

Molecular cloning, expression, purification and *in silico* epitope prediction of cobalamin-independent methionine synthase (Mor a 2), as a novel allergen from *Morus alba* pollen

Yunus Aksüt¹, Murat Pekmez², Semian Karaer Uzuner², Aslı Gelincik³, Suna Büyüköztürk³, And Nazlı Arda^{2,4}

¹Department of Molecular Biology and Genetics, Institute of Graduate Studies in Sciences, Istanbul University, Istanbul, Turkey

²Department of Molecular Biology and Genetics, Faculty of Science, Istanbul University, Istanbul, Turkey

³Division of Immunology and Allergic Diseases, Department of Internal Medicine, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey

⁴Center for Research and Practice in Biotechnology and Genetic Engineering (BIYOGEM), Istanbul University, Istanbul, Turkey

Submitted: 10 October 2020; **Accepted:** 1 May 2021

Online publication: 12 May 2021

Corresponding author:

Yunus Aksüt
Department of Molecular Biology and Genetics
Institute of Graduate Studies in Sciences
Istanbul University
34134, Vezneciler-Istanbul, Turkey
Phone: +905356331085
E-mail: aksut.yns@gmail.com

Arch Med Sci

DOI: <https://doi.org/10.5114/aoms/136322>

Copyright © 2022 Termedia & Banach

Abstract

Introduction: *Morus alba* (white mulberry) pollen is an aero-allergen source that can trigger allergic diseases. Cobalamin-independent methionine synthase (MetE) in *M. alba* pollen has been proved to be one of the major allergens for some patients living in Istanbul (Turkey). The aim of the present study was to carry out recombinant production and identification of MetE (Mor a 2), a novel allergen from *M. alba* pollen. The IgE binding reactivity of rMor a 2 produced for the first time was evaluated and some structural features were investigated by *in silico* methods to better understand its immunogenicity.

Material and methods: The gene encoding Mor a 2 was cloned in fission yeast, *Schizosaccharomyces pombe ura4-D18h* strain, using the pSLF1073 vector. This is the first report of the production of recombinant pollen allergen in *S. pombe*. After the purification, immunoreactivity of rMor a 2 was confirmed by immunoblotting using sera of a patient allergic to *M. alba* pollen. Moreover, B-cell epitopes of rMor a 2 were predicted using various bioinformatic tools, namely Bioinformatics Predicted Antigenic Peptides, BepiPred 2.0 and Immune Epitope Database, whereas T-cell epitopes were estimated using NetMHCIIpan-3.2 and NetMHCII 2.3 servers.

Results: The immunoblotting analysis yielded 11 of 11 positive reactions to rMor a 2. *In silico* predictions exerted seven B-cell epitopes (22-33, 384-394, 407-423, 547-553, 571-577, 671-678, 736-741) and seven T-cell epitopes (54-62, 161-170, 197-205, 347-358, 622-630, 657-665, 756-764).

Conclusions: These findings may support the use of rMor a 2 in the diagnosis and treatment of allergic diseases associated with *M. alba* and/or MetE.

Key words: *Morus alba*, pollen allergy, recombinant allergen, Mor a 2, cobalamin-independent methionine synthase (MetE), *Schizosaccharomyces pombe*, B- and T-cell epitopes.

Introduction

Allergic diseases are among the major health problems of the present time and affect about 20–30% of individuals in all age groups from every continent [1]. One of the most important aeroallergens causing allergic reactions and diseases is pollens. High concentrations of various allergenic pollens in the air trigger a variety of allergic reactions, including seasonal rhinoconjunctivitis and bronchial asthma, and reduce the life quality of susceptible individuals. Allergenic molecules in pollens are predominantly water-soluble proteins or glycoproteins [1–3]. These proteins have a molecular weight of 5–150 kDa, and an isoelectric point between 4 and 9 [3].

Mulberry (*Morus* spp.), a member of the Moraceae family, is a broad-leaved tree with diverse species that grows in temperate and subtropical regions around the world. White mulberry (*Morus alba*) pollen has been shown to cause respiratory allergies and contact urticaria [4, 5]. In particular, the incidence of *M. alba* related allergy cases increases in the pollen season (April-May) [6, 7]. Although small proteins (7–17 kDa) are mostly responsible for allergies associated with many mulberry species [5, 8–11], an 84 kDa protein, cobalamin-independent methionine synthase (MetE), in *M. alba* has been identified by our research group as a potential allergen for some patients living in Istanbul, a metropolitan city of Turkey [12].

MetE, a member of the zinc metalloenzyme family, is an enzyme that catalyzes the transfer of a methyl group from methyltetrahydrofolate to L-homocysteine, which is the last step of methionine biosynthesis [13]. It is found in bacteria, fungi, and plants. MetE has been shown as an allergenic protein in two studies. The first study was conducted in *Brassica napus* pollen by Chardin *et al.* [14] and the second one in *Salsola kali* by Assarehzadegan *et al.* [15].

Allergens are produced recombinantly by advanced genetic engineering techniques for further investigations and clinical use in diagnosis and treatment. Recombinant allergens are structurally and immunologically identical to natural allergens. They can be produced with constant quality, in large amounts, and purely by appropriate host organisms that are determined according to experience available in the laboratory, and the nature of the allergen [16, 17]. Over the years, with these advantages, recombinant DNA technology has popularized the use of recombinant allergens in the characterization of allergens, and also allergy diagnosis and treatment. In particular, the poor quality and insufficiency of natural allergen extracts for allergen immunotherapy have shown the necessity of recombinant production of allergens [18].

Allergen-specific immunotherapy (AIT) is an immunomodulatory treatment of allergic diseases used over a hundred years. The efficacy and safety of AIT have been demonstrated in numerous clinical trials [19]. Ciprandi *et al.* [20, 21] determined the efficacy of the treatment depending on the cytokine levels of allergic patients receiving specific-immunotherapy. To develop AIT, identification of epitopes responsible for allergic responses has recently come to the fore. T- and B-cell epitope peptides have an important role in the design of appropriate hypoallergenic AIT vaccines [19]. However, no study has been conducted on the epitopes of MetE allergen in *M. alba* pollen to date. Thus, we designed the present study in order to better understand the structure and immunogenicity of *M. alba* MetE.

Herein, we cloned, expressed, and purified recombinant MetE (rMetE). IgE binding activities of rMetE were analyzed by immunoblotting analysis. The new allergen (MetE) of the *M. alba* pollen was designated as Mor a 2 according to the WHO/IUIS system of allergen nomenclature. We also used bioinformatics to predict the secondary and tertiary protein structure of Mor a 2 and to identify the B-cell and T-cell epitopes. Our results suggest that the defined Mor a 2 structure, immunogenicity, and epitopes can help in the diagnosis and treatment of allergic diseases associated with *M. alba* and/or MetE. Methodology and the data of the current study are also expected to contribute to the development of new diagnostic tools and/or potential peptide-based vaccine design for white mulberry pollen allergy or other pollen allergies in general.

Material and methods

In the first step of this study, we cloned the *MetE* gene of *M. alba* pollen in fission yeast in order to produce recombinant MetE. Afterward, we checked the IgE binding activity of rMetE by immunoblotting with patients' sera. The allergenic MetE protein from *M. alba* pollen, as a new allergen, was designated Mor a 2 by the WHO/IUIS Allergen Nomenclature Subcommittee (<http://www.allergen.org/viewallergen.php?aid=1044>). In the final step, we determined the secondary and tertiary protein structure and B-cell and T-cell epitopes of Mor a 2 using bioinformatics tools.

Ethics statement

After receiving ethical approval from the Ethics Committee of Istanbul Faculty of Medicine (protocol number:1057 2016/109, 29th August 2016), herein, sera of 11 patients and 3 control groups used in our previous study [12] were tested in the IgE binding capacity.

Total RNA isolation and cDNA synthesis

Total RNA was isolated from *M. alba* pollen collected in our previous study [12] by using the Total RNA Miniprep Purification Kit (Genemark Bio), according to the manufacturer's instructions. The reverse transcription reaction was performed with 1 µg of total RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The reaction was carried out at 25°C for 15 minutes followed by 37°C for 120 minutes and denaturation at 85°C for 5 minutes.

The cDNA was amplified by polymerase chain reaction (PCR) using degenerate primers (F1: 5' ATG KCR TCY CAC RTY GTC GGT 3', R1: 5' TCA CTT GGC ACT GGC CAA YTG 3') that were designed from the known MetE sequences of different plants. The PCR product was isolated from 1% agarose gel by the PureLink Quick Gel Extraction Kit (Invitrogen), sequenced by the Sanger method, and then analyzed by Medsantek Sequence Laboratory (Istanbul, Turkey) for confirmation. The deduced amino acid sequences were analyzed using the Blast program on the NCBI website (<http://www.ncbi.nlm.nih.gov>).

Expression of recombinant Mor a 2

The coding region of Mor a 2 was amplified with *Pfu* DNA polymerase (iNtRON Biotechnology), using two specific primers. The forward primer contained the *Not* I restriction site (underlined) (F2: 5' TTT TGC GGC CGC ATG GCA TCC CAC ATC GTC GGT 3'). The reverse primer contained the *Sal* I restriction site (underlined) (R2: 5' TTG CCC GTC GAC TCA CTT GGC ACT GGC CAA CTG 3'). The amplicon was digested by the *Not* I (Anza 1, Invitrogen) and the *Sal* I (Anza 14, Invitrogen) restriction enzymes according to the manufacturer's protocol. The digested amplicon was ligated into the pSLF1073 plasmid (National BioResource Project). According to Gregan *et al.* [22], the recombinant plasmid was transformed into the *S. pombe* strain *ura4-D18h⁻*. Briefly, yeasts were grown in YE media at 32°C overnight to mid-log phase, and harvested by centrifugation; the pellet was washed with 10 ml of TE/LiAc (10 mM Tris, pH 7.6, 1 mM EDTA, 100 mM lithium acetate) twice and then resuspended in TE/LiAc. For transformation, 150 µl of cells were mixed with 15 µl of recombinant plasmid and 2 µg of salmon sperm DNA. After incubation at room temperature for 10 minutes, 375 µl of TE/LiAc/PEG (40% PEG in TE/LiAc) was added. Cells were incubated at 32°C for 30 minutes, then subjected to heat shock by incubation for 5 minutes at 46°C. Cells were centrifuged, removed from the supernatant, resuspended in 5 ml of YE, then incubated overnight at 32°C with shaking. The transformants were selected

on MMA medium. A single transformant colony was inoculated in MML medium + 10 µM thiamine overnight at 30°C (by shaking at 200 rpm). Cells were harvested at 2000 rpm for 5 min at 4°C, washed twice in dH₂O, and resuspended in 50 ml of MML medium without thiamine, of which 250 µl of cells were used to inoculate 250 ml of MML medium without thiamine.

Protein isolation and purification

Subsequent to growing in 250 ml of MML medium without thiamine at 30°C for 48 hours, the transformant cells were harvested by centrifugation (10,000 rpm for 10 minutes at 4°C), resuspended in lysis buffer (100 mM Tris-HCl pH 8.0, 1 mM DTT, 20% (v/v) glycerol), and then an equal volume of acid-washed glass beads (0.25 mm, B. Braun Melsungen AG) was added into the suspension. The cells were then disrupted by Mikro-Disembrators (Sartorius) for 1 minute, followed by incubation on ice for 1 minute. This step was repeated 5 times. The sample was centrifuged at 10,000 rpm for 10 minutes at 4°C, and the supernatant was stored at -20°C until use.

Recombinant Mor a 2 was fused to the C terminal 8XHis HA double tag. The recombinant protein was purified with the Profinia protein purification system (Bio-Rad) using the Native Immobilized Metal Affinity Chromatography (IMAC) method according to the manufacturer's instructions. The protein concentration of the purified protein was determined using the SMART BCA Protein Assay Kit (iNtRON Biotechnology). Equal quantities of protein (30 µg/well) were separated under reducing conditions by SDS-PAGE gel and stained with Coomassie (Bio-Safe Coomassie Stain, Bio-Rad) or silver staining (Pierce Silver Stain, Thermo Scientific).

Immunoblotting assays

Immunoreactivity of rMor a 2 was analyzed by Western blotting. The purified recombinant protein was separated by 10% SDS-PAGE and transferred onto a polyvinylidene difluoride (PVDF) membrane using the Trans-Blot Turbo Transfer System (Bio-Rad). After transfer, the membrane was blocked with 5% (w/v) non-fat dry milk [prepared in Tris-buffered saline (TBS) - (0.1% (v/v) Tween 20 (TBST))] for 2 hours at room temperature. The membrane was then incubated with HRP (horseradish peroxidase)-conjugated His tag Antibody (Novus Biologicals, NBP1-42785) (diluted 1 : 2000), or patients' sera obtained from our previous study (diluted 1 : 4) at 4°C overnight [12]. The membrane was washed with TBST five times for 5 minutes. The IgE-binding proteins on the membrane were detected using a 1 : 1000 dilution of

HRP-conjugated mouse anti-human IgE (Fc) antibody (Southern Biotech, 9160-05). The membrane was washed with TBST five times for 5 minutes, and then proteins were visualized by enhanced chemiluminescence (ECL) as recommended by the manufacturer (Pierce ECL Plus Western Blotting Substrate, Thermo Scientific) using ChemiDoc Imaging Systems (Bio-Rad).

Physicochemical analysis

Physicochemical analysis, computing the molecular weight, theoretical pI, amino acid composition, instability index, aliphatic index, and the grand average of hydropathicity (GRAVY) of Mor a 2, was performed using the ProtParam tool (<http://web.expasy.org/protparam/>) [23].

Secondary structure prediction

In order to identify conserved substructures, Mor a 2 secondary structural element recognition was assessed by PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>), which compares sequence segments with a template protein structure of high sequence similarity, since it has the highest published score among secondary structure prediction methods [24]. Furthermore, the secondary structure elements were identified and compared with the results obtained with NetSurfP ver. 1.1 (<https://services.healthtech.dtu.dk/service.php?NetSurfP-1.1>) [25] and PredictProtein (<http://www.predictprotein.org>) programs.

Homology modeling and validation

The Mor a 2 protein sequence was searched for homology in the RCSB Protein Data Bank (PDB) (<http://www.rcsb.org/>). The homologous template, which was suitable for Mor a 2, was selected from the SWISS-MODEL server [26]. An initial structural model was generated and checked for recognition of errors in three-dimensional (3D) structure by PROCHECK [27], ERRAT [28], and VERIFY_3D [29] programs in Structural Analysis and Verification Server (SAVES). The quality of the model was assessed by QMEAN [30], and also the protein stereology was tested with ProSA [31].

B-cell epitope prediction

Three immunoinformatics tools, Bioinformatics Predicted Antigenic Peptides (BPAP, <http://imed.med.ucm.es/Tools/antigenic.pl>), BepiPred 2.0 (<http://www.cbs.dtu.dk/services/BepiPred/>) [32], and Immune Epitope Database Analysis Resource (IEDB-AR, <http://tools.iedb.org/bcell>), were used to predict the B-cell associated epitopes of Mor a 2. Antigenicity, surface accessibility, flexibility, and hydrophilicity of the amino acid sequence were

chosen as parameters for B-cell epitopes in IEDB-AR, because B-cell epitope prediction performance is shown to be increased by combining several methods [33]. The data obtained from these immunoinformatics-based servers were combined to produce more accurate results.

T-cell epitope prediction

T-cell epitopes are principally predicted indirectly by identifying the peptide fragments with ability to bind to major histocompatibility complex (MHC) molecules. It has been reported that a sequence of nine amino acids in length (9-mer), known as a core peptide, is essential for MHC-class II binding [33]. The binding strength of each peptide to the given MHC molecule was estimated by NetMHCII 2.3 and NetMHCIIpan-3.2, depending on threshold values. NetMHCII-2.3 (<https://services.healthtech.dtu.dk/service.php?NetMHCII-2.3>) was applied for the human leukocyte antigen (HLA)-DQ-based T-cell epitope prediction by using HLA-DQA10101-DQB10501, HLA-DQA10501-DQB10201, HLA-DQA10501-DQB10301 and HLA-DQA10102-DQB10602 alleles. A certain sequence provided from the consensus result of at least three alleles was considered an epitope. HLA-DR-based T-cell epitopes were predicted by NetMHCIIpan-3.2 (<https://services.healthtech.dtu.dk/service.php?NetMHCIIpan-3.2>), which is an artificial neural network-based alignment method. HLA-DRB-10101, HLA-DRB30101, HLA-DRB40101, and HLA-DRB50101 alleles were intended for HLA-DR-based T-cell epitope prediction. The HLA-DR-based T-cell epitope was exhibited by combining the results obtained for these four alleles. The data obtained from HLA-DQ- and HLA-DR-based T-cell epitope predictions were combined in order to exhibit final T-cell epitopes. Finally, B-cell and T-cell epitopes identified by computational tools were mapped into a linear sequence and onto a 3D model of MetE to determine their positions [34].

Results

Amplification and sequence analysis of Mor a 2

An amplicon of approximately 2298 bp (open reading frame of *Mor a 2* gene) was produced from cDNA synthesized from *M. alba* pollen using degenerate primers designed according to the known MetE gene sequences of plant species. The cDNA sequence of Mor a 2 was submitted to the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov/>) under the accession number MN947650. The deduced amino acid sequence was compared to those of other plant species and found to be highly matched with MetE sequences of *Salsola kali* and *Brassica napus* plants (Figure 1).



Figure 1. Sequence alignment analysis of *Morus alba* MetE, and its comparison to those of different plant species. Blast analysis showed that MetE of *M. alba* shared a high degree of amino acid sequence identity with that Sal k 3 of *Salsola kali* (92%) and MetE of *Brassica napus* (91%). Sequences of T-cell and B-cell epitopes are framed as blue and red color and are labeled as T1-T7 and B1-B7, respectively. Conserved sequence (*), conservative mutations (:), semi-conservative mutations (.), and non-conservative mutations (.)

Cloning of Mor a 2 cDNA sequence

For the production of rMor a 2, firstly the Mor a 2 cDNA sequence was obtained by PCR amplification with specific primers and then cloned into the pSLF1073 vector. The pSLF1073 Mor a 2 vector was transformed into *E. coli*, and identified by single

digestion with *Not* I and by double digestion with *Not* I and *Sal* I. The restriction digest results are presented in Figure 2 A. Recombinant plasmid single digestion produced one band of approximately 10 750 bp, whereas double digestion generated two bands of approximately 8500 and 2300 bp.

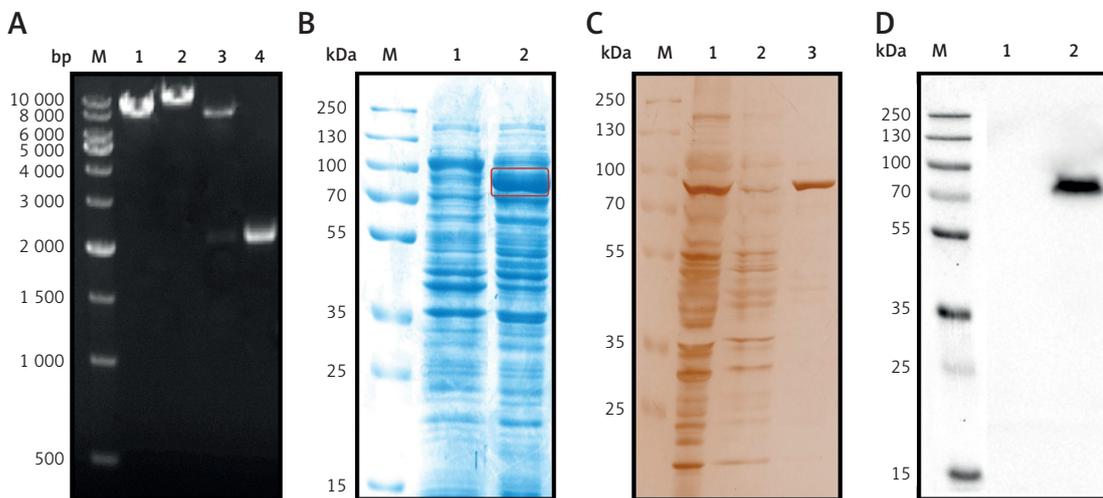


Figure 2. Verification of the recombinant Mor a 2. **A** – Restriction enzyme digestion of the recombinant plasmid showed successful cloning of the MetE gene into the vector pSLF1073; single and double digestion produced 10750 bp and 8500/2300 bp, respectively. Lane M, 1 kb DNA ladder (Dongsheng Biotech, M1181); lane 1, pSLF1073 plasmid digested by *Not* I; lane 2, pSLF1073-Mor a 2 plasmid digested by *Not* I; lane 3, pSLF1073-Mor a 2 plasmid digested by *Not* I and *Sal* I; lane 4, PCR product of Mor a 2 gene. **B** – A band of MW ~85 kDa (rMor a 2, in the red frame) on Coomassie Brilliant Blue stained SDS-PAGE gel. Lane M, PageRuler Plus Prestained Protein Ladder (Thermo Scientific, 266192); lane 1, cell lysate from control cells of *S. pombe*; lane 2, cell lysate from transformant cells of *S. pombe*. **C** – Purified protein band (rMor a 2) after SDS-PAGE and silver staining. Lane M, Protein Ladder; lane 1, cell lysate from transformant cells of *S. pombe*; lane 2, elution with 20 mM imidazole; lane 3, elution with 500 mM imidazole. **D** – His-tagged recombinant proteins were transferred to PVDF membrane and detected by immunoblotting using anti-His tag antibody to confirm rMor a 2 expression. Lane M, Protein Ladder; lane 1, cell lysate from control cells of *S. pombe*; lane 2, cell lysate from transformant cells of *S. pombe*

Recombinant production and purification of Mor a 2

Transformant cells were harvested, and total proteins were isolated by disrupting the cells. His-tagged rMor a 2 protein was purified. Before purification, proteins were separated under denaturing conditions by SDS-PAGE and stained with Coomassie bright blue staining (Figure 2 B). A thick band of approximately 85 kDa molecular weight was detected in the total protein extract. The rMor a 2 was purified by elution from a Ni-column with 250 mM of imidazole. The purity of rMor a 2 was checked by silver staining after SDS-PAGE (Figure 2 C), and its yield was quantified using the BCA protein assay. Approximately 1 mg of rMor a 2 was obtained from 1 l of yeast culture.

The identity of the recombinant protein was confirmed by immunoblot analysis using anti-His tag antibody (Figure 2 D). Serum samples from 11 allergy patients obtained in our previous study [12] were used to investigate the IgE reactivity of rMor a 2 by immunoblotting. IgE antibodies from all serum samples, except that of one patient, intensely reacted with the purified rMor a 2 (Figure 3). Only one serum sample produced weak interaction with rMor a 2, probably due to the low concentration of IgE. No immune reaction was observed with the sera from controls (healthy individuals).

Predicted physicochemical properties

The data obtained from the ProtParam server showed that rMor a 2 had a molecular weight of

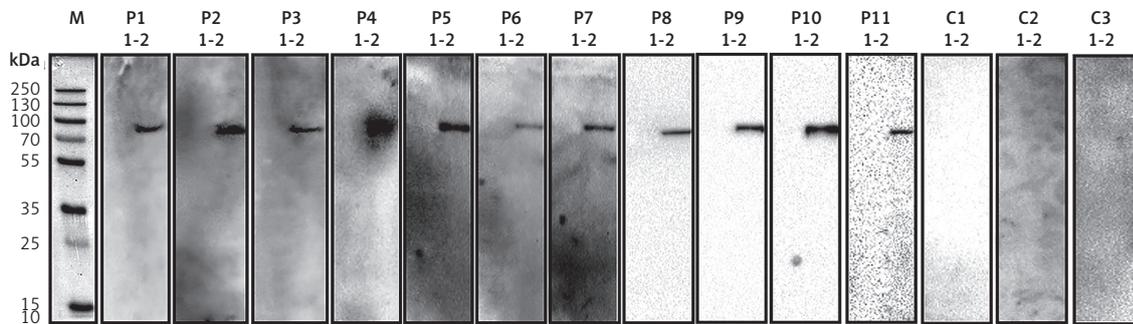


Figure 3. Analysis of specific IgE reactivity of recombinant Mor a 2 by immunoblot assay. Sera collected from patients allergic to *M. alba* pollen and healthy controls in our previous study [12] were used. Lane M, Protein Ladder; lane 1, cell lysate from control cells of *S. pombe*; lane 2, purified rMor a 2; P1-P11, IgE immunoblotting analysis of the sera of patients (P) with *M. alba* allergy; C1-C3, IgE immunoblotting analysis of the sera of controls (C – healthy subject)

Table I. Predicted secondary structure of Mor a 2

Tools	α -helices	β -sheets
PSIPRED	16–25, 31–51, 65–74, 80–83, 91–98, 136–146, 158–164, 177–197, 210–211, 216–232, 249–256, 275–279, 297–298, 303–317, 345–350, 353–369, 374–388, 396–404, 407–410, 416–427, 444–454, 460–481, 496–503, 540–548, 565–566, 576–596, 606–610, 617–634, 653–662, 678–686, 708–721, 724–726, 740–763	3–5, 56–57, 119–120, 131–132, 154–156, 202–206, 238–242, 263–267, 269–270, 287–293, 322–325, 507–510, 513–517, 527–533, 557–559, 562–564, 601–604, 643–648, 667–671, 727–729
Predict protein	15–25, 31–51, 69–72, 138–142, 144–144, 158–166, 177–197, 216–232, 251–256, 277–279, 303–314, 345–350, 354–369, 374–379, 381–381, 444–454, 460–481, 500–503, 540–547, 575–598, 601–601, 607–614, 618–636, 679–684, 709–718, 740–762	3–5, 56–58, 63–67, 73–75, 117–121, 201–206, 209–211, 237–244, 262–268, 287–293, 321–325, 433–437, 486–487, 492–493, 524–529, 557–557, 564–565, 602–603, 667–669, 692–695, 727–729
NetSurfP – 1.1	15–25, 31–51, 64–75, 92–97, 135–146, 158–167, 177–198, 216–232, 249–257, 275–281, 303–317, 345–396, 374–388, 396–406, 416–426, 444–455, 460–480, 494–503, 540–549, 561–569, 576–596, 607–612, 620–637, 653–662, 679–687, 708–722, 740–763	2–5, 56–58, 109–113, 118–124, 129–131, 152–156, 202–205, 209–211, 237–243, 264–268, 287–293, 322–325, 485–487, 507–510, 515–518, 521–523, 527–530, 554–558, 601–604, 643–648, 667–671, 692–697, 727–729
Overall results	15–25, 31–51, 65–74, 91–98, 136–146, 158–164, 177–197, 216–232, 249–256, 275–279, 303–317, 345–350, 353–369, 374–388, 396–404, 416–427, 444–454, 460–481, 496–503, 540–548, 576–596, 606–610, 617–634, 653–662, 678–686, 708–721, 740–763	3–5, 56–57, 119–120, 152–156, 201–206, 237–244, 262–268, 287–293, 321–325, 486–487, 507–510, 513–517, 527–533, 554–558, 601–604, 643–648, 667–671, 692–697, 727–729

84.323 kDa and consisted of 765 amino acids. The theoretical pI was 6.55 and the aliphatic index was 90.14. The GRAVY was -0.161 , indicating the hydrophilic character of Mor a 2. The instability index was 37.72 (< 40), indicating that Mor a 2 was a stable protein.

Predicted secondary structure of Mor a 2

The secondary structure of Mor a 2 predicted by three different tools is presented in Table I. Mor a 2 was estimated to consist of 33 α -helices and 20 β -sheets by PSIPRED, 26 α -helices and 21 β -sheets by PredictProtein and 27 α -helices and 23 β -sheets by NetSurfP v1.1. Overall results predicted by at least two programs were 27 α -helices and 19 β -sheets.

Tertiary structure prediction and validation

According to the results obtained in RCSB PDB, the protein with the accession number of 1u1j.1 (*Arabidopsis thaliana* MetE, AtMetE) showed the highest sequence identity (86.54%) to Mor a 2, and it was used for homology modeling. The overall 3D structure of Mor a 2 is shown in Figure 4 A. As indicated by the Ramachandran plot (Figure 4 B), 88.9% of the amino acid residues of Mor a 2 were within the most favored regions, 9.0% were within the additional allowed region, 1.5% were in generously allowed regions, and 0.5% were in

the disallowed region. The results of the ERRAT program (Figure 4 C) revealed that the overall quality factor was 97.323, which indicated that the structure had a high resolution. As indicated by the VERIFY_3D program, the results revealed that 91.70% of the residues had an average 3D (atomic model)-1D (amino acid sequence) score of > 0.2 ; this result was favorable. The ProSA server showed that the Z-score of Mor a 2 was -13.08 . The QMEAN Q value of Mor a 2 was -3.88 .

Prediction of B-cell epitopes

Fragment flexibility, surface accessibility, and hydrophobicity are important features for predicting antigenic epitopes. The antigenic index directly indicated the epitope-forming capacity of the Mor a 2 sequence. The ultimate results of the immunoinformatic tools (BPAP, BepiPred 2.0, and IEDB-AR) exhibited 7 segments (22-33, 384-394, 407-423, 547-553, 571-577, 671-678, 736-741) for B-cell epitope regions of Mor a 2. Sequences and positions of these epitopes are presented in Table II and projected in Figure 5 A.

Prediction of T-cell epitopes

Individual and total predicted regions of T-cell epitopes of Mor a 2 produced by combining data from HLA-DQ- and HLA-DR-based predictions are listed in Table III. HLA-DQ-based T-cell epitope

A

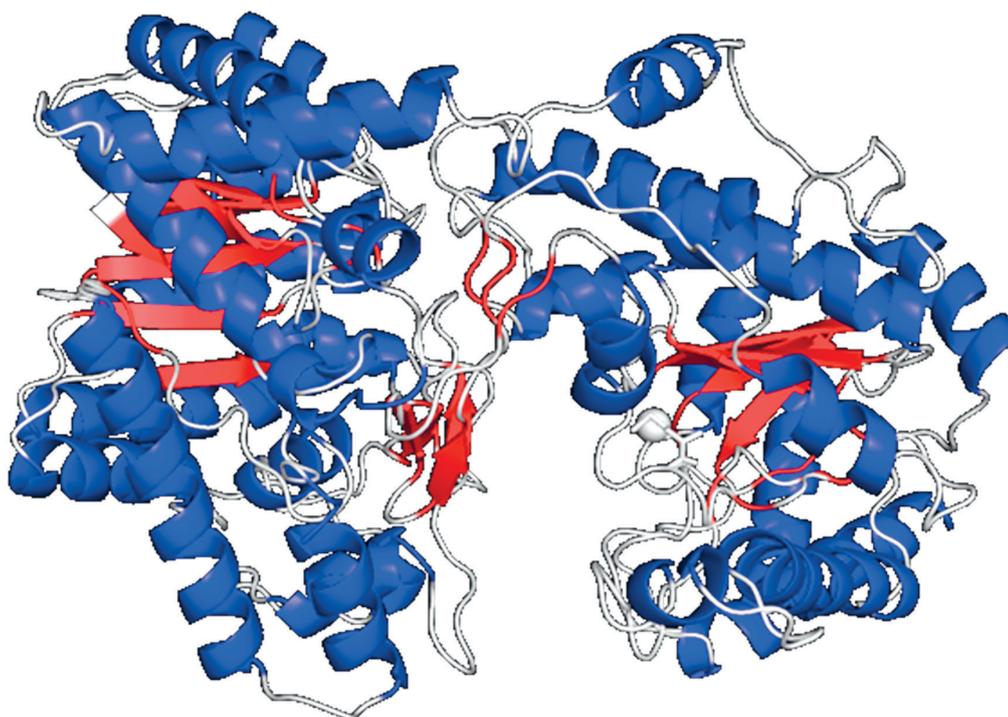


Figure 4. Three-dimensional structure and validation of Mor a 2 homology model. **A** – Protein structure of Mor a 2 homology model through RCSB PDB tool. The blue color indicates α -helices and the red color indicates β -sheets

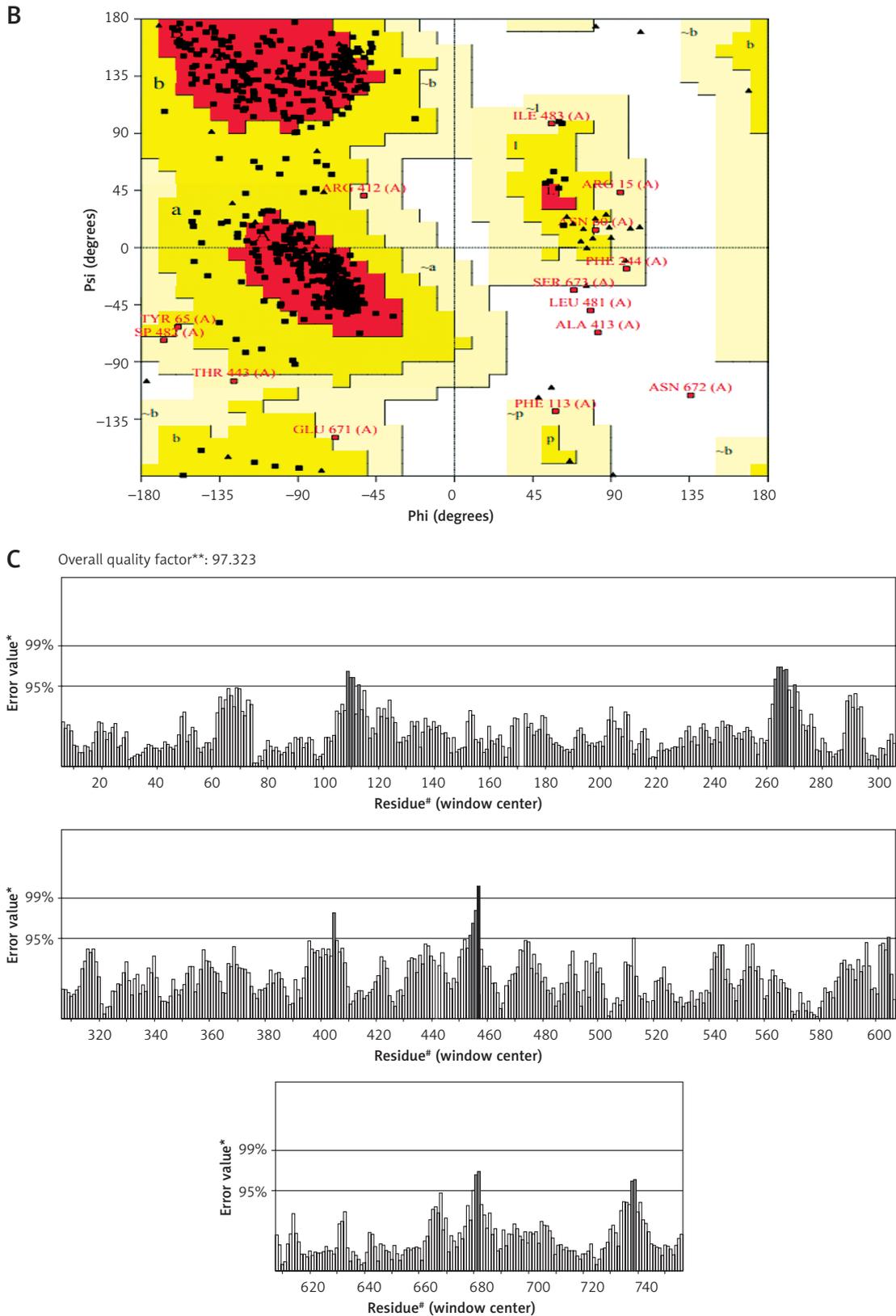


Table II. Predicted B-cell epitopes of Mor a 2

Peptide	Sequence	Position	Tools
B1	GSFWDGKSSAED	22–33	BepiPred-2.0, IEDB-AR**
B2	QASRRSSPRVT	384–394	BepiPred-2.0, IEDB-AR**
B3	GSDHRRATNVSSRLDAQ	407–423	BepiPred-2.0, IEDB-AR**
B4	QSTTKRP	547–553	IEDB-AR**
B5	RNNQPRF	571–577	BepiPred-2.0, IEDB-AR*
B6	ENSR SDEK	671–678	IEDB-AR**
B7	KTRKYS	736–741	BPAP, IEDB-AR*

*antigenicity, **antigenicity, surface accessibility, flexibility

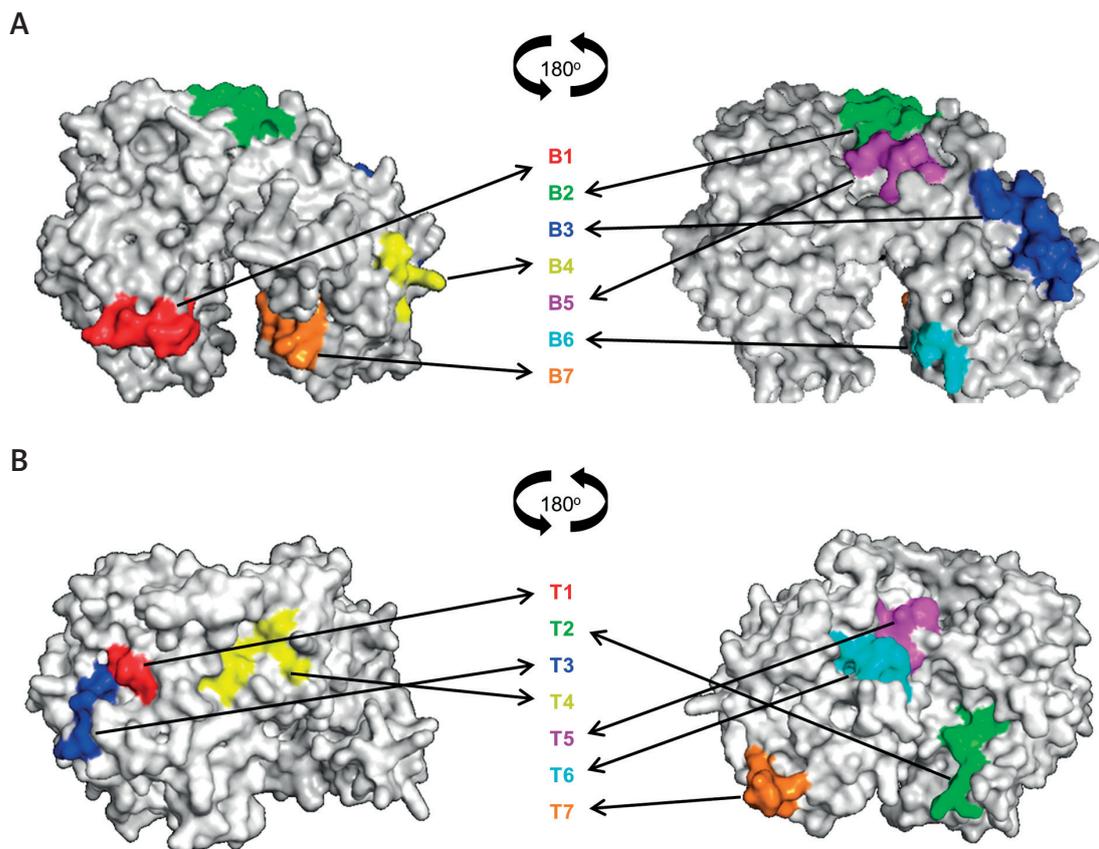


Figure 5. B-cell and T-cell epitope positions identified on the surface of the 3-dimensional structure of the Mor a 2 allergen. **A** – B-cell epitope prediction performed using available tools (BPAP, BepiPred 2.0, and IEDB-AR). Hydrophilicity, flexibility, accessibility, and antigenicity of the amino acid sequence were chosen as parameters for B-cell epitopes in the IEDB. B1-B7 are the predicted B-cell epitopes. **B** – NetMHCII 2.3 and NetMHCIIpan-3.2 were used to predict the core 9-mer T cell epitopes in the Mor a 2 allergen. T1-T7 are the predicted T-cell epitopes

prediction using HLA-DQA10101-DQB10501, HLA-DQA10501-DQB10201, HLA-DQA10501-DQB10301 and HLA-DQA10102-DQB10602 yielded 3 peptides (197-205, 350-358, 657-665). HLA-DR-based T-cell epitope prediction using HLA-DRB10101, HLA-DRB30101, HLA-DRB40101, and HLA-DRB50101 resulted in 5 peptides (54-62, 161-170, 347-358, 622-630, 756-764). Consequently, as shown in Table III and projected in Figure 5 B, Mor a 2 was predicted to have 7 T-cell epitope sequences; those

locations were 54-62, 161-170, 197-205, 347-358, 622-630, 657-665 and 756-764. One sequence located at 347-358 seemed to be mutual.

Discussion

Mulberry pollen is an allergen source associated with asthma, allergic rhinitis, and allergic conjunctivitis [4, 5, 8–11]. White mulberry (*M. alba*) pollen is known to trigger allergic diseases that are rhinoconjunctivitis, asthma, and urticaria [4, 5].

Table III. Predicted T-cell epitopes of Mor a 2

Prediction		Location of the prediction results
T-cell epitope HLA-DQ	HLA-DQA10101-DQB10501	60-68, 61-69, 62-70, 87-95, 88-96, 89-97, 109-117, 115-123, 116-124, 172-180, 179-187, 200-208, 201-209, 202-210, 240-248, 241-249, 242-250, 261-269, 330-338, 347-355, 350-358, 477-485, 478-486, 479-487, 496-504, 503-511, 509-517, 599-607, 618-626, 619-627, 620-628, 621-629, 622-630, 623-631, 628-636, 648-656, 649-657, 650-658, 656-664, 657-665, 660-668, 725-733
	HLA-DQA10102-DQB10602	69-77, 70-78, 101-109, 175-183, 197-205, 231-239, 234-242, 235-243, 298-306, 304-312, 305-313, 326-334, 327-335, 347-355, 349-357, 361-369, 362-370, 395-403, 536-544, 540-548, 620-628, 628-634, 657-665, 713-721, 716-724, 722-730, 745-753, 746-754, 747-755, 748-756, 751-759, 752-760, 753-761
	HLA-DQA10501-DQB10201	61-69, 62-70, 63-71, 64-72, 74-82, 116-124, 117-125, 146-154, 177-185, 180-188, 187-195, 198-206, 222-230, 235-243, 239-247, 240-248, 242-250, 243-251, 244-252, 261-269, 286-294, 330-338, 350-358, 463-471, 474-482, 478-486, 479-487, 492-500, 496-504, 583-591, 586-594, 599-607, 600-608, 656-664, 657-665, 660-668, 719-727
	HLA-DQA10501-DQB10301	49-57, 71-79, 72-80, 73-81, 74-82, 75-83, 84-92, 95-103, 98-106, 99-107, 100-108, 130-138, 144-152, 193-201, 195-203, 196-204, 197-205, 258-263, 278-286, 288-296, 362-370, 378-386, 379-387, 380-388, 396-404, 398-406, 400-408, 401-409, 428-436, 436-444, 504-512, 541-549, 554-562, 555-563, 594-602, 683-691, 688-696, 689-697, 690-698, 691-699, 743-751, 749-757, 751-759
T-cell epitope HLA-DR	HLA-DRB10101	54-62, 94-102, 95-103, 117-125, 128-136, 141-149, 160-168, 161-169, 162-170, 177-125, 194-202, 195-203, 353-361, 361-369, 376-384, 377-385, 252-260, 266-274, 347-355, 350-358, 554-562, 580-588, 581-589, 621-629, 622-630, 625-633, 715-723, 719-727, 743-751, 747-755, 754-762, 756-764
	HLA-DRB30101	37-45, 54-62, 56-64, 62-70, 67-75, 89-97, 111-119, 112-120, 113-121, 180-188, 243-251, 298-306, 299-307, 322-330, 350-358, 499-507, 527-535, 569-567, 601-609, 602-610, 615-623, 621-629, 622-630, 623-631, 652-660, 656-664, 660-668, 661-669, 710-719, 719-727
	HLA-DRB40101	41-49, 54-62, 161-169, 162-170, 177-185, 347-355, 350-358, 443-451, 444-452, 446-454, 471-479, 564-672, 656-664, 679-687, 715-723, 719-727, 750-758, 751-759, 753-761, 754-762, 755-763, 756-764
	HLA-DRB50101	41-49, 54-62, 91-99, 93-101, 94-102, 95-103, 117-125, 128-136, 129-137, 130-138, 141-149, 160-169, 161-169, 162-170, 181-189, 188-196, 251-259, 252-260, 266-274, 347-355, 349-357, 350-358, 364-372, 398-406, 443-451, 446-454, 449-457, 580-588, 609-617, 622-630, 625-633, 747-755, 750-758, 757-765
The final predicted T-cell epitopes		54-62, 161-170, 197-205, 347-358, 622-630, 657-665, 756-764

However, to date, only one specific allergen, MetE, from *M. alba* pollen has been found [12]. To the best of our knowledge, this is the first report in which recombinant production of the MetE was achieved, and its critical features for immunoreactivity were exhibited by an *in silico* approach.

The first study reported in 2001 that suggested MetE is an allergen showed that the amino acid sequence of an allergen protein of about 80 kDa from *Brassica napus* pollen was very similar to methionine synthase of the cobalamin-independent MetE of *Arabidopsis thaliana* [14]. Later on, MetE in *Salsola kali*, namely Sal k 3, was reported as an allergen by Assarehzadegan *et al.* [15]. Crystallographic studies on MetE suggested that almost all biological functions are located in the C-terminal domain [13]. However, in a study to determine which domain of Sal k 3 is an allergen, both the C- and N-terminal domains of the protein were re-

combinantly produced and found to have similar IgE binding capacities [15]. In the present study, we produced MetE protein of *M. alba* pollen recombinantly to contain both terminals, named it as Mor a 2, and found that the Mor a 2 shares a high degree of amino acid sequence identity with that of Sal k 3 (92%) and MetE of *Brassica napus* (91%). These results revealed that MetE proteins from these three plants share common IgE binding epitopes.

The allergenic properties, chemical constituents, and atmospheric concentrations of pollens vary according to climate, geographic location, and environmental conditions. Possibly due to these variations, there are some troubles in diagnosis and treatment of pollen allergy [2, 35]. In our previous study, MetE in *M. alba* pollen collected from Istanbul showed specific IgE susceptibility in 11 patients, while commercial *M. alba* pollen extract did so in only 2 patients [12]. This result suggests that in some cases

commercial extracts may not provide reliable results for patients living in different habitats, since allergenic constituents of pollens may differ, and thus the immune response against emergent allergenic proteins may vary under different climatic and environmental conditions. Therefore, the identification of allergenic proteins in authentic pollen samples is critical for the development of more reliable diagnostic kits and treatment agents.

Although prokaryotic systems are generally used for recombinant allergen production, genetically well-characterized yeast cells can offer higher yields with the advantages they provide including glycosylation, disulfide bond formation, and post-translational modification [16]. Therefore, many yeast species such as *Pichia pastoris* [36], *Saccharomyces cerevisiae* [37], *Yarrowia lipolytica* [38], and *Kluyveromyces lactis* [39] have been used to produce recombinant allergens. However, an important yeast species, *S. pombe*, was not used as a recombinant allergen production system. Genome-scale engineering of *S. pombe* via a large-scale gene deletion approach showed important potential for the development of host strains that have enhanced recombinant protein production ability [40]. Therefore, we used the *S. pombe* (*ura4-D18h*) strain and successfully produced a recombinant allergen in this organism for the first time. Firstly, here we cloned full-length cDNA expressing MetE from *M. alba* pollen in fission yeast for further investigations. Purified rMor a 2 showed specific IgE susceptibility in 11 patients, as in the previous study performed with natural MetE. Although recombinant production is an important step towards discovering the properties of the MetE protein and using it as diagnostic and immunotherapy tools in the future, the clinical results of the recombinant product also need to be evaluated. For example, the immunological response (e.g., inflammatory cytokines and molecules) should be demonstrated in allergic subjects [20, 21]. This is a limitation of our study. Other limitations of the study will be removed by performing mass spectrometric analysis to support the verification of Mor a 2 identification and by determination with ELISA [41] and the basophil activation test [42] whether there is a correlation between the amount of Mor a 2-specific IgE and the severity of allergic reactions.

Bioinformatics plays an important role in predicting the structure, properties, and B-cell and T-cell epitopes of target proteins. *In silico* prediction is a useful tool in selecting B-cell isotopes of allergens. It also correlates well with the experimental approach. Many tools have been developed for predicting B-cell epitopes based on parameters such as antigenicity, hydrophilicity, segmental mobility, flexibility, and accessibility based on the propensity of polypeptide chains [43].

We investigated several properties of Mor a 2 using bioinformatic tools. To further understanding the structure and function of Mor a 2, the basic sequence properties were analyzed. Herein, ProtParam predicted that the GRAVY score of Mor a 2 was -0.161 , which indicated hydrophilic characteristics. Moreover, having an instability index below 40 indicated the stability of Mor a 2 [23]. Homology modeling was applied to investigate the 3D structure of Mor a 2 by using the PDB; the structure of 1u1j.1 (AtMetE) was the most appropriate template with marked identity to Mor a 2. The Ramachandran plot showed that a total of 97.9% of residues of the Mor a 2 model were in favored and allowed regions, indicating that the distribution of the amino acids in the Mor a 2 model was reasonable. The overall quality factor was determined as 97.3226% by the ERRAT program and 91.70% of the residues had an average 3D-1D score ≥ 0.2 according to the VERIFY 3D program. The data revealed that the tertiary structures of Mor a 2 were favorable and had a high resolution. All of the validations showed that the structural model of Mor a 2 seemed to be correct.

In the present study, we predicted the B-cell linear epitopes of Mor a 2 allergen by 3 sequence-based tools (BPAP, BepiPred 2.0, and IEDB-AR) and predicted 7 peptides (22–33, 384–394, 407–423, 547–553, 571–577, 671–678, 736–741) as potential B-cell epitopes. In addition, NetMHCIIpan-3.2 and NetMHCII 2.3 were used to estimate T-cell epitopes of the Mor a 2 allergen. Seven potential T-cell epitope sequences (54–62, 161–170, 197–205, 347–358, 622–630, 657–665 and 756–764) were identified. This is the first report on B- and T-cell epitope prediction for Mor a 2. Our data reveal that both N- and C-domains are involved in the allergenicity of Mor a 2. Molecular characteristics that have been theoretically exhibited may help gain a better understanding of its immunological effects. Overall, this study is expected to contribute to the use of rMor a 2 in the diagnosis and treatment of patients after further clinical trials and may accelerate the discovery and recombinant production of authentic pollen allergens.

Acknowledgments

We cordially thank Aslihan Şengelen, PhD candidate at Istanbul University, for her help during immunoblotting assays.

The study was supported by a grant from the Scientific Research Projects Coordination Unit of Istanbul University, Turkey (grant number 57936).

Conflict of interest

The authors declare no conflict of interest.

References

- Puc M. Characterisation of pollen allergens. *Ann Agric Environ Med* 2003; 10: 143-50.
- Radauer C, Breiteneder H. Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *Allergy Clin Immunol* 2006; 117: 141-7.
- Alessandri C, Zennaro D, Zaffiro A, Mari A. Molecular allergology approach to allergic diseases in the paediatric age. *Ital J Pediatr* 2009; 35: 29.
- Muñoz FJ, Delgado J, Palma J, Giménez MJ, Monteseirín FJ, Conde J. Airborne contact urticaria due to mulberry (*Morus alba*) pollen. *Contact Dermatitis* 1995; 32: 61.
- Navarro AM, Orta JC, Sánchez MC, Delgado J, Barber D, Lombardero M. Primary sensitization to *Morus alba*. *Allergy* 1997; 52: 1144-5.
- Singh BP, Singh AB, Nair PK, Gangal SV. Survey of airborne pollen and fungal spores at Dehra Dun, India. *Ann Allergy* 1987; 59: 229-34.
- Sneller MR, Hayes HD, Pinna JL. Pollen changes during five decades of urbanization in Tucson, Arizona. *Ann Allergy* 1993; 71: 519-24.
- Ciardello MA, Palazzo P, Bernardi ML, et al. Biochemical, immunological and clinical characterization of a cross-reactive nonspecific lipid transfer protein 1 from mulberry. *Allergy* 2010; 65: 597-605.
- Micheal S, Wangorsch A, Wolfheimer S. Immunoglobulin E reactivity and allergenic potency of *Morus papyrifera* (paper mulberry) pollen. *J Investig Allergol Clin Immunol* 2013; 23: 168-75.
- Choi JH, Sim JK, Oh JY, et al. An IgE-mediated allergic reaction caused by mulberry fruit. *Allergy Asthma Immunol Res* 2015; 7: 195-8.
- Aslam MS, Khalid T, Gull I, et al. Identification of major allergens of paper mulberry (*Broussonetia papyrifera*) pollens and purification of novel 40 kDa allergen protein. *Curr Allergy Clin Im* 2015; 28: 36-41.
- Çetereisi D, Karlioglu N, Gelincik A, et al. Proteomic identification of allergenic proteins of *Morus alba* L. *Asian Pac J Allergy Immunol* 2019; 37: 205-11.
- Ferrer JL, Ravel S, Robert M, Dumas R. Crystal structures of cobalamin-independent methionine synthase complexed with zinc, homocysteine, and methyltetrahydrofolate. *Biol Chem* 2004; 279: 44235-8.
- Chardin H, Mayer C, Senechal H, Tepfer M, Desvaux FX, Peltre G. Characterization of high-molecular-mass allergens in oilseed rape pollen. *Int Arch Allergy Immunol* 2001; 125: 128-34.
- Assarehzadegan MA, Sankian M, Jabbari F, Tehrani M, Falak R, Varasteh A. Identification of methionine synthase (Sal k 3), as a novel allergen of *Salsola kali* pollen. *Mol Biol Rep* 2011; 38: 65-73.
- Schmidt M, Hoffman DR. Expression systems for production of recombinant allergens. *Int Arch Allergy Immunol* 2002; 128: 264-70.
- Cui Y, Yu L, Teng F, et al. Expression of recombinant allergen, Der f 1, Der f 2 and Der f 4 using baculovirus-insect cell systems. *Arch Med Sci* 2018; 14: 1348-54.
- Valenta R, Linhart B, Swoboda I, Niederberger V. Recombinant allergens for allergen-specific immunotherapy: 10 years anniversary of immunotherapy with recombinant allergens. *Allergy* 2011; 66: 775-83.
- Jutel M, Kosowska A, Smolinska S. Allergen immunotherapy: past, present, and future. *Allergy Asthma Immunol Res* 2016; 8: 191-7.
- Ciprandi G, De Amici M, Murdaca G, et al. Serum IL-4 as a marker of immunological response to sublingual immunotherapy. *J Biol Regul Homeost Agents* 2008; 22: 117-23.
- Ciprandi G, De Amici M, Negrini S, Marseglia G, Tosca MA. TGF-beta and IL-17 serum levels and specific immunotherapy. *Int Immunopharmacol* 2009; 9: 1247-9.
- Gregan J, Rabitsch PK, Rumpf C, Novatchkova M, Schleiffer A, Nasmyth K. High-throughput knockout screen in fission yeast. *Nat Protoc* 2006; 1: 2457-64.
- Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExpASY server. In: Walker JM (ed.). *The Proteomics Protocols Handbook*. Humana Press, New Jersey 2005; 571-607.
- Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999; 292: 95-202.
- Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct Biol* 2009; 9: 51.
- Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006; 22: 195-201.
- Laskowski RA, MacArthur MW, Moss DS, Thornton MJ. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993; 26: 283-91.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 1993; 2: 1511-9.
- Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 1991; 253: 164-70.
- Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* 2008; 71: 261-77.
- Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007; 35: W407-10.
- Jespersen MC, Peters B, Nielsen M. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017; 45: W24-9.
- Yang X, Yu X. An introduction to epitope prediction methods and software. *Rev Med Virol* 2009; 19: 77-96.
- Wang DW, Ni WW, Zhou YJ, et al. Expression, purification and epitope analysis of Pla 2 allergen from *Platanus acerifolia* pollen. *Mol Med Rep* 2018; 17: 394-9.
- Kurowski M, Majkowska-Wojciechowska B, Wardzyńska A, Kowalski ML. Associations of allergic sensitization and clinical phenotypes with innate immune response genes polymorphisms are modified by house dust mite allergen exposure. *Arch Med Sci* 2011; 7: 1029-36.
- Oeo-Santos C, Mas S, Benedé S, et al. A recombinant isoform of the Ole e 7 olive pollen allergen assembled by de novo mass spectrometry retains the allergenic ability of the natural allergen. *J Proteomics* 2018; 187: 39-46.
- Paddock CD, McKerrow JH, Hansell E, Foreman KW, Hsieh I, Marshall N. Identification, cloning, and recombinant expression of procalin, a major triatomine allergen. *J Immunol* 2001; 167: 2694-9.
- Morín M, Asturias JA, Domínguez A. Expression of Alt a 1 allergen from *Alternaria alternata* in the yeast *Yarrowia lipolytica*. *FEMS Microbiol Lett* 2012; 333: 121-8.
- De Silva C, Dhanapala P, King S, Timothy Doran T, Tang M, Suphioglu C. Immunological comparison of native and

recombinant Hen's egg yolk allergen, chicken serum albumin (Gal d 5), produced in *Kluyveromyces lactis*. *Nutrients* 2018; 10: 757.

40. Sasaki M, Kumagai H, Takegawa K, Tohda H. Characterization of genome-reduced fission yeast strains. *Nucleic Acids Res* 2013; 41: 5382-99.
41. Zhang Z, Cai Z, Hou Y, et al. Enhanced sensitivity of capture IgE-ELISA based on a recombinant Der f 1/2 fusion protein for the detection of IgE antibodies targeting house dust mite allergens. *Mol Med Rep* 2019; 19: 3497-504.
42. Eberlein B. Basophil activation as marker of clinically relevant allergy and therapy outcome. *Front Immunol* 2020; 11: 1815.
43. Nair S, Kukreja N, Singh BP, Arora N. Identification of B cell epitopes of alcohol dehydrogenase allergen of *Curvularia lunata*. *PloS one* 2011; 6: e20020.