Discontinuous epitope prediction based on mimotope analysis

Violaine Moreau, Claude Granier, Sylvie Villard, Daniel Laune and Franck Molina*

CNRS UMR 5160, Centre de Pharmacologie et Biotechnologie pour la Santé, Faculté de Pharmacie, 15 Avenue Charles Flahault, BP 14491, 34093 Montpellier Cedex 5, France

Received on October 21, 2005; revised on January 10, 2006; accepted on January 19, 2006

Advance Access publication January 24, 2006

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Phage display is a widespread technique used to obtain peptide mimotopes selected by binding to a given monoclonal antibody in a similar way as the native epitope. However, the localization of the interaction site mimicked by the mimotopes on the surface of the antigen is not always a straightforward task. MIMOP is a computational tool developed with the aim of helping experimentalists to analyze a set of mimotope sequences and guide them in the identification of the mimicked region.

Results: To predict potential epitopic regions, MIMOP integrates two different approaches combining two- and three-dimensional analyses: MimAlign starts from degenerated alignment analyses, and MimCons is based on consensus identification. The relevance and usefulness of the tool are illustrated by four use cases corresponding to real-life situations.

Availability: upon request.

Contact: franck.molina@cpbs.univ-montp1.fr

1 INTRODUCTION

During antibody–antigen (Ab–Ag) interaction, the paratope of the Ab binds the epitope of the Ag. A protein epitope can either be continuous or discontinuous. While a continuous (also called sequential or linear) epitope is a sequential fragment from the protein sequence, a discontinuous (also called conformational) epitope is composed of several fragments scattered along the protein sequence and brought together in spatial proximity when the protein is folded (Van Regenmortel, 2001). Most epitopes are discontinuous although they are often composed of small continuous elements of the sequence. The Ab interaction with such an element (usually, a synthetic peptide) is often weaker than with the native protein because its conformation is either different from that of the protein or other elements of the discontinuous epitope are energetically missing. Specificity is the hallmark of Ab–Ag interaction. However, specificity is not absolute (Van Regenmortel, 1998). Cross-reactivity refers to the binding of the Ab to an analog of the native Ag against which the Ab has been raised. In this case, both epitopes comprise the same (or analogous) critical residues. Polyspecificity is the ability of an Ab to bind to different molecules using different critical residues. In this case, the epitope of the native Ag is not homologous to the epitope of the other molecules (Keitel et al., 1997). In 1986, Geysen’s group coined the term ‘mimotope’, literally defined as a mimic of an epitope (Geysen et al., 1986). Mimotopes are peptides selected from random libraries for their capacity to bind to an Ab directed against a given Ag. Thus, mimotopes and Ag are both recognized by the same Ab paratope. By definition, mimotopes are expected to imitate part of the ‘true’ epitope, although no clear homology may exist between the sequence of such mimotopes and the epitope of the native Ag. The mimicry is supposed to rely on similarities in physicochemical properties and similar spatial organization.

Mimotopes can have many applications in important axes of research such as, for example, diabetes (Samli et al., 2005), AIDS (Deng et al., 2005) or cancer (Hardy and Raiter, 2005). They can lead to pharmacological target identification, which is interesting for drug design or guided docking. Mimotopes are also very useful for mapping the epitope of the Ag recognized by an Ab. In this way, they can also lead to protein engineering, vaccine design, identification of protein function, etc. However, the precise identification of the ‘true’ epitope mimicked by a mimotope is a difficult task because most epitopes are discontinuous and, although they are functionally equivalent, the epitope and the mimotope often do not share sequence similarities. Pizzi et al. (1995) were the first to combine computational methods and experimental results to assign epitopes from mimotopes. Since then, a few fully automated procedures have been developed, either based on sequence information (including reversals) (Mumey et al., 2003), three-dimensional (3D) information (Schreiber et al., 2005) or on a combination of the two (Enshell-Seijffers et al., 2003; Halperin et al., 2003). However, although they appear to be efficient for an individual analysis, some of these algorithms were not characterized for their capacity to handle large datasets or appropriate data (i.e. mimotopes selected with a monoclonal and not a polyclonal Ab). None of these methods propose an analysis of the mimotope sequences in order to characterize them. For example, consensus patterns displayed by mimotopes need to be taken into account. Such homologies are indicative of the importance of certain residues for molecular recognition and, thus, should advantageously be taken into consideration for prediction. We thus sought to develop a prediction strategy based on the identification of key residues from sets of mimotope sequences and on the matching of those key residues with the accessible amino acids on the Ag surface.

*To whom correspondence should be addressed.
In the present study, we describe MIMOP, a new computational tool that provides a 3D epitopic region prediction from the information content of mimotope peptide sequences. MIMOP is able to predict either continuous or discontinuous epitopes. In addition, this tool provides an environment for mimotope characterization which integrates two main approaches, MimAlign and MimCons, that deliver to the user mimotope analysis results (mimotope clustering, multiple alignments, consensus patterns, key residues, etc.). These results are displayed via a web accessible, user-friendly interface, allowing, among other things, interactive visualization of the predictions directly on the 3D structure of the Ag. To validate the efficacy of MIMOP, we report the results of discontinuous epitope prediction on several Ag models and compare the predictions with the available epitope data.

2 METHODS

2.1 MIMOP description

MIMOP includes two approaches (Fig. 1). MimAlign predicts potential epitopic regions (PER) from mimotope sequences and from the 3D structure of the Ag, if available. MimCons predicts PER from mimotope sequences but requires the 3D structure of the Ag.

MimAlign combines results from four multiple sequence alignments of the Ag and mimotope sequences in a combined alignment. For each position of this combined alignment a frequency and a score are calculated. Convergent positions are then selected and clustered based on their topology. The clusters obtained are considered as PER, then scored and ranked.

MimCons evaluates the similarity of the mimotope sequences and clusters them accordingly. ‘Consensus patterns’ are identified from mimotope sequences of each cluster. The accessible surface of the Ag is scanned to look for all possible exposed consensus patterns. Spatially neighboring amino acids are identified and constitute PER. The different PER are further scored and ranked.

MimAlign can predict continuous epitopes, whereas MimCons cannot (since surface accessible amino acids are scanned for consensus patterns). If the selected convergent positions are displayed in a sequential order in the protein sequence, the fragment is proposed as a PER. MimCons allows separate analysis of different identified clusters. This is useful because mimotopes might mimic different parts of a discontinuous epitope, i.e. key residues of a discontinuous epitope can be present in several clusters of mimotopes from the same dataset.

MimAlign and MimCons can be run independently or with their results combined. Each of the three approaches proposes the first quartile PER to be the most probable regions on the Ag mimicked by a given set of mimotopes.

2.2 Combination of multiple sequence alignments

Multiple sequence alignments of Ag and mimotopes are performed by four different algorithms: DIALIGN (Morgenstern et al., 1998), T-Coffee (Notredame et al., 2000), Multalin (Corpet, 1988) and DCA (Stoye et al., 1997). They have been selected among other softwares for their ability to align sequences of very low similarity and different sizes. The parameters of Multalin and DCA have been optimized to perform degenerated sequence alignments required to align sets of weakly similar mimotope sequences.

To display a combined alignment of all the multiple sequence alignments provided by the four algorithms, the Ag sequence is considered as a reference. The Ag sequence is represented only once and the four alignments follow one after the other (displayed in different colors on the computer screen). If a shift appears in one alignment owing to the insertion of a gap in the aligned Ag sequence, the character ‘@’ is inserted in place of the gap in the other alignments to keep the four multiple alignments in phase.

2.3 Calculation of frequency and score

For each position of the combined alignment, a frequency and a score are calculated. The frequency is defined as the percentage of amino acids from the aligned mimotope sequences occurring at the observed position in the Ag sequence:

\[ f_i = \frac{n_i}{N} \times 100\% \]

where \( n_i \) is the number of amino acids from the mimotopes aligned at Ag position \( i \) of the alignment and \( N \) the number of mimotope sequences aligned in the combined alignment (equal to the number of mimotopes multiplied by the number of sequence alignment softwares).

The score (expressed as a percentage) at a given position is defined by the relative sum of the similarities between the amino acid in the aligned mimotope sequences and the corresponding amino acid in the Ag sequence:

\[ s_i = \frac{\sum_{j=0}^{s_{\text{max}}} s_j}{s_{\text{max}} \times n_i} \times 100\% \]

where \( j = n_i \) is the number of amino acids from mimotopes aligned at position \( i \) of the alignment, \( s_j \) is the similarity score of the amino acid from the mimotope compared with the one from the Ag sequence and \( s_{\text{max}} \) is the highest possible similarity score between two amino acids. \( s_j \) and \( s_{\text{max}} \) depend on the matrix of similarity used. Here we used our own
matrix of similarity based on the physicochemical properties of the amino acids (unpublished data).

### 2.4 Selection and clustering of convergent positions

Positions in the protein sequence of the combined alignment are selected if their frequency, or their score are at least equal to 25% (percentage shown to give the best predictions) and if they are found to be accessible (surface accessibility of amino acids is determined using DSSP (Kabsch and Sander, 1983) in the corresponding 3D structure of the Ag. These positions are named ‘convergent positions’. A clustering is then produced using Phylip package v3.6 (Felsenstein, 1989) after a Euclidian distance matrix calculation based on the 3D coordinates of the convergent positions. In this way, the clusters obtained are considered as PER on the solvent accessible surface of the Ag.

### 2.5 Clustering based on mimotope similarity

Two types of methods were developed to evaluate the similarity between mimotope sequences within a given set of mimotopes (derived from a given experiment). The first method is based on sequence alignments performed either by ClustalW (Thompson et al., 1994) or Multalin (Corpet, 1988). The alignment is then translated into a distance matrix. The second method is the calculation of a distance matrix using Levenshtein distances (Levenshtein, 1966). This distance can be calculated from raw or aligned mimotope sequences.

The distance matrix obtained from either of the two methods is then used to cluster the sequences with Phylip package v3.6 (Felsenstein, 1989) using a threshold value (that allows identification of the clusters) shown to give the best predictions.

### 2.6 Identification of consensus patterns

Consensus patterns from each cluster of mimotope sequences (at least three sequences) are calculated using the algorithm PRATT v2.1 (Jonassen et al., 1995). A consensus pattern is expressed as a PROSITE pattern (Hulo et al., 2004) and identifies the most common residues present in a set of sequences (Fig. 2A). The consensus patterns can be calculated from either raw or aligned mimotope sequences (ClustalW). ‘Degenerated consensus patterns’ are also determined on the basis of those found by PRATT. To this end, amino acids of the consensus pattern are replaced by amino acids with similar physicochemical properties, if they exist. For example, R can be replaced by K and vice versa, or W can be replaced by Y or F. In the same way, ‘global consensus’ patterns are generated which combine the consensus pattern found by PRATT and the degenerated consensus patterns.

### 2.7 Selection of neighboring residues by surface scan

The protein surface is scanned for all the consensus patterns. All amino acids from a consensus pattern [X is not taken into account (Fig. 2A)] are sought among the solvent accessible amino acids. Thus, to each position of the consensus pattern there corresponds a list of accessible amino acids on the surface of the Ag. From these lists, all the possible combinations in accordance with the consensus pattern are calculated. Spatially neighboring amino acids form a combination (a consensus pattern found at the surface of the protein) where all the composing amino acids are inside a sphere of 10 Å from the gravity point (Fig. 2B). The distance of 10 Å is based on an analysis of the Ag. The distance of 10 Å is based on an analysis of Chakrabarti and Janin (2002), where they dissected protein–protein recognition sites and found that the average radius of an interface area is 8.0 ± 1.7 Å. By rounding this value to 10 Å, potential peptide epitopes are not missed. As a result, the selected combinations of amino acids accessible on the surface of the Ag are considered to be PER.

### 2.8 PER combination

The two best quartile PER selected from MimAlign and from MimCons are compared. When amino acids are found in common in two PER, a new PER is calculated comprising surface accessible residues inside the sphere with a 10 Å radius.

---

**Fig. 2.** MimCons illustration. (A) Conserved consensus patterns are calculated using the algorithm PRATT v2.1 from mimotope sequences. Several consensus patterns can be found for a set of mimotopes. (B) A consensus pattern found on the accessible surface of the protein is selected if all the residues (bold and underlined in the example) are inside a sphere with a radius of 10 Å from the gravity point. A PER includes all the solvent accessible amino acids inside the sphere.

### 2.9 PER ranking

To calculate a PER score, the amino acids of a PER are compared twice with the amino acids of the mimotope sequences. The first time, the score increases by 1 when the two amino acids compared are identical. The second time, the rest of the mimotope amino acids are compared with the rest of the amino acids of the PER: the score increases by 0.5 when the two residues are similar. Groups of similar residues are positively charged (R, K, H), negatively charged (D, E), hydrophobic (W, Y, F, L, I, V, M), amidated (Q, N), cysteine similar (S, C), small (A, G) and hydroxylated (S and T). The PER are then ranked.

### 3 RESULTS

MIMOP is a computational method developed with the aim of helping experimentalists to analyze a set of mimotope sequences and guide them in the identification of the mimicked epitopic region.

#### 3.1 MIMOP principle and specifications

MIMOP (Fig. 1) is a bioinformatics tool set which provides the user with mimotope analysis results interactively displayed through a user-friendly interface. The main output is a ranked list of PER. Two approaches have been developed (MimAlign and MimCons) because sets of mimotopes exhibit either similarity, although sometimes very weak, with the Ag sequence or contain a small number of residues always present in the same position in the mimotope sequences. MimAlign is best suited to handle the first situation and MimCons more adapted to the second. Depending on the experimental conditions and the biological system studied, both situations may coexist within a heterogeneous set of mimotopes. To tackle such a case, MIMOP is able to combine the results of the two approaches, leading to convergent predictions.

MimAlign combines multiple sequence alignments and selects convergent positions by scoring them according to their similarity with the Ag sequence. The convergent positions are then clustered...
on the accessible surface of the Ag in order to identify PER. Thus, MimAlign essentially analyses the global similarity of the sequences of mimotopes with the sequence of the antigen. The second approach, MimCons, determines key residues corresponding to consensus residues of the mimotope sequences grouped according to their similarity. These key residues are found on the accessible surface of the Ag and PER are determined by including neighboring amino acids. Thus, MimCons select residues at the antigen surface that are identical or similar to residues conserved within mimotope sequences. Because MimAlign and MimCons are based on different predictive postulates, they may yield different results. If they do so, the results of MimAlign and MimCons can usefully be combined. The PER found by both approaches are compared, and the common residues are used to identify new PER. In all three approaches proposed (MimAlign, MimCons and combined) the PER are scored by comparing their amino acid composition with that of the mimotope sequences, and the top-ranked PER is proposed as the region mimicked by the mimotope set.

To evaluate or to use such a tool, it is recommended to work with a set of mimotopes selected against a unique mAb to be sure that there is only a single targeted epitope (by opposition with mimotopes selected by polyclonal antibodies from which one can expect many epitopes targeted on the Ag surface). All the following cases of the use of MIMOP concern mimotopes selected by an mAb. Furthermore, they show increasing difficulties from a case with available data and a well-known epitope to a case with partially available data and an unknown epitope.

### 3.2 MIMOP use cases

#### 3.2.1 Validation case: the epitope is known from X-ray crystallography

As a pure validation test, we compared the epitope prediction provided by MIMOP with the ‘true’ epitope defined by X-ray crystallography of an Ab–Ag complex. We thus used a set of 27 mimotopes selected against mAb Bo2C11 targeting the C2 domain of factor VIII (FVIII) (Villard et al., 2003), the 3D structure of the C2 domain complexed with Bo2C11 Fab (Spiegel et al., 2001) being available (PDB code: 1IQD). FVIII is a multidomain protein acting as a cofactor in the generation of thrombin by the intrinsic pathway tenase complex. Dysfunction or deficiency in FVIII results in a bleeding disorder called hemophilia A. Bo2C11 is a human immunoglobulin (IgG4κ) derived from an inhibitor-positive patient with hemophilia A (Jacquemin et al., 1998). The paratope of Bo2C11 interacts with the C2 domain of FVIII through a conformational epitope composed of 15 amino acids (Fig. 3). A total of 27 mimotopes have been previously selected using Bo2C11 and phage display (Villard et al., 2003).

We first performed a MimAlign study on the 27 mimotope sequences thanks to the four multiple sequence alignment algorithms proposed. Several residues (N2198, M2199, R2215, R2220, L2252, H2315) of the best cluster identified by MimAlign as a PER were found at the interface of the Bo2C11 mAb-C2 domain complex (Fig. 3).

To tackle the fuzziness of degenerated mimicking information found in mimotope sequences, we used the MIMOP tool which combines both the MimAlign and MimCons approaches. The results of such a combined analysis are presented in Fig. 3. The best ranked PER found was to be located at the interface of the mAb-Ag complex and comprised residues identified as contact residues by X-ray crystallography. Eight amino acids (F2196, T2197, N2198, M2199, F2200, R2215, R2220 and Q2222) out of the 11 predicted were
observed in this epitope, which is composed of 15 residues. The other three (Y2195, T2202 and K2207) are very close to the first eight since they are all inside a sphere with a 10 Å radius. Thus, in this validation study, MIMOP was able to propose an epitope location (a PER) which encloses eight residues of the ‘true’ epitope defined by X-ray crystallography.

3.2.2 Case for which structural and functional information on the epitope are available

We used a 57-mimotope set selected against mAb BCF2 (Gazarian et al., 2003), targeting the scorpion toxin Cn2 for which an NMR structure is available (Pintar et al., 1999). The neurotoxin Cn2 represents the most abundant and one of the most potent toxins in the venom of the scorpion Centruroides noxius. Scorpion toxins are small (4–8 kDa), globular proteins evolutionarily adapted for recognition and blockage of K-, Na- and Ca-membrane channels (Possani et al., 1999a,b). mAb BCF2 is able to neutralize the toxic effects of Cn2 (Licea et al., 1996). Recently, parts of the primary structure of Cn2 involved in the epitope function were identified using continuous and discontinuous synthetic peptides derived from the toxin sequence, and a 3D model of the complex Cn2-BCF2 was constructed (Selisko et al., 1999). From these studies, 17 amino acids were identified as being involved in the epitope (Fig. 4).

MIMOP predictions were conducted using the sequence of Cn2, its 3D structure defined by NMR (PDB code: 1CN2) and the set of 57 mimotope sequences selected for their capacity to bind to the mAb BCF2 (Gazarian et al., 2003). A MimAlign analysis was performed to produce a combined alignment. The best PER (Fig. 4) thus identified was composed of 15 residues with 10 residues (K8, N9, T10, C12, Y14, I56, P61, K63, C65, S66) belonging to the experimentally identified epitope (Gazarian et al., 2003) and the 5 others very close in the 3D structure of the toxin.

A MimCons study was simultaneously performed using the same dataset. To cluster the mimotopes, the similarity was evaluated using the Levenshtein method on sequences aligned with Multalin. The similarity of the mimotopes was found to be low since 22 clusters were identified. Consensus patterns were determined for each cluster comprising more than two mimotopes (12 clusters were composed of only one mimotope and four clusters of two mimotopes). From these, 13 consensus patterns could be identified. The best scored PER identified by topological research on the accessible surface of the Ag was composed of 16 amino acids (Fig. 4). Among them, nine (K8, N9, T10, C12, I56, P61, N62, K63 and C65) were identified by Gazarian et al. (2003) to be part of the epitope.

MimAlign and MimCons approaches were then combined to analyze the same dataset. The best PER is displayed on the 3D structure of the Cn2 toxin (Fig. 4). Out of the 16 predicted amino acids 10 (K8, N9, T10, C12, I56, P61, N62, K63, C65 and S66) were indeed in the published epitope. Therefore, the predictions of MimAlign, MimCons as well as the combined approach of MIMOP were in good agreement with the results found by Gazarian et al. (2003).

3.2.3 Case where no information on the epitope is available

This case is representative of a quite common situation in which no precise 3D information is available in databases, and no information on the epitope is available.

The dermonecrotic factor (DNF) from the venom of the spider Loxosceles intermedia is a 35 kDa protein endowed with dermonecrotic, hemolytic and sphingomyelinase activities (Tambourgi et al., 1998). A recent study has led to the characterization of the anti-DNF neutralizing mAb LimAb7 (Alvarenga et al., 2003). Various experiments were conducted with the aim of mapping the epitope recognized on the DNF. Indeed, four mimotopes which did not show similarity with the DNF sequence were selected by phage display for binding to LimAb7 (Moura et al., manuscript in preparation). MIMOP was run with the available mimotope
sequences and the \textit{L.intermedia} DNF sequence, protein LiD1 (Kalapothakis \textit{et al}., 2002). Since the 3D structure of the Ag was not available, a theoretical model of the DNF was determined based on methods of homology modeling (Moura \textit{et al}., manuscript in preparation). In the MimAlign approach, the four programs of multiple sequence alignment were used. In the MimCons approach, mimotope sequences were clustered according to the Levenshtein distance, and consensus patterns were identified on peptide sequences. The PER proposed by MIMOP (combined analysis of MimAlign and MimCons) comprised residues C197, Y224, W225, and consensus patterns from the mimotope sequences. However, no PER could be predicted since the 3D structure of the Ag is not available. The MimAlign analysis was performed using the four programs of multiple sequence alignment. Selected convergent positions of the Ag were determined on the basis of the combined alignment and highlighted on the Ag sequence (Fig. 6). Most of these positions showed high scores of similarity and low frequency. These residues are continuous in the Ag sequence and map positions 380–387. Since this region is similar to residues 380-389 of gp91phox, the proposal of Burritt \textit{et al}., (1995) is also indicated.

\textbf{3.2.4 Case in which no information on the epitope and no 3D structure of the Ag are available} In this particular case, MIMOP was still able to propose PER even though no 3D analysis could be conducted for identification or scoring of PER on the surface of the protein. To illustrate this typical case, the data from Burritt \textit{et al}., (1995) were analyzed by MIMOP and the prediction compared with these later. In this study, 21 mimotopes were selected by phage display technology using mAb 54.1 with the aim of mapping the epitope on the gp91phox cytochrome \textit{b} subunit. The NADPH-oxidase system of neutrophils is a host-defensive plasma membrane redox system that kills invading microbes and causes damage to tissue. Humans lacking this enzyme system suffer from recurrent bacterial infections, granulomatous lesions of multiple organs and early death. Cytochrome \textit{b} of human neutrophils is the central component of the microbicidal NADPH-oxidase system. This component is a heterodimeric integral membrane protein composed of 91 kDa (gp91phox) and 22 kDa (p22phox) subunits. The mAb 54.1 recognizes specifically the gp91phox cytochrome \textit{b} subunit. Burritt \textit{et al}., (1995) selected 21 mimotopes for their binding to this Ab. Among the selected mimotopes, the authors identified the consensus amino acid sequence PKKAVDGP, which is similar to residues 380-389 of gp91phox. Their mapping data confirmed that the epitope recognized by mAb 54.1 binds amino acids 382-389 of gp91phox.

A MimCons analyses would give information on the similarity and consensus patterns from the mimotope sequences. However, no PER could be predicted since the 3D structure of the Ag is not available. The MimAlign analysis was performed using the four programs of multiple sequence alignment. Selected convergent positions of the Ag were determined on the basis of the combined alignment and highlighted on the Ag sequence (Fig. 6). Most of these positions showed high scores of similarity and low frequency. However, eight residues were striking because they had both high similarity and frequency scores. These residues are continuous and included positions 380–387. Since this region is sequential and with scores higher than the other selected positions, this fragment was considered as a PER. Moreover, this fragment can be C-terminally extended by two positions since G388 and P389 have also been selected, though with lower frequency. Thus, the PER predicted by MIMOP is 380KLPKIAVDGP389, whereas Burritt \textit{et al}., (1995) proposed 382PKIAVDGP389. Therefore, and once again, the prediction of MIMOP was in good agreement with the results found by Burritt \textit{et al}., (1995).
DISCUSSION

It is assumed that the collection of random peptides that bind specifically to a monoclonal antibody defines the structural correlate of the epitope the antibody recognizes on its cognate antigen (Felici et al., 1993). However, inferring the 3D arrangement of an epitope from the linear sequences of mimotope peptides is not trivial. We have developed a new computational tool that provides ranked predictions of PER from a set of mimotope peptide sequences.

We have found four studies in the literature describing a computational method that can be used to predict an epitope on a given protein Ag from mimotope sequences. Two of them, FINDMAP (Mumey et al., 2003) and 3DEX (Schreiber et al., 2005), use only one peptide at a time: the automatic analysis of the total set of peptide sequences does not seem to be possible. FINDMAP is only based on sequence since the aim of the method is slightly different from others. Indeed, the authors’ intention is to develop a method to acquire information on the 3D structure of the protein: by identifying discontinuous epitopes, they can identify residues that are close to each other on the surface of the protein. 3DEX (3D-Epitope-Explorer) is an algorithm which proposes to localize conformational epitopes in 3D protein structures from a set of amino acids. Although this tool appears to be well adapted to the localization of a binding site on a protein surface when the composing residues are known, it appears to be less suited to mapping the epitope from a set of mimotopes. This tool attempts to find all the amino acids of a given mimotope on the Ag surface but with a limited possibility of replacing them by similar amino acids. The amino acids on the Ag surface are then sought for one after the other, in a sequential order; however, all the residues of a mimotope are not needed for interaction with the Ab and are rarely in an identical sequence order on the Ag surface. Both FINDMAP and 3DEX algorithms were applied to particular analyses, but they probably have to be improved to cope with either wider sets of data or appropriate data (i.e. mimotopes selected with a mAb and not a pAb). The two other tools developed by Enshall-Seijffers et al. (2003) and Halperin et al. (2003) allow, in contrast, the analysis of all mimotopes from an experimental set. The method described by Enshall-Seijffers et al. (2003) is based on the identification of pairs of residues in the mimotopes that are then mapped on the protein surface. SiteLight (Halperin et al., 2003) compares amino acids from the mimotopes with amino acids in patches determined at the surface of the protein. Although these tools seem to be well suited for the analysis of mimotopes, we feel they lack some intermediary analyses, sometimes essential for interpretation of the experimental results. For example, different elements of the same epitope can be mimicked by distinct mimotopes from the same set of sequences. All these described methods do not allow testing other approaches if the running analysis does not lead to a relevant prediction.

We have tried to overcome some of these weaknesses by developing MIMOP. This new tool provides an environment for mimotope characterization that delivers to the user different analytical results. MIMOP is able to predict either continuous or discontinuous epitopes. Because it is used via a flexible web interface, the user can easily intervene to adapt the analysis to his dataset or to his specific knowledge of the Ag. In addition, the user can refine the analysis by running MIMOP with a subset of similar mimotopes or with a subset showing a defined consensus pattern. MIMOP provides intermediate results (mimotope clustering, multiple alignments, consensus patterns, key residues, etc.) at any step of the analysis, insuring that the user obtains at least some answers.

A key step in releasing a new prediction tool is the validation step. In the case of epitope prediction, this is not a trivial issue mainly because of the scarcity of useful data. In fact, numerous studies have been performed to identify mimotopes, but the mimicked epitope is rarely known. Validation is made difficult because the predicted epitope must be compared with the reference epitope to assess whether the prediction is correct or not. However, the epitope itself can be depicted in different ways. Indeed, Van Regenmortel (1998) described a structural epitope that comprises 15–20 amino acids in contact with the paratope and a functional (or energetic) epitope composed of only 3–5 amino acids contributing to the binding energy of interaction. Novotny et al. (1989) showed earlier that only a limited set of the contact residues in a structural epitope actually participate in the functional epitope. It is now commonly accepted that interaction between two proteins is more dynamic than static involving molecular breathing, exchange of hydrogen bonds, influence of distant residues, etc. (Janin, 1997). Finally, as a consequence of the ability of antibodies to cross-react, the mimicked epitope can either correspond totally or partially to the reference epitope. For all these reasons, it is difficult to compare a predicted epitope with a reference epitope in order to strictly evaluate algorithms like MIMOP. Because there are now several tools aimed at predicting epitope localization from mimotope sequences and because they have all been validated using different experimental datasets, we suggest that a contest (like CASP for protein structure prediction) be set-up to strictly compare the efficiency of epitope prediction algorithms.

We have chosen to present herein various sets of data that show the ability and usefulness of MIMOP in predicting epitopes and for guiding further experiments. Four different Ag models were used: the C2 domain of factor VIII, the Cn2 scorpion toxin, the DNF enzyme and the cytochrome b molecule. In the two cases where the reference epitope had been experimentally defined previously (C2 domain and Cn2 toxin), MIMOP was able to propose PER that coincided with the epitope. In the case of the enzymatically active DNF molecule, there was no available information on the structure or nature of the epitope recognized by mAb LimAb7. However, the predictions of MIMOP were consistent with the known functional activity of the Ag and readily explained the capacity of mAb LimAb7 to inhibit the catalytic activity of DNF. Thus, although it was not possible to perform a complete, statistically relevant study of the efficiency of the MIMOP algorithm to predict epitopic regions of proteins, the three cases where we were able to check prediction with available structural or functional information yielded predictions that make sense. The prediction obtained with the three methods (MimAlign, MimCons and combined) converged to a unique prediction in the case of the Cn2 epitope only. This is not unexpected because MimAlign and MimCons achieve prediction based on different hypotheses. In two other cases, the combined approach yielded the most accurate prediction with 8/11 and 10/16 of the amino acids of the top-ranked region corresponding to residues of the reference epitope for the C2 domain and the Cn2 toxin, respectively.

As experimental data from phage display increases, the MIMOP tool should be useful in helping experimentalists to interpret their
results. In the future, this tool can also be generalized to predict the interface between two proteins.

ACKNOWLEDGEMENTS

The authors thank Dr S. L. Salti for presubmission editorial assistance. We are grateful to B. Leblanc and B. Jardin for fruitful discussions. The authors also thank U. Cantrelle for his contribution to the validation tests of MIMOP. Funding to pay the Open Access publication charges for this article was provided by CNRS.

Conflict of Interest: none declared.

REFERENCES


Licea,A.F. et al. (1996) Fab fragments of the monoclonal antibody BCF2 are capable of neutralizing the whole soluble venom from the scorpion Centruroides noxius Hoffmann. *Toxicin*, 34, 843–847.


