JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

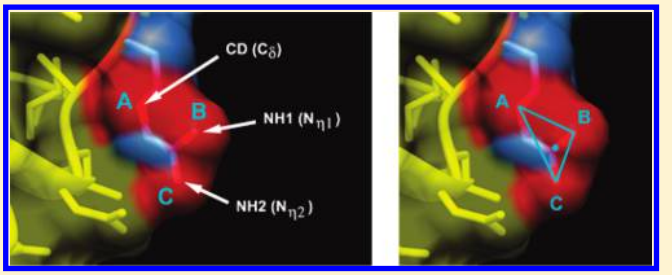
LocaPep: Localization of Epitopes on Protein Surfaces Using Peptides from Phage Display Libraries

Luis F. Pacios,*[,†] Leticia Tordesillas,[‡] Arantxa Palacín,[‡] Rosa Sánchez-Monge,[‡] Gabriel Salcedo,[‡] and Araceli Díaz-Perales[‡]

[†]Department Biotecnología, ETSI Montes, Universidad Politécnica de Madrid, 28040 Madrid, Spain

[‡]Department Biotecnología, ETSI Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid, Spain

ABSTRACT: The use of peptides from a phage display library selected by binding to a given antibody is a widespread technique to probe epitopes of antigenic proteins. However, the identification of interaction sites mimicked by these peptides on the antigen surface is a difficult task. LocaPep is a computer program developed to localize epitopes using a new clusters algorithm that focuses on protein surface properties. The program is constructed with the aim of providing a flexible computational tool for predicting the location of epitopes on protein structures. As a first set of testing results, the localization of epitope regions in eight different antigenic proteins for which experimental data on their antibody interactions exist is correctly identified by LocaPep. These results represent a disparate sample of biologically different systems. The program is freely available at http://atenea.montes.upm.es



1. INTRODUCTION

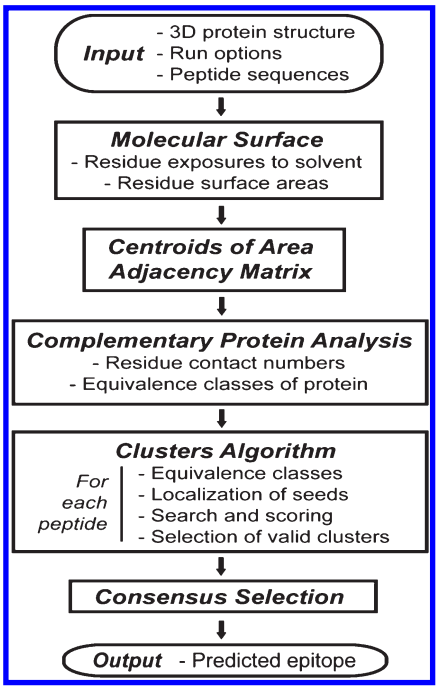
In contrast to the progress made in T-cell epitope predictions, the development of predictive methods for B-cell epitopes has proven a much more challenging task. Recent reviews evaluating the performance of a variety of strategies for predicting B-cell epitopes[1−4] reveal that there is no method available yet to achieve this goal in a reliable way. Insofar as the existing prediction tools use a variety of strategies and as there are a number of databases that store a considerable amount of data on antibody−antigen (Ab−Ag) systems, this status points to more fundamental difficulties. It is widely accepted that epitopes recognized by an Ab are exposed on the protein surface.[5] It has been estimated that >90% of B-cell epitopes are discontinuous (conformational) and composed of several fragments scattered along the sequence and brought together in spatial proximity at the folded 3D structure,[6] although they often include small continuous (linear) segments in the sequence. Other physicochemical properties, such as flexibility, hydrophilicity, and polarity, have been also considered in order to identify amino acid propensities at the interfaces of Ab−Ag interactions in complexes of known structure.[7] However, it is still not clear what makes a specific set of antigenic residues and how the structural environment affects their selection.

While the most accurate way to determine an epitope is solving the 3D structure of the complex formed by the antigenic protein and its Ab, this resource is usually not available for most researchers. Phage display technology[8,9] has become an epitope mapping method frequently used. A phage display is a mixture of phage clones each carrying a different peptide insertion into their surface protein. In this way, each phage displays a different peptide on its surface. This technology is used to select from a set of random peptides those with a high binding affinity to an Ab.[10,11] The selected peptides are assumed to mimic the epitope in terms of spatial organization and physicochemical properties. The term 'mimotope' (literally 'mimic of an epitope') was proposed in 1986.[12] Mimotopes and Ag are both recognized by the same Ab paratope, and therefore mimotopes are expected to imitate part of the 'true' epitope, even though no clear homology may exist between their sequences and those of the epitope of the native Ag.

However, it is difficult to identify with precision the actual epitope mimicked by mimotopes because although they show a partial functional equivalence, they often do not share sequence similarities. The algorithmic goal is to utilize the information contained in the set of peptides for correctly predicting the corresponding epitope on the surface of the antigen. The first work combining computational methods and experimental results to identify epitopes from mimotopes was reported in 1995.[13] Since then, a few automated procedures, either computer programs or web servers, have been developed based on different approaches which can be classified into three categories: Programs in the first category, such as Findmap, only work with sequence information from mimotopes and antigen.[14] Methods in the second category use both sequence data and the antigen 3D structure. SiteLight,[15] 3DEX,[16] Mapitope,[17−19] and PepSurf[20] belong to this category. Finally, web servers, such as MIMOX,[21] MIMOP,[22] and Pepitope[23] make the third category integrating in different ways the two preceding approaches.

Received: February 8, 2011
Published: May 03, 2011

Scheme 1. Flowchart of LocaPep



While these computational tools appear to be efficient for particular analyses, they often predict several alternatives to the putative epitope. Some of them rely more on information regarding mimotope sequences than on relationships between protein surface characteristics and features of every peptide. Sometimes, their use is not flexible enough for fast in silico experiments which may be more feasible running standalone programs with full control of input instead of automated servers. In the present report we describe LocaPep, a new computer program which uses an algorithm that focuses on protein surface features for predicting potential epitopes from the input given by phage display selected peptides. The main differences between LocaPep and other servers and programs are briefly described in the Discussion Section below.

2. METHODS

LocaPep is written in Fortran90 without resort to any specific library. The source included in the distribution can be compiled with several Fortran compilers including the freely available open-source g95 or gFortran. The program runs in command-line mode and is organized (Scheme 1) as follows.

2.1. Input. The program starts reading input from a plain-text file named 'locapep.inp'. The first line specifices the name of the file with the 3D protein structure in PDB format. The four following lines give values for parameters that control run options: KRun, KSea, KMin, and KPrn. KRun controls whether the SES surface (Section 2.2), the most time-consuming step, is calculated (KRun = 0) or the surface data from a previous run are read instead (KRun $\neq$ 0). KSea selects the strategy to organize the searching process in the clusters algorithm (Section 2.5). KMin sets the minimum number of residues localized in a cluster to consider it valid (Section 2.5). KPrn is a parameter that just controls the amount of printing in output. Next line gives NSeq, number of peptide sequences to be read, and finally a block of NSeq lines give these sequences in one-letter format. A NSeq = 0 run (no peptides read) can be used to just compute the molecular surface and related properties as well as the adjacency matrix (Section 2.3) for subsequent KRun $\neq$ 0 calculations. Details on input are given in a brief manual included in the distribution.

2.2. Molecular Surfaces. Molecular surfaces are calculated using the GEPOL analytic procedure[24] fully implemented in LocaPep in a similar way to that done before with other software.[25,26] The solvent-accessible surface (SAS) is first obtained to compute percentages of exposure to solvent of protein residues using SAS stochastic standard states of amino acids[27] for reference. The solvent-excluded surface (SES), the actual molecular surface, is then computed using the GEPOL tessellation algorithm with options NDIV = 3, OFAC = 0.80, and RMIN = 0.50, which guarantee a highly reliable representation of the protein surface,[24] and probe radius of 1.4 Å. The main reason to use GEPOL in our work is the capability of this algorithm to partition any molecular surface (SAS or SES) into atomic contributions, an issue central to our representation of protein residues as centroids of area (Section 2.3) upon which LocaPep is based. Compared with other computational tasks performed by our program, this is the only time-consuming step in a full calculation. Computer times for calculating the SES increase with protein size: about 4 s for a 100-residue protein, yet about 40 s for a 300-residue protein. For medium or large proteins, it should be advisible to run first a KRun = 0, NSeq = 0 case to just compute surfaces data and then run as many Krun $\neq$ 0, NSeq $\neq$ 0 calculations as needed to analyze different peptide sets or explore option parameters. Once SAS and SES data obtained in a previous run are read, a whole calculation with standard options for a 300-residue protein and about 50 peptides takes only a few tenths of seconds. The GEPOL procedure allows to partition the molecular surface into atomic and hence residue contributions. With SES residue areas and SAS percentages of exposure to solvent, LocaPep computes minimum, maximum, and mean values of the 20 amino acids in the protein. These values are then used for scoring as explained below (Section 2.6).

2.3. Centroids of Area and Adjacency Matrix. After obtaining the atomic contributions to the SES, each residue is represented by a centroid of area defined with the three atoms having the largest areas s_A, s_B, and s_C. Centroid coordinates $\{X,Y,Z\}_i$ for residue i are thus

$$X_i = \frac{S_A X_A + S_B X_B + S_C X_C}{S_A + S_B + S_C} \qquad (\text{ibid.} Y_i, Z_i)$$

where x_A, x_B, and x_C are X coordinates of atoms A, B, and C. Since these centroids are used to determine which residues are adjacent on the protein surface, a correction is introduced for long side chains with great exposure to solvent. Given that in these cases A, B, and C are often atoms distant from backbone, the centroid of area is then defined with atoms closer to the peptide chain to avoid excessively long distances from neighbor residues at the surface. The selection of atoms A, B, and C is thus corrected taking into account: (i) atom areas, (ii) side chain length, and (iii) the magnitude of exposure to solvent as compared with mean values in the protein and reference data.[27] Once centroids of area are defined for the N protein residues, an $N \times N$ symmetric (0,1) matrix is constructed as follows: Entry (i,j) is 1 if residues i and j are adjacent in the protein surface and 0 otherwise. Two residues are considered adjacent if the distance

between their centroids of area is ≤ TolDist. A threshold value TolDist = 8.5 Å, which agrees with previous results,[22,28] was set after exploring several proteins with different surface topography. Only residues with exposure to solvent ≥1% are used to construct this matrix.

2.4. Complementary Protein Analysis. LocaPep also uses residue contact numbers. For a residue *i*, this property is defined as the number of Cα atoms within a distance of 10 Å of Cα atom in *i*. It has been reported that residues identified as part of epitopes have significantly lower contact numbers compared to nonepitope residues in a data set of 76 crystal structures of Ab-Ag complexes.[29] As done with residue areas and exposures, minimum, maximum and mean values of contact numbers are computed for the 20 different amino acids in the protein and used for scoring (Section 2.6). In order to identify equivalent amino acids in protein and mimotopes, equivalence classes are defined grouping amino acids by taking into consideration chemical characteristics of side chains as follows:

1 − acidic: D, E	2 − basic: R, K	3 − aromatic: F, W, Y	
4 − polarl: N, Q	5 − polar2: S, T	6 − nonpolar: I, L, V	
7 − sulfur: C, M	8 − unique1: H	9 − unique2: P	10 − small: A, G

Equivalence classes in each peptide together with their distribution in the protein constitute the central information to accomplish the search for clusters as explained next.

2.5. Clusters Algorithm. The main LocaPep task to map mimotopes data onto antigen surfaces is to search for clusters, i.e., patches of adjacent residues in the surface equivalent to amino acids in mimotopes. A cluster is assumed to form around a seed residue. The algorithm is applied to each peptide in input and consists of two steps: (1) selection of seeds and (2) searching for clusters.

Step (1) selects protein residues for a number of equivalence classes in each peptide fixed according to KSea input parameter (Section 2.1). If KSea = 0, then the program selects up to 5 equivalence classes by picking classes 1−3 plus those of classes 4−9 for which the protein had five residues or less. Class 10 is never used for setting seeds. If the protein lacks amino acids in some classes and 5 have not been selected yet, then the program picks as many classes 4−9 as needed regardless the number of protein residues in them. If KSea = *M*, the user must give in input the *M* class numbers desired. The protein surface is then scanned for selecting seeds among residues belonging to each equivalence class. To act as seed, a residue has to display significantly large surface area and exposure and small contact number as compared with mean and extreme values in the protein for the corresponding amino acid. All residues in each equivalence class are analyzed, and those not fulfilling these constraints are discarded unless the protein had three or less residues in that class in which case, all of them are valid seeds.

Step (2) searches in residues adjacent to each seed (one entry in seed row of the adjacency matrix) for all those with same equivalence classes than peptide amino acids. Once a valid residue is localized, the program computes an individual score for it (Section 2.6). If after exploring this first adjacency ring around a seed, no residues are found for a class, then equivalent residues are searched for in the second adjacency ring (one entry in rows of the adjacency matrix corresponding to residues in the first ring). If more residues than needed are found for an equivalence class, then the selection to set a final cluster is done on the basis of greater residue scores.

A cluster is valid only if the number of residues localized is greater than a minimum predetermined value and is discarded otherwise. This value is fixed by KMin parameter (Section 2.1). If KMin = 0, only clusters with at least $(Q/2) + 2$ residues localized are valid, Q being the number of identifiable peptide amino acids (those belonging to equivalence classes present in the protein). KMin may be modified in input to change this minimum. For 12-mer peptides in usual phage display libraries, default KMin = 0 leads to keep only clusters with at least 8 residues found (provided that all peptide amino acids had possible equivalent partners in the protein). If for a peptide set it happens that too few clusters (or no clusters at all) are found, then new searches to keep clusters with ≤7 residues should be run.

2.6. Calculation of Residue Scores. For each residue *i* of amino acid type *k* localized in a cluster, a total score is calculated as the weighted sum of the following four scores defined in the range (0,1):

(1) Area score, scoA_i. Let AMin, AMax, and AMean be minimum, maximum, and mean values, respectively, of surface areas of type *k* in the protein. A tolerance factor is first computed as ATol = (AMin + AMean)/2. Let A_i be the surface area of residue *i*. If A_i > ATol, then the area score is calculated as scoA_i = A_i/AMax. If A_i ≤ ATol, scoA_i = 0.0

(2) Exposure score, scoE_i. It is computed with EMin, EMax, EMean, and E_i values corresponding to percentages of exposure exactly as scoA_i.

(3) Contacts score, scoC_i. Let CMin and CMean be minimum and mean values, respectively, of contact numbers of type *k* in the protein. A tolerance factor is first computed as CTol = (CMin + CMean)/2. Let C_i be the contact number of residue *i*. If C_i < CTol, then this score is calculated as scoC_i = (CMean C_i) /CTol. If C_i ≥ CTol, then scoC_i = 0.0, unless *i* is the only protein amino acid of type *k* in which case, scoC_i = 1.0.

(4) Distance score, scoD_i. LocaPep calculates first minimum distances of each residue to its closer neighbor on surface, DMin. Let D_i be the distance between residue *i* and seed residue of the cluster. If D_i < 10 Å, then this score is calculated as scoD_i = DMin/D_i. If D_i ≥ 10 Å, then scoD_i = 0.0.

Total score for residue *i* is finally computed as:

$$\text{score}_i = 0.20\text{sco}A_i + 0.30\text{sco}E_i + 0.30\text{sco}C_i + 0.20\text{sco}D_i$$

where weight factors aim to emphasize slightly the two properties that have been repeatedly highlighted in studies on the nature of epitope amino acids. Irrespective of values for properties 1−4, score$_i$ is set to 1.0 if residue *i* is a seed or if it belongs to an equivalence class for which the protein has a number of residues equal to or less than the number of equivalent amino acids in the peptide. This scoring is directly applied to clusters: a cluster composed of *n* residues has a total score$_T$ = ∑score$_i$, *i* = 1,*n*.

2.7. Consensus Selection = Predicted Epitope. To select the residues that define the predicted epitope, two final tasks are performed. First, all score$_i$ values corresponding to all times that residue *i* has been localized in valid clusters are added to obtain a total score TotSco$_i$. Residues are then sorted by decreasing TotSco$_i$, and those having values greater than a certain threshold (which depends on the number of peptides and clusters found), make set {$\mathbf{S}_i$}. Second, for all seed residues corresponding to best clusters (those showing the greatest score$_T$ for each peptide), a parameter is calculated as TotSco$_i$ multiplied by the number of

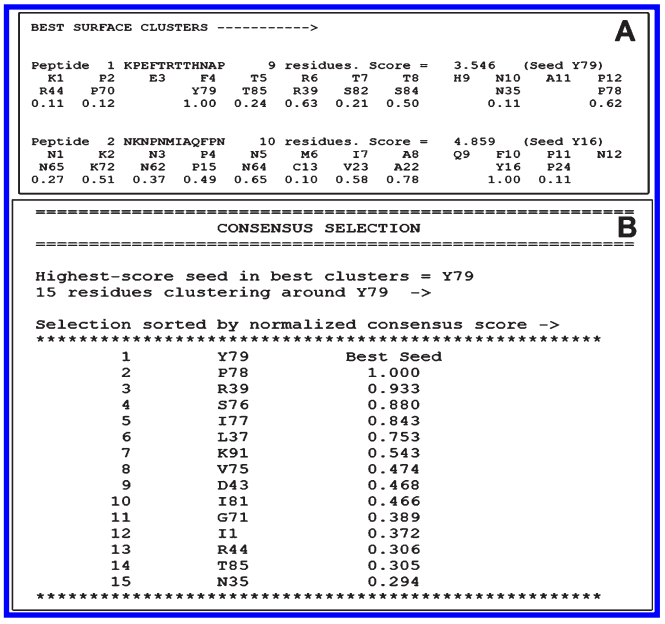


Figure 1. Two fragments of LocaPep output. (A) Example of output on best clusters. (B) Example of output on final consensus selection (predicted epitope).

times each seed appears in the collection of best clusters. The greatest value of this parameter selects the best seed for the final consensus cluster, BS. $\{S_i\}$ residues that are within a 2TolDist distance from BS and have greater areas and exposures and lower contacts than mean values of their amino acid types make the consensus selection that forms the predicted epitope. To avoid too large a number of residues in the epitope, this selection is repeated within a more restricted domain defined by a (TolDist + 4) distance from BS if more than 20 residues were formerly selected. With the relative value of those properties, a composite factor ConFac is calculated for each residue within the consensus group. Excluding the leading BS residue and normalizing to the highest factor, ConFac values define a normalized consensus score (Figure 1) that permits to assess quantitatively the relative weight of each residue in the predicted epitope.

2.8. Availability of LocaPep. Source code, input files, manual, and binaries for both x86_64 Linux and 32-bits Windows are freely available at: http://atenea.montes.upm.es.

3. RESULTS

LocaPep is a computer program aimed to identify amino acids in a putative epitope mapped onto the antigen surface by a set of mimotopes. The output is a basic text file. Since excellent software exists to render molecular graphics for proteins (most freely available to academia), the task to present this information in a graphic form is best managed by this software. A brief overview of LocaPep output follows (Section 3.1), and then a sample of results for eight antigens of different biological nature for which experimental data on their epitopes location exist (Section 3.2) is presented.

3.1. LocaPep Output. For a full KRun = 0 calculation output starts with information on GEPOL surfaces. Data needed for the clusters algorithm are written to separate files that store atom area contributions, residue properties, and adjacency matrix. For a KRun $\neq$ 0 calculation, LocaPep starts reading these files, and the first output is information on the antigen. Next task is the localization of protein residues by applying the clusters algorithm to each peptide. The amount of information on this process written to output is controlled by KPrn input parameter. Minimum printing (KPrn = 0) only lists valid and discarded clusters, whereas maximum printing (KPrn = 2) includes detailed information on all trials attempted in the search process. This can produce a considerable amount of printing for large proteins and large peptide sets. Upon completion of the search for clusters, a summary of results follows. Valid clusters for each peptide are listed indicating number of amino acids identified, seed residue, and cluster score. Then, best clusters for each peptide displaying mapping between equivalent peptide amino acids and antigen residues with individual scores are listed (Figure 1). All protein residues in valid clusters with score greater than a threshold fixed by the program are then gathered into two lists, one sorted by sequence number and another sorted by score. Final output is the consensus selection that defines the predicted epitope. After a list of these amino acids written by protein sequence number, the consensus selection is given showing BS (Section 2.7), and residues arranged by decreasing order of the normalized consensus score (Figure 1).

3.2. Sample of LocaPep Results. The following sample of results is aimed at presenting a set of validation cases and consists of eight antigenic proteins for which experimental data on the location of epitope regions are available. Six proteins correspond to complexes with known crystal structure (five Ab−Ag and one receptor−protein complex), and two are uncomplexed proteins for which evidence not relying on complex data exists on epitope locations. Except where indicated, default KSearch = 0, KMin = 0 runs were used. No metrics have been used in the comparison with experiment for two reasons: First, even when a crystal structure exists for an Ab−Ag complex, the composition of an epitope has usually no definite limits. Second, the use of a mimotope library is basically a procedure intended to mimic the whole or part of a true epitope. Hence, as far as LocaPep provides a single prediction with an internal consistent scoring and the problem studied lack closed solution in terms of a definite, unique collection of protein residues, the performance of our program is illustrated in the following case systems without resort to external metrics.

3.2.1. VEGF Protein. Vascular endothelial growth factor (VEGF), a heparin-binding protein, is an angiogenic factor in development as well as in the growth of tumors. The high-resolution X-ray structure of the complex between a neutralizing Ab (Fab-12; IgG form known as rhuMAb VEGF) and the receptor binding domain of VEGF[30] was chosen for this first case (PDB ID 1BJ1). In a combined research making use of crystal analyses, alanine scannings, phage libraries, and site-directed mutagenesis, an optimized anti-VEGF Ab was obtained.[31] We used the set of 42 mimotopes (37 hexapeptides, 3 pentapeptides, and 2 tetrapeptides) from these phage-displayed libraries available at the benchmark data set in Pepitope server[23] to probe VEGF protein. LocaPep found nine residues at or near the Ab-binding interface (Figure 2A). In particular, three residues that play a key role in Ab-binding, identified in both crystal and alanine scanning analyses (Q79, G92, and E93), are correctly identified as they are included in the predicted epitope.

3.2.2. HIV-1 Capsid Protein p24. A few examples of highly cross-reactive and neutralizing anti-HIV specific monoclonal antibodies (mAbs) have been described over the years. The identification in the viral antigen of epitopes of clinically interesting mAbs is a task of great importance. HIV-specific mAbs

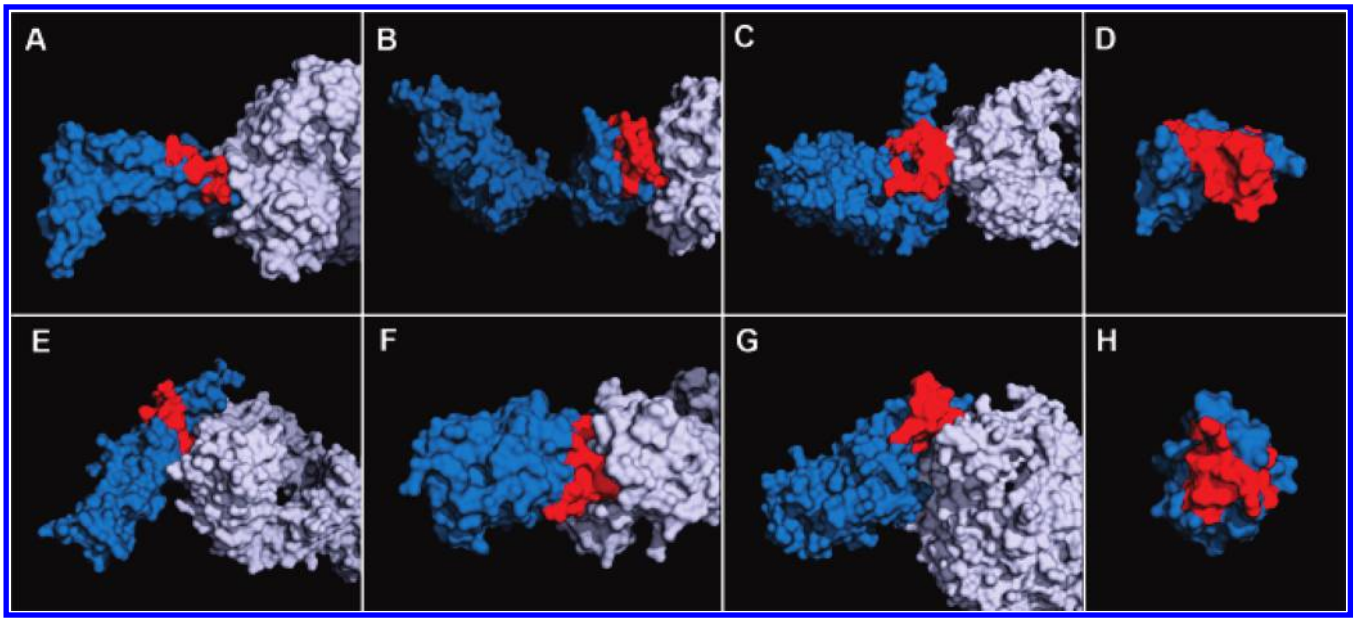


Figure 2. Epitope sites predicted by LocaPep (red) on the surface of antigens (blue) complexed with Fab fragments (A−C, E, and F) or receptor (G) (bluewhite) and two uncomplexed proteins (D and H). PDB ID codes for the 3D structures are given in parentheses (A) VEGF protein (1BJ1). (B) HIV-1 capsid protein p24 (1E6J). (C) HIV-1 envelope protein gp120 (1GC1). (D) Scorpion neurotoxin Cn2 (1CN2). (E) Her-2 protein (1N8Z). (F) Human factor VIII C_2 domain (1IQD). (G) SARS coronavirus spike protein (2AJF). (H) Peach Pru p 3 protein (2B5S). Surfaces rendered with PyMOL 1.3 (http://pymol.org).

have been employed to screen phage display peptide libraries and to use then mAb-specific phages as a database to analyze structures of viral antigens in order to identify epitopes.[17] One of the systems studied this way was the complex between mAb 13b5 and the HIV-1 capsid protein p24. As a second test, we used the crystal structure of p24 in the 13b5/p24 complex[32](PDB ID 1E6J) and the set of 16 different 13b5-specific dodecapeptides isolated[17] to scan the p24 surface. LocaPep identified a 14-residue epitope. Except one (that having the lowest normalized consensus score), the other 13 amino acids locate between positions 196 and 220 (Figure 2B). This prediction agrees with the actual experimental epitope of p24 that spans the outer region between positions 195 and 217.[17]

3.2.3. HIV-1 Envelope Protein gp120. Directly related with the preceding system is the ternary complex of the HIV-1 envelope protein gp120 with a truncated part of its soluble receptor CD4 and a Fab fragment of the mAb 17b, crystallized at 2.5 Å resolution.[33] As a third test, we used the gp120 structure of this complex (PDB ID 1GC1) and 11 random dodecapeptides isolated through comprehensive screening of three combinatorial phage display peptide libraries.[17] LocaPep identified a 15-residue epitope composed of residues 113, 114, 119, 120, 122, 199−206, 429, and 435 (Figure 2C). This prediction is in fair agreement with the genuine epitope identified in the experimental structure of gp120 which comprises residues 119, 120, 122, 200−205, 419, 421−423, 432, 435, and 437. Regarding the somewhat peculiar location of this epitope on the gp120 surface, it should be noted that it was derived from the 17b peptide library, but it is known that it overlaps the epitope corresponding to mAb CG10 as long as 17b and CG10 compete for CD4/gp120 binding.[17]

3.2.4. Scorpion Neurotoxin Cn2. Scorpion toxins are small proteins evolutionary adapted to recognize and block ion membrane channels. Cn2 is the most abundant and one of the most potent toxins in the venom of scorpion *Centruroides noxius*. mAb BCF2 is able to neutralize the toxic effects of Cn2, and hence, the Cn2-BCF2 system has been investigated on the development of vaccines against scorpion venom.[34] However, no experimental structure for this Ab−Ag system exists, and only a NMR solution structure of Cn2 is available.[35] A large collection of mimotopes that seem to reproduce the antigenic and immunogenic specificity of the Cn2 epitope was isolated from a phage displayed 12-mer peptide library.[34] By using a 27-mimotope set from this collection and the structure of Cn2 (PDB ID 1CN2), LocaPep identified an epitope that comprises residue 1 plus 11 amino acids spanning sequence positions 7−10, 56−58, and 60−63 (Figure 2D). Eight (K8, N9, T10, I56, L60, P61, N62 and K63) out of these 12 residues are among those experimentally localized by Gazarian et al.[34] On the other hand, residues 1−15 and 54−66 were also identified in experiments with synthetic peptides.[36]

3.2.5. Human Epidermal Factor Receptor Her-2. The human epidermal growth factor receptor 2 (Her-2 or Her-2/neu) is a member of the epidermal growth factor receptor (EGFR) family of receptor tyrosine kinases. EGFR receptors are essential in regulation of cell proliferation. Hence, their overexpression and uncontrolled activation is associated with key features of cancer. Her-2 contains a large extracellular domain accessible to components of the immune system so that mAbs targeting Her-2 have been generated.[37] The Ab-binding site of Her-2 was first characterized using the crystal structure of this extracellular domain complexed with trastuzumab,[38] a member of these mAbs which inhibits growth of Her-2 overexpressing tumor cells. With 5 decapeptides defined for the epitope recognized by trastuzumab, the binding site was identified to comprise three loop regions in Her-2: residues 557−561, 570−573, and 593−603.[39] We used the structure of Her-2 in the complex (PDB ID 1N8Z), and these 5 decapeptides to localize that epitope. With a standard

KMin = 0 run, LocaPep found no clusters because the minimum number of residues to validate clusters using decapeptides is by default set to 7, and this case involves a very small set of mimotopes. Upon decreasing this minimum to 5, the program found a consensus selection composed of 11 residues: 567, 568, 572, 573, 575−577, 580, 596, 597, and 604 (Figure 2E). Although only five predicted residues are strictly within the experimental epitope, the six others happen to be very close in Her-2 surface.

3.2.6. Human Factor VIII (FVIII) C2 Domain. FVIII is a multidomain cofactor in the generation of thrombin. Dysfunction or deficiency in FVIII results in hemophilia. Therapeutic administration of exogenous FVIII to hemophiliacs gives rise to emergence of Abs that inhibit FVIII procoagulant activity. Inhibitor Abs are preferentially directed against epitopes in A_2, A_3 and, especially, C_2 domains. A crystal structure at 2.0 Å resolution of the C_2 domain complexed with Bo2C11, a human IgG mAb representative of anti-C_2 inhibitors, is available[40] (PDB ID 1IQD). Using a phage display peptide library, a set of 27 mimotopes (4 decapeptides and 23 dodecapeptides) was selected against Bo2C11 targeting the C_2 domain of FVIII.[41] We used this peptide set and the crystal structure of FVIII C_2 domain in the Bo2C11−FVIII complex as a test. LocaPep localized an epitope composed of 18 residues that span sequence positions 2195−2203, 2214−2216, 2220, 2222, 2250, 2251, 2253, and 2315 (Figure 2F). The experimental Ag-binding site identified in the X-ray structure comprises 15 residues at positions 2196−2200, 2215, 2220, 2222, 2223, 2250−2253, 2315, and 2316. Our prediction agrees thus remarkably well with experimental data. Moreover, the flexibility of LocaPep in enabling to localize surface sites for mapping a single peptide allowed us to carry out a supplementary validation test. Villard et al.[41] observed that one particular peptide (107 in their notation) had the most potent neutralizing capacity toward Bo2C11-mediated FVIII inhibition both in vitro and in vivo. This led the authors to propose peptide 107 in possible therapeutic applications. Upon running LocaPep with this single peptide, the consensus selection was composed of the following 10 residues: 2197−2202, 2214−2216, and 2220. This result permits to identify these amino acids as candidate to participate in the Ab-binding to peptide 107.

3.2.7. SARS Coronavirus Spike Protein. Severe acute respiratory syndrome (SARS) emerged as an infectious disease causing a serious worldwide outbreak in recent years. A novel SARS coronavirus (SARS-CoV) was associated with progressive pneumonia and respiratory failure.[42] The receptor for SARS-CoV was promptly identified, and neutralizing Abs that block receptor association were developed.[43] Two independent panels containing 42 and 18 peptides were isolated to localize the receptor-binding site of CoV spike protein.[18] However, mapping the information provided by both peptide panel 1 (9 16-mer, 10 15-mer, 19 14-mer, and 4 13-mer peptides) and panel 2 (18 15-mer peptides) onto the surface of CoV spike protein proved challenging. The Mapitope algorithm employed for this task predicted for both panels three clusters of comparable significance located at separate regions corresponding to positions 450−480 (cluster A), 318−330 and 440−445 together (cluster B), and 334−384 (cluster C).[18] During the course of this study, the X-ray structure of the SARS-CoV receptor-binding domain complexed with its receptor, angiotensin-converting enzyme-2 (ACE2), was made available.[44] This allowed the authors to assess the genuine site which happens to comprise three segments: residues 455−463, 463−472, and 473−476, validating thus Mapitope cluster A.

We chose this system as it posed a stringent test to LocaPep which gives the opportunity to explore further our approach. We used the structure of the CoV spike protein in that complex (PDB ID 2AJF) and the two peptide panels of Tarnovitski et al.[18] to localize the receptor-binding site. Not enough clusters were found with standard KMin = 0 runs. Since applied to 15- or 16-mer peptides this default choice would validate only clusters with at least 9 residues localized, we decrease this value to 8 and performed new runs. Using standard KSea = 0 option to select equivalence classes, LocaPep identified for both panels a site composed of residues spanning positions 323, 325, 348−356, 366−375, 382, 499, and 501. This site agrees with a combined prediction of Mapitope cluster C and the larger part of cluster B but not with the genuine site. After analyzing the amino acid distribution of CoV spike and noting the anomalously great number of aromatic residues, we removed class 3 in the default selection of equivalence classes and performed new KSea = 4 runs keeping the classes that were selected before except number 3. LocaPep identified then for panel 1 a site composed of 15 residues at positions 443, 445, 460−466, 468, 470−473, and 475 (Figure 2G). For panel 2, a 13-residue site which matches panel 1 site except positions 471 and 472 was predicted. If the default KSea option had been utilized, the program obviously had failed to localize the correct epitope. However, the information available for this especial case[18] served to illustrate a further use of LocaPep. In identifying the role played by equivalence classes in mapping a large set of mimotopes, it gave valuable hints on the origin of the existence of two distinct antigen regions predicted in silico by two completely different algorithms

3.2.8. Peach Pru p 3 Allergen. Pru p 3, the major peach allergen, belongs to the lipid transfer protein (LTP) panallergen family, widely distributed throughout the plant kingdom. LTPs are not only important allergens in plant foods but they are also involved in some pollinosis and cross-reactivity between pollens and plant foods.[45] Pru p 3 is the main food allergen involved in allergic reactions in the adult population from the Mediterranean area,[46] and it has been chosen as a suitable model to undertake identification of IgE epitopes in plant foods. Following a combined approach based on structural analyses of allergens, site-directed mutagenesis, synthetic peptides, and mimotope mapping, a potential IgE-binding site composed of residues 35−37, 39, 40, 42−44, 46, 70 and 76−79 has been identified in Pru p 3.[47,48] A crystal structure is available for the peach allergen[49] but no structure for its IgE complex exists. As a final test, we used this structure (PDB ID 2B5S) and a 17-peptide set from the mimotopes selected using Pru p 3-specific IgE.[48] LocaPep identified an epitope composed of 15 residues: 1, 35, 37, 39, 43, 44, 71, 75−79, 81, 85, and 91 (Figure 2H). Eleven residues are within or immediately close to the experimental epitope.

4. DISCUSSION

The selection of peptides that bind with high affinity specifically an antibody of interest from a pool of random peptides has become a technique increasingly used to map epitopes. The collection of selected peptides is expected to mimic the whole or part of the true epitope. However, inferring the location of epitope residues on an antigen from mimotope sequences is far from trivial because no clear homology may exist between these sequences and the epitope. This imitation is supposed to rely on

similarities in spatial organization, surface characteristics, and physicochemical properties, yet without knowing exactly which features count and how they do. These difficulties explain the existence of distinct approaches proposed to solve this mapping problem. We have added to this effort by developing LocaPep, a computer program to predict an unique epitope site from a collection of peptides.

A number of computational resources are available for mimotope-based epitope prediction. Findmap[14] aligns one single consensus sequence ('probe') derived from a set of mimotopes to the antigen sequence, allowing for permutations of the probe. SiteLight[15] divides the antigen surface into overlapping patches and then aligns each mimotope sequence with each patch scoring them. High-scoring patches are selected iteratively until a fourth of the surface is covered. 3DEX[16] searches for a linear peptide sequence on the protein surface allowing for identical but not for potential matches. Amino acids on the antigen are sought for one after the other in a sequential order. The program is applied for a single peptide so that results from different peptides are not combined. Mapitope[17−19] is an algorithm that deconvolutes mimotope sequences into a set of overlapping amino acids. Statistically significant pairs are then identified on overlapping pairs and mapped onto the antigen structure. These pairs are finally clustered into a set of different epitope candidates. MIMOX[21] starts deriving a single consensus sequence from the set of mimotopes. It then searches for neighbor residue pairs (determined on the basis of distances between their Cα and Cβ atoms) in the surface until all possible ways of forming the input sequence are made. MIMOP[22] integrates two different methodologies: MimAlign and MimCons or a combination of them. MimAlign employs several alignment methods to align the set of peptides to the antigen sequence. Positions in protein sequence best aligned to peptide sequences are grouped according to their 3D distance to define candidate epitopes. MimCons identifies consensus patterns within peptide sequences, and the surface is then scanned to localize exposed regions that encompass residues similar to consensus patterns. The antigen surface is considered only following alignments, and each peptide is weighted equally. PepSurf algorithm,[20] implemented together with Mapitope in the Pepitope server,[23] aligns each peptide to a graph which represents the antigen surface. In this graph, vertices represent exposed residues and edges connected residues. Residue similarities are scored using a substitution matrix, but unmatched peptide amino acids are allowed by combining gap costs in the alignment (the substitution matrix and gap costs can be adjusted by the user). Each aligned peptide corresponds to a path of residues on the protein surface similar to the input peptide.

These tools employ a wide range of methodological approaches to solve the mapping problem. However, an user interested in having a single epitope prediction from a set of mimotopes often finds that they provide several alternatives. These alternatives can correspond to different regions on the antigen surface, and although information for making choices is usually given, this may be an inconvenience in managing the output. Some ot these methods offer no flexibility to identify characteristics of a particular peptide within the whole set of mimotopes or to explore in different ways this set. The prominent methodological weight given to alignments in many of these tools can lead them to fail to localize surface regions of potential interest that align poorly with a particular sequence (consensus or not). Criteria used to identify neighbor residues relies usually on interatomic distances in space instead of on the surface, which may result in false connections in the antigen.

We have tried to overcome some of these weaknesses by developing LocaPep. The clusters algorithm devised to reach a workable solution to the mimotope mapping problem is based on the idea that some exposed residues can act as 'seeds' around which patches of amino acids ('clusters') that are mimicked by the mimotopes are formed in the surface. Upon comparing properties computed for each residue in a cluster (surface area, exposure, distance to seed and contact number) with maximum, minimum, and mean values of its amino acid type in the protein, the program assigns an individual score to each residue that is equivalent to an amino acid in the peptide. Given that this scoring is composed of four different terms, it is flexible enough to rank candidate residues on an individual basis. Repeating the search for clusters over all the possible protein residues that may act as seeds for relevant classes of equivalence between protein and peptide amino acids, the sum of residue scores is used to rank clusters. Insofar as this algorithm is applied to each peptide in the input set, the program permits to assess separately their suitability and identify thus 'better' and 'worse' peptides. The consensus selection that yields the final prediction is made on an individual, residue-based basis which allows to assign a suitability score to each of the putative protein amino acids in the predicted epitope. Since this selection is constructed with score-selected candidate residues within a reduced domain around the best seed found for the whole set of mimotopes, the final prediction is a single surface patch composed of a number of amino acids that is also given by the program.

LocaPep relies mainly on surface features with very little resort to sequence information. Because of the characteristics of the algorithm devised, the program provides an unique prediction that, by construction, is composed of close residues forming a single patch in the antigen surface. Moreover, since these consensus residues are ranked with a final normalized score, the user has information on the relative suitability of each residue in the predicted epitope. No difference between using one single peptide or a large collection of peptides exists other than the smaller or greater number of candidate residues found and the concomitant lower or higher reliability associated to total scores in each case. With vey few input parameters that control in an extremely simple way distinct options, the program permits to explore different scans of the antigen surface in the search for clusters. It must be mentioned that our algorithm bears resemblance to ProBiS,[50,51] a method intended to detect similar protein binding sites on the basis of local structural alignments developed in a different context of protein−protein interactions studies and recently implemented as a web server.[52]

Summarizing, the overall reliability shown by the varied sample of results presented as well as the features mentioned make LocaPep a computational tool of interest to obtain workable solutions to the prediction of epitope sites upon mapping mimotopes onto antigen surfaces. Given the methodological differences with respect to other existing tools, the information provided by LocaPep may offer a valuable complement and help to widen our understanding of the mapping problem.

AUTHOR INFORMATION

Corresponding Author

*E-mail: luis.fpacios@upm.es.

■ ACKNOWLEDGMENT

This study was supported with funding from the Spanish Ministerio de Ciencia e Innovación, Secretaría de Estado de Investigación, Project BIO2009-07050.

■ REFERENCES

(1) Blythe, M. J.; Flower, D. R. Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci.* **2005**, *14*, 246–248.

(2) Greenbaum, J. A.; Andersen, P. H.; Blythe, M.; Bui, H. H.; Cachau, R. E.; Crowe, J.; Davies, M.; Kolaskar, A. S.; Lund, O.; Morrison, S.; Mumey, B.; Ofran, Y.; Pellequer, J. L.; Pinilla, C.; Ponomarenko, J. V.; Raghava, G. P. S.; van Regenmortel, M. H. V.; Roggen, E. L.; Sette, A.; Schlessinger, A.; Sollner, J.; Zand, M.; Peters, B. Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J. Mol. Recognit.* **2007**, *20*, 75–82.

(3) Roggen, E. L. B-cell epitope engineering: A matter of recognizing protein features and motives. *Drug Discovery Today: Technol.* **2008**, *5*, e49–e55.

(4) Salimi, N.; Fleri, W.; Peters, B.; Sette, A. Design and utilization of epitope-based databases and predictive tools. *Immunogenetics* **2010**, *62*, 185–196.

(5) Batori, V.; Friis, E. P.; Nielsen, H.; Roggen, E. L. An in silico method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context. *J. Mol. Recognit.* **2006**, *19*, 21–29.

(6) Van Regenmortel, M. H. V. Mapping epitope structure and activity: From one-dimensional prediction to four-dimensional description of antigenic specificity. *Methods* **1996**, *9*, 465–472.

(7) Soga, S.; Kuroda, D.; Shirai, H.; Kobori, M.; Hirayama, N. Use of amino acid composition to predict epitope residues of individual antiboides. *Protein Eng., Des. Sel.* **2010**, *23*, 441–448.

(8) Scott, J. K.; Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **1990**, *249*, 386–390.

(9) Smith, G. P.; Petrenko, V. A. Phage display. *Chem. Rev.* **1997**, *97*, 391–410.

(10) Cwirla, S. E.; Peters, E. A.; Barrett, R. W.; Dower, W. J. Peptides on phage: a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6378–6382.

(11) Cortese, R.; Felici, F.; Galfre, G.; Luzzago, A.; Monaci, P.; Nicosia, A. E. Epitope discovery using peptide libraries displayed on phage. *Trends Biotechnol.* **1994**, *12*, 262–267.

(12) Geysen, H. M.; Rodda, S. J.; Mason, T. J. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol.* **1986**, *23*, 709–715.

(13) Pizzi, E.; Cortese, R.; Tramontano, A. Mapping epitopes on protein surfaces. *Biopolymers* **1995**, *36*, 675–680.

(14) Mumey, B. M.; Bailey, B. W.; Kirkpatrick, B.; Jesaitis, A. J.; Angel, T.; Dratz, E. A. A new method for mapping discontinuous antibody epitopes to reveal structural features of proteins. *J. Comput. Biol.* **2003**, *10*, 555–567.

(15) Halperin, I.; Wolfson, H.; Nussinov, R. SiteLight: binding-site prediction using phage display libraries. *Protein Sci.* **2003**, *12*, 1344–1359.

(16) Schreiber, A.; Humbert, M.; Benz, A.; Dietrich, U. 3D-Epitope-Explorer (3DEX): Localization of conformational epitopes within three-dimensional structures of proteins. *J. Comput. Chem.* **2005**, *26*, 879–887.

(17) Enshell-Seijffers, D.; Denisov, D.; Groisman, B.; Smelyanski, L.; Meyuhas, R.; Gross, G.; Denisova, G.; Gershoni, J. M. The mapping and reconstitution of a conformational discontinuous B-cell epitope of HIV-1. *J. Mol. Biol.* **2003**, *334*, 87–101.

(18) Tarnovitski, N.; Matthews, L. J.; Sui, J.; Gershoni, J. M.; Marasco, W. A. Mapping a neutralizing epitope on the SARS coronavirus spike protein: Computational prediction based on affinity-selected peptides. *J. Mol. Biol.* **2006**, *359*, 190–201.

(19) Bublil, E. M.; Tarnovitski, N.; Mayrose, I.; Penn, O.; Roitburd-Berman, A.; Rubinstein, N. D.; Pupko, T.; Gershoni, J. M. Stepwise prediction of conformational discontinuous B-cell epitopes using the Mapitope algorithm. *Proteins* **2007**, *68*, 294–304.

(20) Mayrose, I.; Shlomi, T.; Rubinstein, N. D.; Gershoni, J. M.; Ruppin, E.; Sharan, R.; Pupko, T. Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. *Nucleic Acids Res.* **2007**, *35*, 69–78.

(21) Huang, J.; Gutteridge, A.; Honda, W.; Kanehisa, M. MIMOX: a web tool for phage display based epitope mapping. *BMC Bioinformatics* **2006**, *7*, 451–460.

(22) Moreau, V.; Granier, C.; Villard, S.; Laune, D.; Molina, F. Discontinuous epitope prediction based on mimotope analysis. *Bioinformatics* **2006**, *22*, 1088–1095.

(23) Mayrose, I.; Penn, O.; Erez, E.; Rubinstein, N. D.; Shlomi, T.; Tarnovitski, N.; Bublil, E. M.; Ruppin, E.; Sharan, R.; Gershoni, J. M.; Martz, E.; Pupko, T. Pepitope: epitope from affinity-selected peptides. *Bioinformatics* **2007**, *23*, 3244–3246.

(24) Pascual-Ahuir, J. L.; Silla, E.; Tuñon, I. GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface. *J. Comput. Chem.* **1994**, *15*, 1127–1138.

(25) Pacios, L. F. ARVOMOL/CONTOUR: Molecular surface areas and volumes on personal computers. *Comput. Chem.* **1994**, *18*, 377–385.

(26) Pacios, L. F. Distinct molecular surfaces and hydrophobicity of amino acid residues in proteins. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1427–1435.

(27) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. Hydrophobicity of amino acid residues in globular proteins. *Science* **1985**, *229*, 834–838.

(28) Chakrabarti, P.; Janin, J. Dissecting protein-protein recognition sites. *Proteins* **2002**, *47*, 334–343.

(29) Andersen, P. H.; Nielsen, M.; Lund, O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.* **2008**, *15*, 2558–2567.

(30) Muller, Y. A.; Chen, Y.; Christinger, H. W.; Li, B.; Cunningham, B. C.; Lowman, H. B.; de Vos, A. M. VEGF and the Fab fragment of a humanized neutralizing antibody: Crystal complex at 2.4 Å resolution and analysis of the interface. *Structure* **1998**, *6*, 1153–1167.

(31) Chen, Y.; Wiesmann, C.; Fuh, G.; Christinger, H. W.; McKay, P.; de Vos, A. M. Selection and analysis of an optimized anti-VEGF antibody: Crystal structure of an affinity-matured Fab in complex with antigen. *J. Mol. Biol.* **1999**, *293*, 865–881.

(32) Berthet-Colominas, C.; Monaco, S.; Novelli, A.; Sibai, G.; Mallet, F.; Cusack, S. Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid protein (p24) complexed with a monoclonal antibody Fab. *EMBO J.* **1999**, *18*, 1124–1136.

(33) Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **1998**, *393*, 648–659.

(34) Gazarian, T. G.; Selisko, B.; Gurrola, G. B.; Hernandez, R.; Possani, L. D.; Gazarian, K. G. Potential of peptides selected from random phage-displayed libraries to mimic conformational epitopes: A study on scorpion toxin Cn2 and the neutralizing monoclonal antibody BCF2. *Comb. Chem. High Throughput Screening* **2003**, *6*, 119–132.

(35) Pintar, A.; Possani, L. D.; Delepierre, M. Solution structure of toxin 2 from *Centruroides noxius* Hoffmann, a β-scorpion neurotoxin acting on sodium channels. *J. Mol. Biol.* **1999**, *287*, 359–367.

(36) Calderon-Aranda, E. S.; Selisko, B.; York, E. J.; Gurrola, G. B.; Stewart, J. M.; Possani, L. D. Mapping of an epitope recognized by a neutralizing monoclonal antibody specific to toxin Cn2 from the scorpion *Centruroides noxius*, using discontinuous synthetic peptides. *Eur. J. Biochem.* **1999**, *264*, 746–755.

(37) Fendly, B. M.; Winget, M.; Hudziak, R. M.; Lipari, M. T.; Napier, M. A.; Ullrich, A. Characterization of murine monoclonal antibodies reactive to either the human epidermal growth factor receptor or HER2/neu gene product. *Cancer Res.* **1990**, *50*, 1550–1558.

(38) Cho, H. S.; Mason, K.; Ramyar, K. X.; Stanley, A. M.; Gabelli, S. B.; Denney, D. W.; Leahy, D. J. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **2003**, *421*, 756–760.

(39) Riemer, A. B.; Kraml, G.; Schenier, O.; Zielinski, C. C.; Jensen-Jarolina, E. Matching of trastuzumab (Herceptin®) epitope mimics onto the surface of Her-2/neu - A new method of epitope definition. *Mol. Immunol.* **2005**, *42*, 1121–1124.

(40) Spiegel, P. C.; Jacquemin, M.; Saint-Remy, J. M.; Stoddard, B. L.; Pratt, K. P. Structure of a factor VIII C_2 domain-immunoglobulin $G4_\kappa$ Fab complex: identification of an inhibitory antibody epitope on the surface of factor VIII. *Blood* **2001**, *98*, 13–19.

(41) Villard, S.; Lacroix-Desmazes, S.; Kieber-Emmons, T.; Piquer, D.; Grailly, S.; Benhida, A.; Kaveri, S. V.; Saint-Remy, J. M.; Granier, C. Peptide decoys selected by phage display block in vitro and in vivo activity of a human anti-FVIII inhibitor. *Blood* **2003**, *103*, 949–952.

(42) Peiris, J. S. M.; Lai, S. T.; Poon, L. L.; Guan, Y.; Yam, L. Y.; Lim, W.; Nicholls, J.; Yee, W. K.; Yan, W. W.; Cheung, M. T.; Chan, K. H.; Tsang, D. N.; Yung, R. W.; Ng, T. K.; Yuen, K. Y. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **2003**, *361*, 1319–1325.

(43) Sui, J.; Wenhui, L.; Murakami, A.; Tamin, A.; Matthews, L. J.; Wong, S. K.; Moore, M. J.; Tallarico, A. S. C.; Olurinde, M.; Choe, H.; Anderson, L. J.; Bellini, W. J.; Farzan, M.; Marasco, W. A. Potent neutralization of SARS coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2536–2541.

(44) Li, F.; Li, W.; Farzan, M.; Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **2005**, *309*, 1864–1868.

(45) Salcedo, G.; Sanchez-Monge, R.; Barber, D.; Diaz-Perales, A. Plant nonspecific lipid transfer proteins: an interface between plant defence and human allergy. *Biochim. Biophys. Acta* **2007**, *1771*, 781–791.

(46) Cuesta-Herranz, J.; Lazaro, M.; Figueredo, E.; Igea, J. M.; Umpierrez, A.; De-Las-Heras, M. Allergy to plant-derived fresh foods in a birch- and ragweed-free area. *Clin. Exp. Allergy* **2000**, *30*, 1411–1416.

(47) Garcia-Casado, G.; Pacios, L. F.; Diaz-Peralez, A.; Sanchez-Monge, R.; Lombardero, M.; Garcia-Selles, F. J.; Polo, F.; Barber, D.; Salcedo, G. Identification of IgE-binding epitopes of the major peach allergen Pru p 3. *J. Allergy Clin. Immunol.* **2003**, *112*, 599–605.

(48) Pacios, L. F.; Tordesillas, L.; Cuesta-Herranz, J.; Compes, E.; Sanchez-Monge, R.; Palacin, A.; Salcedo, G.; Diaz-Perales, A. Mimotope mapping as a complementary strategy to define allergen IgE-epitopes: Peach Pru p 3 allergen as a model. *Mol. Immunol.* **2008**, *45*, 2269–2276.

(49) Pasquato, N.; Berni, R.; Folli, C.; Folloni, S.; Cianci, M.; Pantano, S.; Helliwell, J. R.; Zanotti, G. Crystal structure of peach Pru p 3, the prototypic member of the family of plant non-specific lipid transfer protein pan-allergens. *J. Mol. Biol.* **2006**, *356*, 684–694.

(50) Konc, J.; Janezic, D. Protein-protein binding-sites prediction by protein surface structure conservation. *J. Chem. Inf. Model.* **2007**, *47*, 940–944.

(51) Konc, J.; Janezic, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26*, 1160–1168.

(52) Konc, J.; Janezic, D. ProBiS: a web server for detection of structurally similar protein binding sites. *Nucleic Acids Res.* **2010**, *38*, W436–W440.