weight spots (± 8 kDa) (spots 19 and 20) which, based on their molecular weight, cannot represent full-size variant 1 and/or variant 2. These proteins may represent one of the newly identified Api m 10 chimeras, although protein degradation of the larger variants cannot be excluded. The VREQMAGILSR peptide, a variant 1- and variant 2-specific peptide (Fig. 3), was not detected in these spots. The only identified tryptic peptide (KNVDTVLVLP-SIER) is part of variant 1 to 5 and based on the theoretical pI/MW this spot may be assigned to variant 5.

Proteomic evidence for the smallest isoforms (isoform 9–11) can hardly be obtained using our approach, as detectable tryptic peptides are not generated. Also, due to the poor resolution in

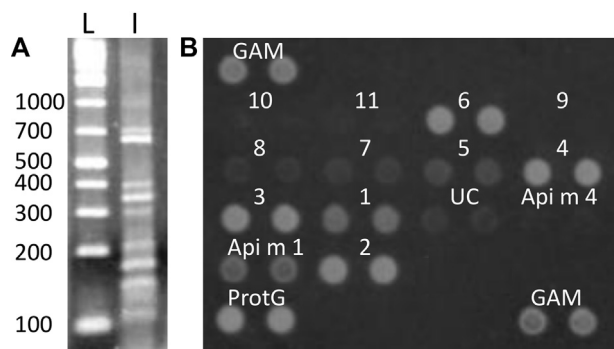


Fig. 1. IgE recognition of identified Api m 10 isoforms. (A) RT-PCR on honeybee venom gland tissue reveals multiple Api m 10 isoform transcripts. Base pair lengths of ladder proteins are shown. L = GeneRuler 1 kb DNA ladder (Fermentas); I = Api m 10 amplicons. (B) IgE recognition of Api m 10 isoforms evaluated by protein array technology. Numbers represent spotted Api m 10 variants 1 to 11, printed as synthetic constructs or recombinants. Also a negative control protein (UC) and two positive control proteins [purified native honeybee venom allergens Api m 1 and Api m 4 (Sigma–Aldrich, St. Louis, MO)] were included. Additionally, protein G (ProtG) and Goat Anti-Mouse (GAM) allowed to assess the quality of biochip development and spot normalization. The array shown was developed with serum of patient 1 (see Table 1 for result).

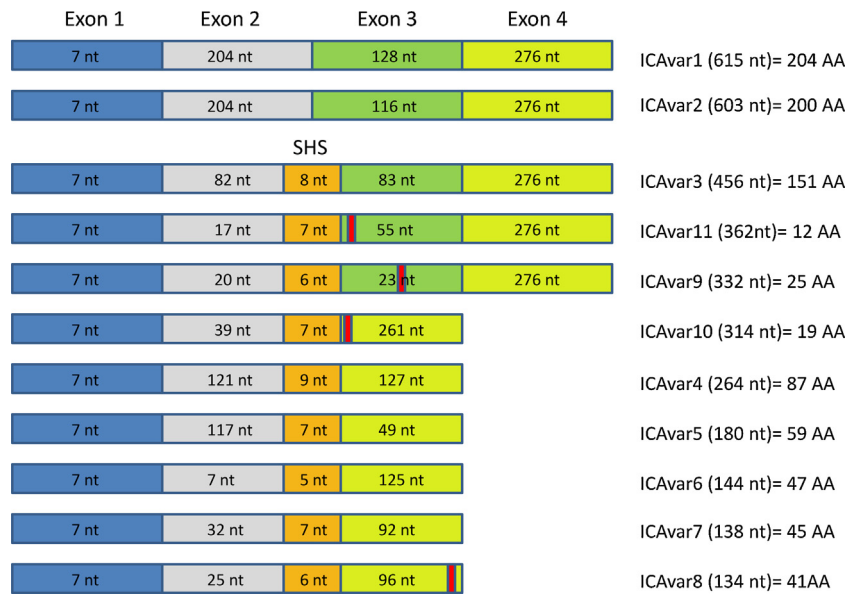


Fig. 2. Exon structure of Api m 10 variants. Schematic figure showing the exon structure of sequenced Api m 10 amplicons, named Api m 10 variant 1 to 11 (ICAv1 to ICAv11). Different exons are shown as colored boxes. Distinct short homologous sequences (SHSs) are found at the junction sites of the chimeric transcripts (orange boxes). Red boxes present alternative stop codons, which generate C-terminally truncated variants. The full coding sequence lengths are shown: number of nucleotides (nt), number of amino acids (AA).

classical SDS-PAGE separation they may not be visualized on a 2D-gel. Besides, due to the process of nonsense-mediated mRNA decay, these smallest isoforms may not be expressed (Conti and Izaurralde, 2005).

3.3. Testing IgE reactivity to Api m 10 isoform

We additionally explored the effect of the observed Api m 10 protein heterogeneity on IgE reactivity by protein array technology. Sera of HBV allergic patients were pre-screened for the presence of specific IgE using Api m 10 ImmunoCAP research prototypes as recently described (Köhler et al., 2014). Api m 10 used for generating the ImmunoCAPs corresponded to variant 2 which has been recombinantly produced using the Sf9 insect cell line. As this cell line produces glycans that lack or show significantly reduced α -1,3-core fucosylation (Blank et al., 2011b; Seismann et al., 2010), the expressed Api m 10 is devoid of CCDs and allows to assess if sera possess IgEs binding to Api m 10 variant 2 independent of CCD reactivity. For spotting on the array, all 11 variants were produced by either peptide synthesis or bacterial expression (Doc. S3) and therefore lack glycosylations including CCD epitopes.

As unambiguous protein evidence for a number of Api m 10 isoforms was not obtained, the complete panel of isoforms was evaluated. Variants were printed on nitrocellulose-coated glass slides (Fig. 1). In contrast to other immunoassays, the applied protein array technology is a multiplex assay, which allows testing a broad protein panel with minute amounts of serum (Lebrun et al., 2005) and can reliably detect allergen-specific IgE below 0.35 kU_A/L, the current World Health Organization standard cut-off for sensitization (Lebrun et al., 2005).

In the first protein array experiment, IgE reactivity to Api m 10 isoforms was analyzed using sera of 18 HBV allergic patients sensitized to Api m 10 variant 2 as determined by ImmunoCAP (Table S1). ImmunoCAP IgE reactivity of variant 2 was confirmed using the protein array technology in sixteen out of eighteen patients (Table 1). Discrepant results in two patients may be related to the use of differently expressed Api m 10 variant 2 in the two assay systems. While the ImmunoCAPs were loaded with Api m 10 expressed in Sf9 cells, i.e. with glycosylation but devoid

of α -1,3-core fucosylation, the protein array used *E. coli* expressed i.e. non-glycosylated Api m 10 variant 2. This could indicate that absence of multiple carbohydrates (variant 2 contains 2 N-linked glycosylation sites) may have altered the protein conformation and specific conformational epitopes. A direct comparison of non-glycosylated and glycosylated (without CCD) variant 2 using an ELISA based assay system however showed comparable IgE reactivity to both forms of Api m 10 (Blank et al., 2011b). Alternatively, it may simply be an issue of different assay sensitivity of the two systems used for analysis. Both discrepant sera displayed rather low Api m 10-specific IgE titers (0.59 and 0.60 kU_A/L) by ImmunoCAP, which may have been too low to be detected by the protein array.

IgE recognition was found to be both isoform- and patient-specific (Table 1). In general, three groups of isoforms can be distinguished. Variants 1 to 4 are IgE recognized by most of the sera, variants 5 to 8 are recognized by less than half of the sera, and variants 9 to 11 are recognized by none of the sera. The observed differences in IgE reactivity between isoforms can be explained in two ways.

First, in case only the alternative splice variants 1 and 2 are present in the venom, the IgE recognition of several other variants may be explained by their conserved N- and C-terminal regions (Fig. 3). These regions may contain linear or even conformational IgE epitopes identical to those of variant 2. Consequently, in addition to variant 2, other variants will be recognized by IgE from sera that have been selected based on IgE reactivity to variant 2 in the ImmunoCAP. One of these conserved epitopes may correspond to the mouse IgE epitope of the venom Api m 10 of the Asian honeybee (*A. cerana*) (Wong et al., 2012), which is highly similar to *A. mellifera* Api m 10 variant 1. The experimentally determined antigenic serine is conserved in the variants 1 to 4 and 6 and 7 (Fig. 3), which explains their IgE reactivity in the array experiment. The observed IgE recognition of variant 5, which lacks this antigenic site serine, indicates that additional epitopes may exist, which are conserved between variant 2 and 5. The smallest variants, variant 9, 10 and 11 have only limited sequence identity with variant 2 which may explain why they are not recognized.

Second, besides variants 1 and 2, additional variants may be present in the venom. For example, the N-terminal region of the



Fig. 3. Protein sequence alignments of Api m 10 isoforms generated without (A) and with (B) frame shifts. (A) Yellow boxes show all identified tryptic peptides by searching honeybee venom MS/MS data from preceding research (Peiren et al., 2005; Van Vaerenbergh et al., 2013, 2014) against the RefSeq honeybee protein database extended with all translated Api m 10 isoform sequences. The antigenic site serine of the mouse IgE epitope from the venom Api m 10 of the Asian honeybee (*Apis cerana*) (Wong et al., 2012) is indicated by an arrow. (B) Isoform-specific sequences are indicated in gray.

Table 1

Result of the first protein array experiment, conducted with 18 variant 2-sensitized sera (ImmunoCAP Api m 10 IgE titers > 0.35 kU_A/L). IgE recognition of the spotted recombinant and synthetic Api m 10 variants is indicated for each serum sample: x = recognition; 0 = no recognition.

Patients No.	ImmunoCAP		Results protein array experiment										
	Api m 10		Recombinants							Synthetic peptides			
	kU _A /L	Class	VAR1	VAR2	VAR3	VAR4	VAR5	VAR7	VAR6	VAR8	VAR9	VAR10	VAR11
1	14.90	3	x	x	x	x	x	0	x	0	0	0	0
2	11.20	3	x	x	x	x	0	x	x	0	0	0	0
3	10.60	3	x	x	x	x	x	0	x	0	0	0	0
4	8.88	3	x	x	x	x	0	0	x	0	0	0	0
5	8.31	3	x	x	x	x	x	x	x	0	0	0	0
6	7.17	3	x	x	x	x	x	0	x	x	0	0	0
7	4.41	3	x	x	x	x	x	x	x	0	0	0	0
8	3.39	2	0	x	x	x	x	0	0	0	0	0	0
9	3.05	2	x	x	x	x	0	0	x	0	0	0	0
10	3.03	2	x	x	x	x	x	0	x	x	0	0	0
11	2.84	2	0	x	x	x	0	x	x	0	0	0	0
12	1.79	2	0	x	x	x	x	x	0	0	0	0	0
13	1.40	2	0	x	x	x	0	0	0	0	0	0	0
14	0.94	2	x	x	x	x	0	0	x	x	0	0	0
15	0.84	2	x	x	x	0	0	0	0	0	0	0	0
16	0.60	1	x	0	0	0	0	0	0	0	0	0	0
17	0.59	1	x	0	0	0	0	x	0	0	0	0	0
18	0.58	1	x	x	x	0	0	0	0	x	0	0	0
No. of positive sera			14	16	16	14	8	6	11	4	0	0	0

Table 2

Partial results of the second protein array experiment, performed with variant 2-non-sensitized sera. Only the results of arrays analyzed with sera containing ImmunoCAP Api m 10 IgE titers between 0.1 and 0.35 kU_A/L are shown. For each serum sample, IgE recognition of the spotted recombinant and synthetic Api m 10 variants is indicated: x recognition; 0 = no recognition. Besides, none of the variants were recognized by 14 sera with ImmunoCAP Api m 10 IgE titers below 0.1 kU_A/L (results not shown).

Patients No.	ImmunoCAP		Results protein array experiment										
	Api m10		Recombinants						Synthetic peptides				
	kU _A /L	Class	VAR1	VAR2	VAR3	VAR4	VAR5	VAR7	VAR6	VAR8	VAR9	VAR10	VAR11
1	0.34	0	0	0	x	x	0	0	0	0	0	0	0
2	0.29	0	0	0	0	0	0	0	0	0	0	0	0
3	0.27	0	0	0	0	0	0	0	0	0	0	0	0
4	0.27	0	0	0	0	x	0	0	0	0	0	0	0
5	0.26	0	0	x	0	0	0	0	0	0	0	0	0
6	0.26	0	0	0	0	0	0	0	0	0	0	0	0
7	0.24	0	0	0	0	0	0	0	0	0	0	0	0
8	0.23	0	0	0	0	0	0	0	0	0	0	0	0
9	0.23	0	0	0	0	0	0	0	0	0	0	0	0
10	0.22	0	0	0	0	x	0	0	0	0	0	0	0
11	0.20	0	0	0	0	0	0	0	0	0	0	0	0
12	0.18	0	0	0	0	0	0	0	0	0	0	0	0
13	0.18	0	0	0	0	x	0	0	0	0	0	0	0
14	0.17	0	0	0	0	0	0	0	0	0	0	0	0
15	0.12	0	0	0	0	0	0	0	0	0	0	0	0
No. of positive sera			0	1	1	4	0	0	0	0	0	0	0

variant 8 (Table 1) recognized by IgE and the non-recognized variant 10 are identical (Fig. 3). Therefore, this experiment suggests that variant 8 contains a unique IgE epitope within the variant-specific region. The lack of IgE recognition of the small variants 9 to 11 by all sera indicates that they have either a low antigenicity or that they are not expressed. Variants 1 to 7 may have common IgE epitopes as their N- and C-terminal regions are conserved (Fig. 3). In addition, as each variant lacks a different internal region, these variants have a different conformation and possibly variant-specific conformational IgE epitopes. To investigate if these variants possess variant-specific IgE epitopes and confirm that they are present in the venom, we conducted a second protein array experiment, analyzing the IgE reactivity of these variants with sera of HBV allergic patients lacking variant 2-specific IgE.

In the first protein array experiment, analyses were performed using sera containing variant 2-specific IgE levels higher than 0.35 kU_A/L, i.e. confirmed sensitization to variant 2. However, the ImmunoCAP system allows linear detection down to 0.1 kU_A/L, although the cut-off for sensitization is maintained at 0.35 kU_A/L. In the second protein array experiment, IgE reactivity of the Api m 10 variants was analyzed using 15 sera with variant 2-specific IgE titers between 0.10 and 0.35 kU_A/L, and 14 sera with variant 2-specific IgE titers lower than 0.10 kU_A/L (Table S1). This experiment showed that variant 3 and 4 are IgE recognized by 1 and 4 of the variant 2-negative sera, respectively. These sera belong to the group with variant 2-specific IgE titers between 0.10 and 0.35 kU_A/L (Table 2), while the sera with IgE titers lower than 0.10 kU_A/L did not recognize any variants. This observation indicates that a subset of patients displayed variant 3- and 4-specific IgE reactivity directed toward unique epitopes not present in variant 2 or other variants. In addition, we interpret this as indirect evidence that variant 3 and 4 are present in the venom.

3.4. Clinical consequences

More than 100 compounds have previously been found within HBV (Van Vaerenbergh et al., 2014). The successful identification of nine new Api m 10 chimeric transcripts and the indirect evidence for their presence in the HBV proteome, shows that the venom complexity has not yet been fully unraveled. Besides its consequences for the protein function and activity, protein heterogeneity may have important consequences for diagnosis and immunotherapy of HBV allergy. Allergen heterogeneity has already been observed

for the HBV allergen Api m 6, for which four allelic protein isoforms have been described differing in their primary structure at the amino and carboxy terminus by a maximum of six amino acids. However, immunoblot analyzes revealed no isoform-specific IgE (Kettner et al., 2001). In contrast, our study shows that Api m 10 variants are recognized differently by patient IgE, an observation which has been described for several other major allergens derived from different allergen sources (Wurtzen, 2001). The biological mechanism at the level of effector cell activation triggered by interactions between individual IgEs and different isoallergens however remains mostly unresolved (Christensen et al., 2010).

Currently, rApi m 1 is the only HBV allergen commercially available for component-resolved diagnosis (CRD) (Sturm et al., 2012). However, sensitization to rApi m 1 in HBV allergy is less frequent than previously anticipated. Depending on the patient population, between 20% and 42% of the HBV allergic patients lacks IgE reactivity to rApi m 1 (Hofmann et al., 2011b; Korosec et al., 2011, 2012; Sturm et al., 2011; Jakob et al., 2012). Therefore, additional HBV allergens appear to be relevant in HBV allergy. Indeed, using a panel of 6 different HBV allergens, we recently demonstrated that more than 94% of HBV allergic patients can be detected (Köhler et al., 2014). In this context, Api m 10 has been recognized as novel major allergen with IgE reactivity in up to 62% of HBV allergic patients (Blank et al., 2011b; Köhler et al., 2014). Our present study shows that the observed Api m 10 protein heterogeneity may have important consequences for diagnostic tests as IgE recognition is both isoform- and patient-specific. Variant 2 was previously demonstrated to be a good biomarker for IgE recognition of Api m 10 (Blank et al., 2011b; Köhler et al., 2014), which was confirmed by the present study. In addition, our study found that variant 3 and especially variant 4 may be of additional relevance for the diagnosis of HBV allergy in those patients that do not react to variant 2.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

Matthias Van Vaerenbergh is funded by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). Didier Ebo is a Senior Clinical Researcher of

the Research Foundation Flanders (FWO: 1800614N). The authors want to thank Andy Vierstraete (Department of Biology, Ghent University, Belgium) for sequencing the samples and Griet Debyser (L-PROBE, Ghent University, Belgium), Dr. Gonzalez Van Driessche (L-PROBE, Ghent University, Belgium) and Kim De Crem for technical assistance.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.molimm.2014.09.018>.

References

- Birnboim, H.C., Doly, J., 1979. Rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7, 1513–1523.
- Blank, S., Michel, Y., Seismann, H., Plum, M., Greunke, K., Grunwald, T., Bredehorst, R., Ollert, M., Braren, I., Spillner, E., 2011a. Evaluation of different glycoforms of honeybee venom major allergen phospholipase A2 (Api m 1) produced in insect cells. *Protein Pept. Lett.* 18, 415–422.
- Blank, S., Seismann, H., Michel, Y., McIntyre, M., Cifuentes, L., Braren, I., Grunwald, T., Darsow, U., Ring, J., Bredehorst, R., Ollert, M., Spillner, E., 2011b. Api m 10, a genuine *A. mellifera* venom allergen, is clinically relevant but underrepresented in therapeutic extracts. *Allergy* 66, 1322–1329.
- Casado-Vela, J., Lacal, J.C., Elortza, F., 2013. Protein chimerism: novel source of protein diversity in humans adds complexity to bottom-up proteomics. *Proteomics* 13, 5–11.
- Christensen, L.H., Riise, E., Bang, L., Zhang, C.Q., Lund, K., 2010. Isoallergen variations contribute to the overall complexity of effector cell degranulation: effect mediated through differentiated IgE affinity. *J. Immunol.* 184, 4966–4972.
- Conti, E., Izaurralde, E., 2005. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* 17, 316–325.
- de Graaf, D.C., Brunain, M., Scharlaken, B., Peiren, N., Devreese, B., Ebo, D.G., Stevens, W.J., Desjardins, C.A., Werren, J.H., Jacobs, F.J., 2010. Two novel proteins expressed by the venom glands of *Apis mellifera* and *Nasonia vitripennis* share an ancient C1q-like domain. *Insect Mol. Biol.* 19 (Suppl. 1), 1–10.
- Frenkel-Morgenstern, M., Lacroix, V., Ezkurdia, I., Levin, Y., Gabashvili, A., Prilusky, J., del Pozo, A., Tress, M., Johnson, R., Guigo, R., Valencia, A., 2012. Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. *Genome Res.* 22, 1231–1242.
- Hofmann, S.C., Pfender, N., Weckesser, S., Huss-Marp, J., Jakob, T., 2011a. Added value of IgE detection to rApi m 1 and rVes v 5 in patients with Hymenoptera venom allergy. *J. Allergy Clin. Immunol.* 127, 265–267.
- Hofmann, S.C., Pfender, N., Weckesser, S., Blank, S., Huss-Marp, J., Spillner, E., Jakob, T., 2011b. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy. *J. Allergy Clin. Immunol.* 128, 248.
- Jakob, T., Köhler, J., Blank, S., Huss-Marp, J., Spillner, E., Lidholm, J., 2012. Comparable IgE-reactivity to nApi m 1 and rApi m 1 in CCD negative bee venom allergic patients. *J. Allergy Clin. Immunol.* 130, 276–278.
- Kettner, A., Hughes, G.J., Frutiger, S., Astori, M., Roggero, M., Spertini, F., Corradin, G., 2001. Api m 6: a new bee venom allergen. *J. Allergy Clin. Immunol.* 107, 914–920.
- Köhler, J., Blank, S., Müller, S., Bantleon, F., Frick, M., Huss-Marp, J., Lidholm, J., Spillner, E., Jakob, T., 2014. Component resolution reveals additional major allergens in bee venom allergic patients. *J. Allergy Clin. Immunol.* 133 (5), 1383–1389.
- Korosec, P., Valenta, R., Mittermann, I., Celesnik, N., Erzen, R., Zidarn, M., Kosnik, M., 2011. Low sensitivity of commercially available rApi m 1 for diagnosis of honeybee venom allergy. *J. Allergy Clin. Immunol.* 128 (3), 671–673.
- Korosec, P., Valenta, R., Mittermann, I., Celesnik, N., Silar, M., Zidarn, M., Kosnik, M., 2012. High sensitivity of CAP-FEIA rVes v 5 and rVes v 1 for diagnosis of *Vespa* venom allergy. *J. Allergy Clin. Immunol.* 129, 1406–1408.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lebrun, S.J., Petchpud, W.N., Hui, A., McLaughlin, C.S., 2005. Development of a sensitive, colorimetric microarray assay for allergen-responsive human IgE. *J. Immunol. Methods* 300, 24–31.
- Li, X., Zhao, L., Jiang, H.F., Wang, W., 2009. Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J. Mol. Evol.* 68, 56–65.
- Muller, U.R., 2005. Bee venom allergy in beekeepers and their family members. *Curr. Opin. Allergy Clin. Immunol.* 5, 343–347.
- Peiren, N., Vanrobaeys, F., de Graaf, D.C., Devreese, B., Van Beeumen, J., Jacobs, F.J., 2005. The protein composition of honeybee venom reconsidered by a proteomic approach. *Biochim. Biophys. Acta* 1752, 1–5.
- Peiren, N., de Graaf, D.C., Brunain, M., Bridts, C.H., Ebo, D.G., Stevens, W.J., Jacobs, F.J., 2006. Molecular cloning and expression of icarapin, a novel IgE-binding bee venom protein. *FEBS Lett.* 580, 4895–4899.
- Schenk, M.F., Gilissen, L.J.W.J., Esselink, G.D., Smulders, M.J.M., 2006. Seven different genes encode a diverse mixture of isoforms of Bet v 1, the major birch pollen allergen. *BMC Genomics* 7, 168.
- Seismann, H., Blank, S., Braren, I., Greunke, K., Cifuentes, L., Grunwald, T., Bredehorst, R., Ollert, M., Spillner, E., 2010. Dissecting cross-reactivity in hymenoptera venom allergy by circumvention of alpha-1,3-core fucosylation. *Mol. Immunol.* 47, 799–808.
- Sturm, G.J., Hemmer, W., Hawranek, T., Lang, R., Ollert, M., Spillner, E., Blank, S., Bokanovic, D., Aberer, W., 2011. Detection of IgE to recombinant Api m 1 and rVes v 5 is valuable but not sufficient to distinguish bee from wasp venom allergy. *J. Allergy Clin. Immunol.* 128, 247–248.
- Sturm, G.J., Bilo, M.B., Bonadonna, P., Hemmer, W., Caruso, B., Bokanovic, D., Aberer, W., 2012. Ves v 5 can establish the diagnosis in patients without detectable specific IgE to wasp venom and a possible north-south difference in Api m 1 sensitization in Europe. *J. Allergy Clin. Immunol.* 130, 817.
- Van Vaerenbergh, M., Cardoen, D., Formesyn, E.M., Brunain, M., Van Driessche, G., Blank, S., Spillner, E., Verleyen, P., Wenseleers, T., Schoofs, L., Devreese, B., de Graaf, D.C., 2013. Extending the honey bee venom with the antimicrobial peptide apidaecin and a protein resembling wasp antigen 5. *Insect Mol. Biol.* 22, 199–210.
- Van Vaerenbergh, M., Debyser, G., Devreese, B., de Graaf, D.C., 2014. Exploring the hidden honeybee (*Apis mellifera*) venom proteome by integrating a combinatorial peptide ligand library approach with FTMS. *J. Proteomics* 99, 169–178.
- Wagner, S., Radauer, C., Bublin, M., Mann-Sommergruber, K.H., Kopp, T., Greisenegger, E.K., Vogel, L., Vieths, S., Scheiner, O., Breiteneder, H., 2008. Naturally occurring hypoallergenic Bet v 1 isoforms fail to induce IgE responses in individuals with birch pollen allergy. *J. Allergy Clin. Immunol.* 121, 246–252.
- Wong, K.L., Li, H., Wong, K.K.K., Jiang, T., Shaw, P.C., 2012. Location and reduction of icarapin antigenicity by site specific coupling to polyethylene glycol. *Protein Pept. Lett.* 19, 238–243.
- Wurtzen, P.A., 2001. Grass allergen-specific T-cells of atopic patients. *APMIS* 109, 561–571.
- Xiong, F., Gao, J.J., Li, J., Liu, Y., Feng, G.Y., Fang, W.L., Chang, H.F., Xie, J., Zheng, H.T., Li, T.Y., He, L., 2009. Noncanonical and canonical splice sites: a novel mutation at the rare noncanonical splice-donor cut site (IVS4 + 1A > G) of SEDL causes variable splicing isoforms in X-linked spondyloepiphyseal dysplasia tarda. *Eur. J. Hum. Genet.* 17, 510–516.

Doc. S1: A) Nucleotide sequences and GenBank accession numbers of sequenced mature Api m 10 amplicons, generated by reverse transcription-PCR on honeybee venom gland tissue. Alternative stop codons producing C-terminally truncated Api m 10 peptides are indicated in red.

>KC952889|ICA_var1

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTCTCGTCCTACCGTCCA
TTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTTTGAGCTTCGAGGACAGCGACGAGGGCTC
CAACTGGAAGTGAACACGCTGCTCAGACCTAACTTTTTGGATGGCTGGTACCAGACGTTACAAAGTGCAATTT
CAGCTCACATGAAGAAAGTGAGGGAGCAGATGGCTGGCATTGTGTCACGTATTCCCGAGCAAGGTGTTGTGA
ATTGGAACAAGATTCCTGAAGGAGCGAACACTACCTCTACCACCAAGATCATCGATGGACACGTGGTAACCAT
TAATGAAACGACTTATACCGATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGA
CCCCAGAATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACTTTGCCTACACTTA
TCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGGATTTGACAACGAGATAC
CGAAGAACCAAGGAGATGTATTAAGTCTTGA

>KC952890|ICA_var2

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTCTCGTCCTACCGTCCA
TTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTTTGAGCTTCGAGGACAGCGACGAGGG
CTCCAACTGGAAGTGAACACGCTGCTCAGACCTAACTTTTTGGATGGCTGGTACCAGACGTTACAAACT
CACATGAAGAAAGTGAGGGAGCAGATGGCTGGCATTGTGTCACGTATTCCCGAGCAAGGTGTTGTGAATT
GGAACAAGATTCCTGAAGGAGCGAACACTACCTCTACCACCAAGATCATCGATGGACACGTGGTAACCAT
TAATGAAACGACTTATACCGATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTG
AGACCCAGAATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACTTTGCCTA
CACTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGGATTTGACAA
CGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA

>KC952891|ICA_var3

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTCTCGTCCTACCGTCCA
TTGAAAGAGATCAAATGATGGCTGGCATTGTGTCACGTATTCCCGAGCAAGGTGTTGTGAATTGGAACAAGAT
TCCTGAAGGAGCGAACACTACCTCTACCACCAAGATCATCGATGGACACGTGGTAACCATTAATGAAACGACT
TATACCGATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCAGAATGAAA
CAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACTTTGCCTACACTTATCGGAAAGAATGA
GACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGGATTTGACAACGAGATACCGAAGAACCAAGG
AGATGTATTAAGTCTTGA

>KC952892|ICA_var4

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTCTCGTCCTACCGTCCA
TTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTTTGAGCTTCGAGGACAGCGATGTGACCAC
TTTGCCTACACTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGGATTT
GACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA

>KC952893|ICA_var5

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTCTCGTCCTACCGTCCA
TTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTTTGAGCTTCGAGGATTTGACAACGAGAT
ACCGAAGAACCAAGGAGATGTATTAAGTCTTGA

>KC952894|ICA_var6

TTCCCTGGTGCACACGATGTGACCACTTTGCCTACACTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAG
GAGTGTGGAAAGCGTCGAGGATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA

>KC952895|ICA_var7

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGAGACCAGCACCCAATCTTCAAGGAGT
GTGGAAAGCGTCGAGGATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA

>KC952896|ICA_var8

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGG
AAAGCGTCGAGGATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAACTGCTTGA

>KC952897|ICA_var9

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGCGAACACTACCTCTACCACCAAGATCATCGATGGACACGT
GGTAAACCATTAATGAAACGACTTATACCGATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATC
GACGTGAGACCCAGAAATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACTTTG
CCTACACTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGGATTTGACA
ACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA

>KC952898|ICA_var10

TTCCCTGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACGTGGTAAACCATTAATGAAA
CGACTTATACCGATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCAGAA
TGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACTTTGCCTACACTTATCGGAAAG
AATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGGATTTGACAACGAGATACCGAAGAAC
CAAGGAGATGTATTAAGTCTTGA

>KC952899|ICA_var11

TTCCCTGGTGCACACGATGAGGATAGCAAGGTGTTGTGAATTGGAACAAGATTCCTGAAGGAGCGAACACTA
CCTCTACCACCAAGATCATCGATGGACACGTGGTAACCATTAATGAAACGACTTATACCGATGGTAGCGACGA
CTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCAGAAATGAAACAATCTTGACAACGGTGAGC
AGCGAAGCGGACAGCGATGTGACCACTTTGCCTACACTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAA
GGAGTGTGGAAAGCGTCGAGGATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA

Doc. S1: B) All translated protein sequences of sequenced mature icarapin amplicons.

>ICA_var1

FPGAHDSEDSKEERKNVDTVLVLPISIERDQMMMAATFDFPSLSFEDSDEGSNWNWNTLLRPNFLDGWYQTLQSAISA
HMKKVREQMAGILSRIPEQGVVNWNIPEGANTTSTTKIIDGHVVTINETTYTDGSDDYSTLIRVRVIDVRPQNETIL
TTVSSEADSDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLTA

>ICA_var2

FPGAHDSEDSKEERKNVDTVLVLPISIERDQMMMAATFDFPSLSFEDSDEGSNWNWNTLLRPNFLDGWYQTLQTHM
KKVREQMAGILSRIPEQGVVNWNIPEGANTTSTTKIIDGHVVTINETTYTDGSDDYSTLIRVRVIDVRPQNETILTTV
SSEADSDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLTA

>ICA_var3

FPGAHDSEDSKEERKNVDTVLVLPISIERDQMMAGILSRIPEQGVVNWNIPEGANTTSTTKIIDGHVVTINETTYTDG
SDDYSTLIRVRVIDVRPQNETILTTSSEADSDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLTA

>ICA_var4

FPGAHDSEDSKEERKNVDTVLVLPISIERDQMMMAATFDFPSLSFEDSDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIP
KNQGDVLTA

>ICA_var5

FPGAHDSEDSKEERKNVDTVLVLPISIERDQMMMAATFDFPSLSFEDFDNEIPKNQGDVLTA

>ICA_var6

FPGAHDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLTA

>ICA_var7

FPGAHDSEDSKEERKNETSTQSSRSVESVEDFDNEIPKNQGDVLTA

>ICA_var8

FPGAHDSEDSKEERMRPAPNLQGVWKASRISTTRYRRTKEMY

>ICA_var9

FPGAHDSEDSKERTLPLPRSSMDTW

>ICA_var10

FPGAHDSEDSKEERKNVDTW

>ICA_var11

FPGAHDSEDSKVL

Doc. S2 shows the alignment of detected Api m 10 variants against the Api m 10 gene structure (Gene ID: 503505) on chromosome 1. Exons are shown in green, while introns are not colored. Intron 5'GT donor and 3'AG acceptor splice sites, conform to the general canonical GT-AG splicing rule, are presented in purple. For the chimeric transcripts, yellow boxes show short homologous sequences found at the junction sites of the source sequence.

EXON 1

chrom1	ATGAAGACCCTTGGCGTTCTATTTCATTGCAGCGTGGTTCATCGCATGCACGCACAGTTTC	60
ICAVar1	ATGAAGACCCTTGGCGTTCTATTTCATTGCAGCGTGGTTCATCGCATGCACGCACAGTTTC	60
ICAVar2	ATGAAGACCCTTGGCGTTCTATTTCATTGCAGCGTGGTTCATCGCATGCACGCACAGTTTC	60
ICAVar4	-----TTTC	3
ICAVar5	-----TTTC	3
ICAVAR11	-----TTTC	3
ICAVar9	-----TTTC	3
ICAVAR3	-----TTTC	3
ICAVAR10	-----TTTC	3
ICAVAR6	-----TTTC	3
ICAVAR8	-----TTTC	3
ICAVAR7	-----TTTC	3

chrom1	CCTGTAAGTACATGTTTTTTTACACATGGTCACAATAAATCAAACCTTTATTCTTTTTTC	120
ICAVar1	CCTG-----	64
ICAVar2	CCTG-----	64
ICAVar4	CCTG-----	7
ICAVar5	CCTG-----	7
ICAVAR11	CCTG-----	7
ICAVar9	CCTG-----	7
ICAVAR3	CCTG-----	7
ICAVAR10	CCTG-----	7
ICAVAR6	CCTG-----	7
ICAVAR8	CCTG-----	7
ICAVAR7	CCTG-----	7

EXON 2

chrom1	TCCACAAAGGTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTTC	24540
ICAVar1	-----GTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTTC	115
ICAVar2	-----GTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTTC	115
ICAVar4	-----GTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTTC	58
ICAVar5	-----GTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTTC	58
ICAVAR11	-----GTGCACACGATGAGGATAGCAAGG-----	31
ICAVar9	-----GTGCACACGATGAGGATAGCAAGGAG-----	33
ICAVAR3	-----GTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACACTGTTTC	58
ICAVAR10	-----GTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATGTGGACAC-----	53
ICAVAR6	-----GTGCACACGATG-----	19
ICAVAR8	-----GTGCACACGATGAGGATAGCAAGGAGGAAAG-----	38
ICAVAR7	-----GTGCACACGATGAGGATAGCAAGGAGGAAAGGAAGAATG-----	46

chrom1	TCGTCCTACCGTCCATTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTT	24600
ICAVar1	TCGTCCTACCGTCCATTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTT	175
ICAVar2	TCGTCCTACCGTCCATTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTT	185
ICAVar4	TCGTCCTACCGTCCATTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTT	118
ICAVar5	TCGTCCTACCGTCCATTGAAAGAGATCAAATGATGGCTGCAACTTTTCGATTTCCCGAGTT	118
ICAVAR11	-----	
ICAVar9	-----	
ICAVAR3	TCGTCCTACCGTCCATTGAAAGAGATCAAATGATGGCTG-----	97
ICAVAR10	-----	
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	TGAGCTTCGAGGACAGCGACGAGGGCTCCAACCTGGAACCTGGAACACGCTGCTCAGACCTA	24660
ICAVar1	TGAGCTTCGAGGACAGCGACGAGGGCTCCAACCTGGAACCTGGAACACGCTGCTCAGACCTA	235
ICAVar2	TGAGCTTCGAGGACAGCGACGAGGGCTCCAACCTGGAACCTGGAACACGCTGCTCAGACCTA	235
ICAVar4	TGAGCTTCGAGGACAGCGA-----	137
ICAVar5	TGAGCTTCGAGGA-----	131
ICAVAR11	-----	
ICAVar9	-----	
ICAVAR3	-----	
ICAVAR10	-----	
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	ACTTTTTGGATGGCTGGTACCAGACGTTACAAAGTAAAGGATAAAATATTTTCTTATAAT	24720
ICAVar1	ACTTTTTGGATGGCTGGTACCAGACGTTACAAA-----	268
ICAVar2	ACTTTTTGGATGGCTGGTACCAGACGTTACAAA-----	268
ICAVar4	-----	
ICAVar5	-----	
ICAVAR11	-----	
ICAVar9	-----	
ICAVAR3	-----	
ICAVAR10	-----	
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

EXON 3

chrom1	TATGCAATGAAAAATTCATTATATTCTTATCAATGTAATAATATCTTTTAGGTGCAATT	38760
ICAVar1	-----GTGCAATT	276
ICAVar2	-----	
ICAVar4	-----	
ICAVar5	-----	
ICAVAR11	-----	
ICAVar9	-----	
ICAVAR3	-----	
ICAVAR10	-----	
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	TCAGCTCACATGAAGAAAGTGAGGGAGCAGATGGCTGGCATTGTCACGTATTCCCGAG	38820
ICAVar1	TCAGCTCACATGAAGAAAGTGAGGGAGCAGATGGCTGGCATTGTCACGTATTCCCGAG	336
ICAVar2	--CTCACATGAAGAAAGTGAGGGAGCAGATGGCTGGCATTGTCACGTATTCCCGAG	324
ICAVar4	-----	
ICAVar5	-----	
ICAVAR11	-----	
ICAVar9	-----	
ICAVAR3	-----GCATTGTCACGTATTCCCGAG	120
ICAVAR10	-----	
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	CAAGGTGTTGTGAATTGGAACAAGATTCCTGAAGGAGCGAACACTACCTCTACCACCAAG	38880
ICAVar1	CAAGGTGTTGTGAATTGGAACAAGATTCCTGAAGGAGCGAACACTACCTCTACCACCAAG	396
ICAVar2	CAAGGTGTTGTGAATTGGAACAAGATTCCTGAAGGAGCGAACACTACCTCTACCACCAAG	384
ICAVar4	-----	
ICAVar5	-----	
ICAVAR11	-----TGTTGTGAATTGGAACAAGATTCCTGAAGGAGCGAACACTACCTCTACCACCAAG	86
ICAVar9	-----CGAACACTACCTCTACCACCAAG	56
ICAVAR3	CAAGGTGTTGTGAATTGGAACAAGATTCCTGAAGGAGCGAACACTACCTCTACCACCAAG	180
ICAVAR10	-----	
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	GTAGTCACGCGTAAATATTATGCTCATTTCTCGCTCGGCCAATCAGTTCAGTCACTGACTCA	38940
ICAVar1	---	
ICAVar2	---	
ICAVar4	-----	
ICAVar5	-----	
ICAVAR11	---	
ICAVar9	-----	
ICAVAR3	---	
ICAVAR10	-----	
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

EXON 4

chrom1	CATTGTTTTTTAGATCATCGATGGACACGTGGTAACCATTAATGAAACGACTTATACCG	39896
ICAVar1	-----ATCATCGATGGACACGTGGTAACCATTAATGAAACGACTTATACCG	442
ICAVar2	-----ATCATCGATGGACACGTGGTAACCATTAATGAAACGACTTATACCG	430
ICAVar4	-----	
ICAVar5	-----	
ICAVAR11	-----ATCATCGATGGACACGTGGTAACCATTAATGAAACGACTTATACCG	132
ICAVar9	-----ATCATCGATGGACACGTGGTAACCATTAATGAAACGACTTATACCG	102
ICAVAR3	-----ATCATCGATGGACACGTGGTAACCATTAATGAAACGACTTATACCG	226
ICAVAR10	-----GTGGTAACCATTAATGAAACGACTTATACCG	84
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	ATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCCAGA	39956
ICAVar1	ATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCCAGA	502
ICAVar2	ATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCCAGA	490
ICAVar4	-----	
ICAVar5	-----	
ICAVAR11	ATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCCAGA	192
ICAVar9	ATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCCAGA	162
ICAVAR3	ATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCCAGA	286
ICAVAR10	ATGGTAGCGACGACTACAGTACACTGATCCGCGTCCGTGTGATCGACGTGAGACCCCAGA	144
ICAVAR6	-----	
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	ATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACCTTTGCCTA	40016
ICAVar1	ATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACCTTTGCCTA	562
ICAVar2	ATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACCTTTGCCTA	550
ICAVar4	-----TGTTGACCACCTTTGCCTA	154
ICAVar5	-----	
ICAVAR11	ATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACCTTTGCCTA	252
ICAVar9	ATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACCTTTGCCTA	222
ICAVAR3	ATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACCTTTGCCTA	346
ICAVAR10	ATGAAACAATCTTGACAACGGTGAGCAGCGAAGCGGACAGCGATGTGACCACCTTTGCCTA	204
ICAVAR6	-----TGTTGACCACCTTTGCCTA	34
ICAVAR8	-----	
ICAVAR7	-----	

chrom1	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	40076
ICAVar1	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	622
ICAVar2	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	610
ICAVar4	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	214
ICAVar5	-----	
ICAVAR11	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	312
ICAVar9	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	282
ICAVAR3	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	406
ICAVAR10	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	264
ICAVAR6	CACCTTATCGGAAAGAATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	94
ICAVAR8	-----AATGAGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	84
ICAVAR7	-----AGACCAGCACCCAATCTTCAAGGAGTGTGGAAAGCGTCGAGG	88

chrom1	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	40126
ICAVar1	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	672
ICAVar2	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	660
ICAVar4	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	264
ICAVar5	-TTTCGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	180
ICAVAR11	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	362
ICAVar9	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	332
ICAVAR3	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	456
ICAVAR10	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	314
ICAVAR6	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	144
ICAVAR8	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	134
ICAVAR7	ATTTGACAACGAGATACCGAAGAACCAAGGAGATGTATTAAGTCTTGA	138

Doc. S3 shows the isoform sequences produced as synthetic constructs and as prokaryotic recombinants. His-tags are shown in red.

1. Prokaryotic recombinants

>ICA_VAR1

MAHHHHHHVDDDDKM FPGAHDEDSKEERKNVDTVLVLPSIERDQMM AATFDFPSLSFEDSDEGSNWNW
NTLLRPNFLDGWYQTLQSAISAHMKKVREQMAGILSRIPEQGVVNWNKIPEGANTTSTTKIIDGHVVT
INETTYTDGSDDYSTLIRVRVIDVRPQNETILTTVSSEADSDVTTLPTLIGKNETSTQSSRSVESVED
FDNEIPKNQGDVLT A

>ICA_VAR2

FPGAHDEDSKEERKNVDTVLVLPSIERDQMM AATFDFPSLSFEDSDEGSNWNWNTLLRPNFLDGWYQ
T LQTHMKKVREQMAGILSRIPEQGVVNWNKIPEGANTTSTTKIIDGHVVTINETTYTDGSDDYSTLIRV
RVIDVRPQNETILTTVSSEADSDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLT A

>ICA_VAR3

MAHHHHHHVDDDDKM FPGAHDEDSKEERKNVDTVLVLPSIERDQMM AAGILSRIPEQGVVNWNKIPEGA
NTTSTTKIIDGHVVTINETTYTDGSDDYSTLIRVRVIDVRPQNETILTTVSSEADSDVTTLPTLIGKN
ETSTQSSRSVESVEDFDNEIPKNQGDVLT A

>ICA_VAR4

MAHHHHHHVDDDDKM FPGAHDEDSKEERKNVDTVLVLPSIERDQMM AATFDFPSLSFEDSDVTTLPTL
IGKNETSTQSSRSVESVEDFDNEIPKNQGDVLT A

>ICA_VAR5

MAHHHHHHVDDDDKM FPGAHDEDSKEERKNVDTVLVLPSIERDQMM AATFDFPSLSFEDFDNEIPKNQ
GDVLT A

>ICA_VAR7

MAHHHHHHVDDDDKM FPGAHDEDSKEERKNETSTQSSRSVESVEDFDNEIPKNQGDVLT A

>Uterocalin

MAHHHHHHVDDDDKM LHMGPDPNFDEKLVKGKWF SVALASNEPKFIAKDTDMKFFIHKIQVTPESLQ
FHFHRKVRGMCVPTMMTAHKTKKKFQYTVNHSGHKTIFLEKVDPKHFVIFCAHSMKHGKETVVVTLFS
RTPTVSPDVMWMFKKYCKTHGIHTSNIVDLTQTDRCLHARH

2. Synthetic constructs

>ICA_VAR6

FPGAHDVTTLPTLIGKNETSTQSSRSVESVEDFDNEIPKNQGDVLT A

>ICA_VAR8

FPGAHDEDSKEERM RPAPNLQGVWKASRISTTRYRRTKEMY

>ICA_VAR9

FPGAHDEDSKERTLPLPRSSMDTW
2870.15 Da

>ICA_VAR10

FPGAHDEDSKEERKNVDTW
2260.36 Da

>ICA_VAR11
FPGAHDEDSKVL