Research Journal of Agricultural Science, 58 (3), 2026; ISSN: 2668-926X;

http://doi.org/10.59463/RJAS.2026.3.05

IN SILICO PREDICTION OF B-CELL EPITOPES ON THE MAJOR RAGWEED ALLERGEN Amb a 1 AND COMPARISON WITH EXPERIMENTALLY MAPPED IgE-BINDING SITES

Ștefania-Eugenia BĂLĂȘOIU[1] (ORCID: 0009-0002-8265-4026); Alexandru MURARU[1]; Florin PRUNAR[1]; Silvia PRUNAR[1]

[1]University of Life Sciences "King Mihai I" from Timișoara, Faculty of Agriculture, Departament of Biology

Corresponding author: balasoiu.stefania-eugenia.fa@usvt.ro

Abstract. *Amb a 1, a pectate lyase, is the major allergen of common ragweed (Ambrosia artemisiifolia) pollen, and the identification of its IgE-binding regions is a prerequisite for component-resolved diagnostics and for the design of safer allergen immunotherapy. This study used a reproducible Python pipeline to predict linear B-cell epitopes on the 396-residue Amb a 1 isoform E1XUL2 through the API of the IEDB Analysis Resource, combining the BepiPred-2.0 random forest model with the classical Parker hydrophilicity scale, and compared the predictions with published experimental data on Amb a 1. At the default BepiPred-2.0 threshold of 0.5, 194 residues (49.0%) were assigned as epitope residues, which were grouped into six regions of at least four residues (4–79 residues long). Two regions were unusually long (63 and 79 residues) and are best interpreted as clusters of adjacent exposed segments rather than single epitopes. The predicted signal was only slightly higher in the IgE-dominant β chain than in the α chain (mean score 0.508 vs. 0.488). Of the four conformational IgE-binding sites previously mapped with mimotopes (K104, S110, H214 and W312 in mature-protein numbering, corresponding to residues 129, 135, 239 and 337 of the precursor), one lay inside a predicted region, two flanked a short and weak predicted segment, and one was not captured. These results show that sequence-based predictors give only partial, low-resolution agreement with experimentally mapped conformational epitopes of Amb a 1, and that they should be used to generate hypotheses to be tested experimentally rather than as a substitute for epitope mapping.*

Keywords: *Amb a 1, ragweed pollen, B-cell epitope prediction, BepiPred-2.0, allergen immunotherapy*

INTRODUCTION

Common ragweed (*Ambrosia artemisiifolia* L., Asteraceae) is an annual weed whose pollen is a major cause of seasonal allergic rhinitis, conjunctivitis and asthma in late summer and autumn. Its major allergen, Amb a 1, is recognised as such by the WHO/IUIS Allergen Nomenclature (POMÉS et al., 2018); it sensitises the large majority of ragweed-allergic patients and accounts for a large share of the ragweed-specific IgE (WOPFNER et al., 2005; JAHN-SCHMID et al., 2012).

Amb a 1 is an acidic, non-glycosylated protein of about 38 kDa, encoded by a multigene family (RAFNAR et al., 1991; WOPFNER et al., 2005). It belongs to the pectate lyase family and folds into a parallel β-helix. Its enzymatic function, the cleavage of α-1,4-linked polygalacturonate in plant cell walls, is unrelated to its allergenicity, which depends on which surface regions of the protein are recognised by the human immune system. During purification, natural Amb a 1 is cleaved into a 26 kDa C-terminal α chain and a 12 kDa N-terminal β chain that remain non-covalently associated (WOPFNER et al., 2005). The two chains differ immunologically: the β chain keeps most of the IgE reactivity, whereas the α chain binds IgE more weakly but retains T-cell reactivity, which is why it has been proposed as a naturally occurring hypoallergen (WOPFNER et al., 2009).

Two classes of epitopes shape the allergic response. B-cell epitopes are the surface regions bound by IgE; cross-linking of receptor-bound IgE on mast cells and basophils triggers the release of histamine and other mediators. T-cell epitopes are peptides generated by antigen-presenting cells and displayed on MHC class II molecules to CD4+ T helper cells, which drive Th2 polarisation and class switching to IgE. B-cell epitopes can be linear, formed by a contiguous stretch of residues, or conformational, formed by residues that are far apart in the sequence but close together in the folded protein; sequence-based predictors can only be expected to capture the first kind well (JESPERSEN et al., 2017).

Knowledge of IgE epitopes is a prerequisite for the rational design of molecular approaches to the diagnosis and immunotherapy of allergy, for example component-resolved diagnostics, which test sensitisation to individual allergen molecules instead of crude extracts, and immunotherapy with hypoallergenic variants or peptides that lower the risk of IgE-mediated side effects (ZAHIROVIĆ et al., 2019). For Amb a 1, however, experimental information on IgE-binding sites is still very limited (ZAHIROVIĆ et al., 2019).

Experimental epitope mapping, by crystallography of allergen–antibody complexes, overlapping peptides or phage display, is slow and costly, and computational predictors are therefore widely used as a first screen. BepiPred-2.0 is a random forest model trained on epitopes derived from crystal structures of antibody–antigen complexes, which returns a score for each residue from the amino acid sequence alone (JESPERSEN et al., 2017). Older propensity scales, such as the Parker hydrophilicity scale, rely on the observation that antigenic regions tend to be hydrophilic and exposed on the protein surface (PARKER et al., 1986). Both methods are available through the Immune Epitope Database (IEDB) Analysis Resource (VITA et al., 2019). For Amb a 1, data on chain-level IgE reactivity (WOPFNER et al., 2009) and on four mimotope-derived IgE-binding sites (ZAHIROVIĆ et al., 2019) make it possible to benchmark such predictions against experiment.

The aims of this study were: (i) to predict linear B-cell epitopes on the Amb a 1 sequence with BepiPred-2.0 and, as a simple baseline, with the Parker hydrophilicity scale, using a reproducible Python pipeline that queries the IEDB API; (ii) to define and characterise discrete candidate regions; (iii) to compare these regions with the published chain-level IgE reactivity and with the four experimentally mapped IgE-binding sites, taking into account the different residue numbering conventions used in the literature; and (iv) to locate the regions and the experimental sites on a three-dimensional model of Amb a 1 – keep only if the ChimeraX step is completed.

MATERIAL AND METHODS

Protein sequence and residue numbering

The Amb a 1 sequence was retrieved from UniProt (UniProt Consortium, 2023; accession E1XUL2, pectate lyase from *Ambrosia artemisiifolia*). It has 396 residues. Residues 1–25 form the N-terminal signal peptide, which is absent from the mature protein (residues 26–396). Compared with UniProt entry P27759 (Amb a 1.1), E1XUL2 differs at a single position (Asp92 instead of Glu). In this paper, all residues are numbered according to the full-length precursor (Met1 = 1). The IgE-binding sites reported by Zahirović et al. (2019) as K104, S110, H214 and W312 are numbered in the mature protein; adding 25 gives positions 129, 135, 239 and 337, which carry exactly Lys, Ser, His and Trp in the E1XUL2 sequence, so that the offset was confirmed. The α and β chains were defined as residues 181–396 and approximately 26–180, respectively (WOPFNER et al., 2009).

B-cell epitope prediction

The sequence was submitted to the B-cell epitope prediction tools of the IEDB Analysis Resource (VITA et al., 2019) through its REST API (tools-cluster-interface.iedb.org/tools_api/bcell/), using a Python 3 script (requests, pandas and matplotlib) that is available from the authors. Two methods were run on the full-length precursor sequence. (i) BepiPred-2.0 (JESPERSEN et al., 2017), a random forest model that returns a score for each residue; residues with a score of at least 0.5, the default threshold of the tool, were assigned as epitope residues (E). (ii) The Parker hydrophilicity scale (PARKER et al., 1986), computed with a sliding window of 7 residues. The Parker profile was used only as a simple baseline for visual comparison and no threshold was applied to it. The API was queried in August 2026.The output tables index residues from zero; positions were converted to residue numbers by adding 1, and the residue letters in the output were checked against the input sequence. The signal peptide was left in the input sequence, so the scores of residues 1–25 and of the first residues of the mature protein may be influenced by it.

Definition of candidate regions

Consecutive residues assigned E were merged into regions. Regions shorter than four residues were discarded to limit noise (one segment of three residues, 272–274, was removed). For every region we report the length, the mean and maximum BepiPred-2.0 score, the position of the maximum, the amino acid sequence and the chain in which it lies. The fraction of residues assigned E and the mean score were also calculated separately for the signal peptide (1–25), the β chain and the α chain.

Comparison with experimental data

The predictions were compared with two published data sets. At chain level, Wopfner et al. (2009) reported that the β chain is strongly recognised by IgE from most patients, whereas the α chain binds IgE weakly. At residue level, Zahirović et al. (2019) selected mimotopes from phage-displayed random peptide libraries using Amb a 1-specific IgG, mapped them with the EpiSearch algorithm on a homology model of Amb a 1, and proposed four potential epitopic sites centred at K104, S110, H214 and W312; peptides matching these sites bound IgE from pooled sera of ragweed-allergic patients. For each site we recorded the BepiPred-2.0 score at the central residue, whether it was assigned E, and the distance in sequence to the nearest edge of a predicted region (0 if inside). Because these sites are conformational patches defined on a model, they were treated as approximate locations and not as exact linear epitopes. As a reference for chance agreement, we calculated the share of the sequence covered by the predicted regions.

Three-dimensional visualisation

The AlphaFold model of E1XUL2 (AF-E1XUL2-F1; Jumper et al., 2021; Varadi et al., 2022) was retrieved and visualised in UCSF ChimeraX (Pettersen et al., 2021; version X.X). Residues 1–25 (signal peptide) were hidden, and the predicted regions and the four experimental residues (129, 135, 239 and 337) were coloured on the molecular surface. The solvent-accessible surface area of each region was calculated with the measure sasa command.

RESULTS AND DISCUSSION

Predicted B-cell epitope regions

At the 0.5 threshold, BepiPred-2.0 assigned 194 of the 396 residues (49.0%) as epitope residues (189 of 371 residues, 50.9%, when the signal peptide is excluded). Five of the 25 signal-peptide residues were assigned E, all at its C-terminal end (residues 21–25), where the profile rises into the first region. After the removal of segments shorter than four residues, six candidate regions remained (Table 1), covering 191 residues (48.2% of the sequence).

Research Journal of Agricultural Science, 58 (3), 2026; ISSN: 2668-926X;
http://doi.org/10.59463/RJAS.2026.3.05

Table 1

Candidate B-cell epitope regions predicted on Amb a 1 by BepiPred-2.0 (threshold 0.5)

Region	Residues	Length (aa)	Chain	Mean score	Max. score (residue)	Sequence
1	21–83	63	SP/β	0.57	0.65 (41)	QPVRSAEDLQ…FGKGTVGGKD
2	92–104	13	β	0.62	0.70 (97)	DLDDDVANPKEGT
3	131–134	4	β	0.51	0.52 (132)	MVVN
4	171–194	24	β/α	0.62	0.71 (187)	VKVNPGGLIKSNDGPAAPRAGSDG
5	249–256	8	α	0.57	0.62 (252)	DENIEDRG
6	314–392	79	α	0.58	0.67 (364)	APDERSKKNV…GESALSLTSSAGVLSCQP

Residues are numbered in the full-length precursor (Met1 = 1). SP, signal peptide (residues 1–25); β, residues ≈26–180; α, residues 181–396. Scores are BepiPred-2.0 scores averaged over the region (mean) or the highest value found in it (max.). Long sequences are abbreviated.

Regions 1 (63 residues) and 6 (79 residues) are much longer than a typical linear epitope, which is usually a short peptide. They are best read as clusters of adjacent exposed segments in which the score stays above the threshold, not as single epitopes. Region 3, in contrast, has only four residues and barely exceeds the threshold (maximum score 0.52), so it represents a weak signal.

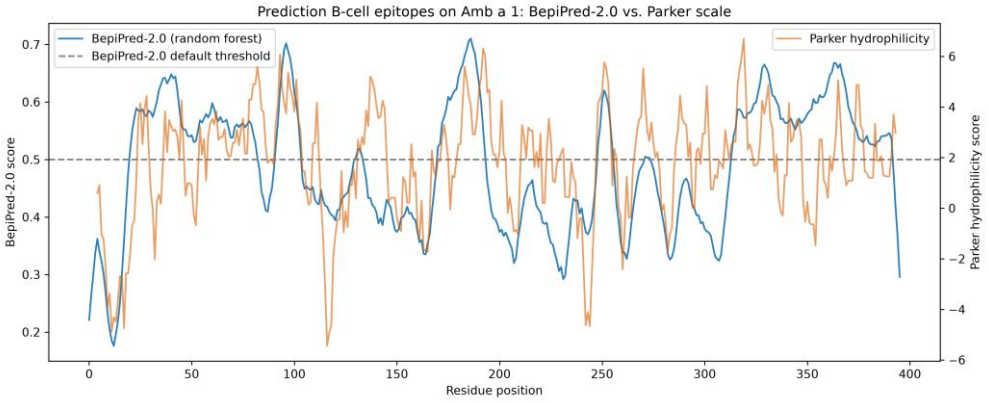


Figure 1. BepiPred-2.0 scores (blue line, left axis) and Parker hydrophilicity scores computed with a 7-residue window (orange line, right axis) along the Amb a 1 sequence. The dashed line indicates the BepiPred-2.0 threshold (0.5). The residue axis starts at 0 in the plot (position = residue number − 1).

The two profiles in Figure 1 follow similar broad trends in some segments (for example around residues 20–45, 170–190 and 320–370) but disagree elsewhere: the Parker scale shows sharp minima around residues 115–120 and 240–245, where the BepiPred-2.0 score declines more gradually, and several Parker peaks between residues 260 and 300 lie in a segment where the BepiPred-2.0 score is at or below the threshold. The comparison is qualitative; no correlation coefficient was calculated.

At chain level, 85 of the 155 residues of the β chain (54.8%) were assigned E, with a mean score of 0.508, against 104 of the 216 residues of the α chain (48.1%), with a mean score of 0.488. This difference is small, and the profile alone would not identify the β chain as the IgE-

dominant one. This is expected, because experimental IgE reactivity (WOPFNEr et al., 2009) reflects the folded protein, whereas BepiPred-2.0 uses only the amino acid sequence.

Location of the predicted regions in the three-dimensional structure

Visualisation of the AlphaFold model in UCSF ChimeraX confirmed that the major predicted regions (such as Regions 1, 2, 4, and 6) are extensively exposed on the surface of the parallel β-helix structure. Residue Trp337 (W312) is located directly within the large surface patch corresponding to Region 6 (pink surface). Residues Lys129 (K104) and Ser135 (S110) are positioned close to each other adjacent to Region 3 (blue), whereas His239 (H214) lies on a distinct exposed loop segment.

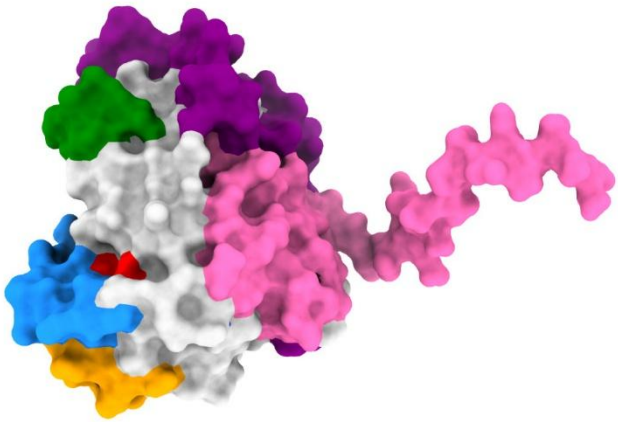


Figure 2. AlphaFold model of mature Amb a 1 (isoform E1XUL2) displayed as a molecular surface. Predicted B-cell epitope regions are highlighted in distinct colours (Region 1 in magenta, Region 2 in green, Region 3 in blue, Region 4 in orange, Region 5 in red, and Region 6 in pink). The four experimentally mapped IgE-binding residues, Lys129 (K104), Ser135 (S110), His239 (H214), and Trp337 (W312), are indicated on the surface.

Comparison with experimentally mapped IgE-binding sites

Table 2 compares the predicted regions with the four IgE-binding sites proposed by ZAHIROVIĆ et al. (2019). Because these authors numbered residues in the mature protein, the positions had to be shifted by 25 residues before the comparison; without this correction, the sites would be assigned to the wrong residues (position 104 of the precursor, for example, is Thr and not Lys) and the comparison would be misleading.

Table 2

Comparison of predicted regions with the IgE-binding sites mapped by ZAHIROVIĆ et al. (2019)

Experimental site (mature numbering)	Position in precursor	Chain	BepiPred-2.0 score	Nearest predicted region	Distance (residues)
K104	Lys129	β	0.47 (not E)	Region 3 (131–134)	2
S110	Ser135	β	0.49 (not E)	Region 3 (131–134)	1
H214	His239	α	0.43 (not E)	Region 5 (249–256)	10
W312	Trp337	α	0.59 (E)	Region 6 (314–392), inside	0

Site numbering as in ZAHIROVIĆ et al. (2019), i.e. mature protein; precursor position = mature position + 25. Distance is the number of residues between the site and the nearest edge of a predicted region. E, residue assigned as epitope residue.

Trp337 (W312) lies inside Region 6, with a BepiPred-2.0 score of 0.59. However, Region 6 covers 79 residues and the predicted regions together cover 48% of the sequence, so that a site is expected to fall inside a predicted region about every second time by chance. This agreement is therefore not, by itself, evidence of predictive power. Lys129 (K104) and Ser135 (S110) are not assigned E (scores 0.47 and 0.49), but they flank Region 3 (residues 131–134), the only predicted segment between residues 105 and 170. The K104/S110 patch therefore coincides with a local and weak maximum of the BepiPred-2.0 profile, a modest agreement that rests on a four-residue segment just above the threshold. Region 2 (residues 92–104), which stands out in the profile, ends 25 and 31 residues before these two sites in the sequence; whether it is spatially close to them can only be judged on the three-dimensional structure. His239 (H214) has a low score (0.43) and was not captured; the nearest predicted segment, Region 5, is 10 residues away. This is the behaviour expected from a sequence-based method for a conformational site, in which residues are brought together only by folding, and the mimotope peptides themselves aligned only weakly with the primary sequence (ZAHIROVIĆ et al., 2019).

Two of the four sites (H214 and W312) lie in the α chain. This indicates that the α chain does contain IgE-binding sites, which fits the weaker, but not absent, IgE binding of this chain with patient sera (WOPFNER et al., 2009), and it shows why Regions 5 and 6 should not be dismissed as false positives only because the α chain is less IgE-reactive overall.

Several limitations must be kept in mind. First, BepiPred-2.0 was run on the precursor sequence, including the signal peptide. Second, the experimental reference sites were derived from mimotopes selected with Amb a 1-specific IgG and mapped on a homology model, with IgE binding tested on pooled sera; they are therefore putative sites and not epitopes defined by a crystal structure of an allergen–IgE complex. Third, no statistical test was applied, and only one isoform was analysed. In practice, the predictions can be used to prioritise segments for peptide synthesis and IgE-binding tests, particularly those that are both predicted and exposed on the surface in the three-dimensional model, but the present data do not justify the choice of any peptide as a diagnostic or therapeutic target.

CONCLUSIONS

A reproducible pipeline combining BepiPred-2.0 and the Parker hydrophilicity scale predicted six candidate B-cell regions on Amb a 1, which together cover about half of the sequence. Agreement with experimentally mapped IgE-binding sites was partial: W312 fell inside a long predicted region, K104 and S110 flanked a short and weak one, and H214 was missed. The difference between the β and α chains in the predicted signal was small. A correct conversion between the numbering of the precursor and that of the mature protein, an offset of 25 residues, was essential for the comparison. Sequence-based B-cell predictors therefore give

low-resolution and only partly consistent information on the epitope landscape of Amb a 1, and should be combined with structure-based tools such as ElliPro (Ponomarenko et al., 2008) applied to a three-dimensional model, with T-cell epitope prediction using NetMHCIIpan-4.0 (Reynisson et al., 2020), and with experimental testing of selected peptides.

BIBLIOGRAPHY

JAHN-SCHMID B., HAUSER M., WOPFNER N., BRIZA P., BERGER U.E., ASERO R., EBNER C., FERREIRA F., BOHLE B., 2012 – Humoral and cellular cross-reactivity between Amb a 1, the major ragweed pollen allergen, and its mugwort homolog Art v 6. Journal of Immunology, 188(3): 1559–1567.

JESPERSEN M.C., PETERS B., NIELSEN M., MARCATILI P., 2017 – BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Research, 45(W1): W24–W29.

JUMPER J., EVANS R., PRITZEL A., ET AL., 2021 – Highly accurate protein structure prediction with AlphaFold. Nature, 596: 583–589.

PARKER J.M.R., GUO D., HODGES R.S., 1986 – New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry, 25(19): 5425–5432.

PETTERSEN E.F., GODDARD T.D., HUANG C.C., ET AL., 2021 – UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Science, 30(1): 70–82.

POMÉS A., DAVIES J.M., GADERMAIER G., ET AL., 2018 – WHO/IUIS Allergen Nomenclature: providing a common language. Molecular Immunology, 100: 3–13.

PONOMARENKO J., BUI H.H., LI W., ET AL., 2008 – ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics, 9: 514.

RAFNAR T., GRIFFITH I.J., KUO M.C., BOND J.F., ROGERS B.L., KLAPPER D.G., 1991 – Cloning of Amb a I (antigen E), the major allergen family of short ragweed pollen. Journal of Biological Chemistry, 266(2): 1229–1236.

REYNISSON B., ALVAREZ B., PAUL S., PETERS B., NIELSEN M., 2020 – NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Research, 48(W1): W449–W454.

UNIPROT CONSORTIUM, 2023 – UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research, 51(D1): D523–D531.

VARADI M., ANYANGO S., DESHPANDE M., ET AL., 2022 – AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research, 50(D1): D439–D444.

VITA R., MAHAJAN S., OVERTON J.A., ET AL., 2019 – The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Research, 47(D1): D339–D343.

WOPFNER N., GADERMAIER G., EGGER M., ASERO R., EBNER C., JAHN-SCHMID B., FERREIRA F., 2005 – The spectrum of allergens in ragweed and mugwort pollen. International Archives of Allergy and Immunology, 138(4): 337–346.

WOPFNER N., JAHN-SCHMID B., SCHMIDT G., CHRIST T., HUBINGER G., BRIZA P., RADAUER C., FERREIRA F., BOHLE B., 2009 – The alpha and beta subchain of Amb a 1, the major ragweed-pollen allergen, show divergent reactivity at the IgE and T-cell level. Molecular Immunology, 46(11): 2090–2097.

ZAHIROVIĆ A., ŠTRUKELJ B., KOROŠEC P., LUNDER M., 2019 – Epitope mapping of major ragweed allergen Amb a 1. Acta Chimica Slovenica, 66(1): 37–44. DOI: 10.17344/acsi.2018.4516.