

Research

Open Access

The MEPS server for identifying protein conformational epitopes

Tiziana Castrignanò¹, Paolo D'Onorio De Meo¹, Danilo Carrabino¹,
Massimiliano Orsini², Matteo Floris² and Anna Tramontano^{*2,3,4}

Address: ¹CASPUR, Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca, Via dei Tizii, 6/b, 5 I-00185 Rome, Italy, ²Center for Advanced Studies, Research and Development in Sardinia (CRS4), Bioinformatics Unit, Parco Scientifico e Tecnologico, POLARIS, Edificio 1, 09010 PULA (CA), Italy, ³Department of Biochemical Sciences, University 'La Sapienza', P.le Aldo Moro, 5, I-00185 Rome, Italy and ⁴Istituto Pasteur – Fondazione Cenci Bolognetti, University 'La Sapienza', P.le Aldo Moro, 5, I-00185 Rome, Italy

Email: Tiziana Castrignanò - Tiziana.Castrignano@caspur.it; Paolo D'Onorio De Meo - donorio@caspur.it;
Danilo Carrabino - carrabino@caspur.it; Massimiliano Orsini - orsini@crs4.it; Matteo Floris - floris@crs4.it;
Anna Tramontano* - anna.tramontano@uniroma1.it

* Corresponding author

from Italian Society of Bioinformatics (BITS): Annual Meeting 2006
Bologna, Italy. 28–29 April, 2006

Published: 8 March 2007

BMC Bioinformatics 2007, 8(Suppl 1):S6 doi:10.1186/1471-2105-8-S1-S6

This article is available from: <http://www.biomedcentral.com/1471-2105/8/S1/S6>

© 2007 Castrignanò et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: One of the most interesting problems in molecular immunology is epitope mapping, i.e. the identification of the regions of interaction between an antigen and an antibody. The solution to this problem, even if approximate, would help in designing experiments to precisely map the residues involved in the interaction and could be instrumental both in designing peptides able to mimic the interacting surface of the antigen and in understanding where immunologically important regions are located in its three-dimensional structure. From an experimental point of view, both genetically encoded and chemically synthesised peptide libraries can be used to identify sequences recognized by a given antibody. The problem then arises of which region of a folded protein the selected peptides correspond to.

Results: We have developed a method able to find the surface region of a protein that can be effectively mimicked by a peptide, given the structure of the protein and the maximum number of side chains deemed to be required for recognition. The method is implemented as a publicly available server. It can also find and report all peptide sequences of a specified length that can mimic the surface of a given protein and store them in a database.

The immediate application of the server is the mapping of antibody epitopes, however the system is sufficiently flexible for allowing other questions to be asked, for example one can compare the peptides representing the surface of two proteins known to interact with the same macromolecule to find which is the most likely interacting region.

Conclusion: We believe that the MEPS server, available at <http://www.caspur.it/meps>, will be a useful tool for immunologists and structural and computational biologists. We plan to use it ourselves to implement a database of "surface mimicking peptides" for all proteins of known structure and proteins that can be reliably modelled by comparative modelling.

Background

Antibodies recognize their cognate antigen by establishing energetically favourable interaction with its amino acids [1]. In some cases, the recognition site is represented by a continuous segment of the antigen sequence, but much more often the epitope is "conformational", i.e. the antibody recognizes the location and type of exposed antigen side chains that are not necessarily contiguous in the antigen's sequence, but brought together by its three-dimensional structure [2]. Identifying the location of the epitope on the protein, i.e. mapping it, is of paramount importance because it is instrumental for the development of diagnostic or biotechnological tools and of recombinant vaccines [3-7].

Epitope mapping can be achieved by experimental techniques. If the target protein is known and cloned, one can fragment or sub-clone the protein and identify which region is still recognized by the antibody [3]. Clearly only continuous epitopes, or epitopes that contain at least one reasonably long stretch of contiguous target amino acids, can be identified with this strategy.

The correct identification of a conformational epitope can be obtained by structural determination of the antibody antigen complex, a time and labour consuming procedure that cannot be guaranteed to succeed in all cases.

However, molecular biology and peptide chemistry provide two powerful tools for mapping the epitopes of a protein, given the recognizing antibody: combinatorial libraries.

Both genetically encoded and chemically synthesised peptide libraries have been used with great success in this area [8,9]. The great diversity contained in a library allows the selection of peptides able to bind to an antibody. The analysis of their sequences can allow the location of linear epitopes to be immediately identified by just comparing the sequences of the peptides with that of the antigen [10].

It is worth emphasizing that the library selection strategy does not require the knowledge of the antigen, but only the availability of an antibody, thereby it can also be used for identifying unknown agents provided antibodies against them can be elicited or found. Furthermore, although antibodies are the most commonly used selectors in combinatorial library experiments, because they can be obtained from patient sera or by immunizing experimental animals, it should be apparent that the technique can be, and is, applied using any receptor as selector molecule.

The experimental strategy does not necessarily limit the results to peptides mimicking a continuous epitope, since peptides able to mimic the discontinuous surface of a conformational epitope can be selected as well. In these cases, though, a straightforward sequence comparison is not sufficient to identify the original epitope on the target protein even more so if the task is not only the identification of the epitope on a given protein, but also the identification of the recognized protein.

The problem in this case is to find a surface region of a protein that can be effectively mimicked by one or more selected peptides, given the structure of the protein and the maximum number of side chains deemed to be required for recognition.

A method for solving the problem has been previously described by some of us [11], but at the time technology and computer power was not adequate to implement the method as a publicly available server and for extending its use to a large number of proteins, and therefore it could only be used locally for the detection of epitopes of a known antigen and it was not suitable for the identification of unknown antigens.

The MEPS server, described below, has been designed to fulfil these requirements.

Results

Generating the surface ensemble

We define here a surface ensemble as the collection of all peptides of a given length L that can position their side chains in such a way that at least m ($1 < m < L$) of their side chains are able to mimic exposed regions of the protein surface.

Given the structure of a target protein, we first select all solvent exposed amino acids. In the current implementation the threshold for minimum solvent accessible surface is set to 40 \AA^2 . Next, we compute the distance between the $C\beta$ ($C\alpha$ for glycines) of each pair of exposed amino acids and store them in a matrix. The choice of using $C\beta$ derives from the need of taking into account the direction of the side chain.

The matrix is used to build a graph where each node represents a surface amino acid, and an edge connects two nodes if their distance is lower than a maximum distance threshold d . The graph is represented as a collection of adjacency lists: there is a list for each amino acid and each list contains a pair $[\textit{neighbour}, \textit{neigh_distance}]$ if $\textit{neigh_distance}$ is lower than d .

Table 1: Searching pattern with mismatches. Example of patterns with up to three mismatches for the input peptide AYKTK

mismatches	1	2	3
Patterns	*YKTK A*KTK AY*TK AYK*K AYKT*	**KTK *Y*TK *YK*K *YKT* A**TK A*K*K A*K*T* AY**K AY*T* AYK**	***TK **K*K **KT* *Y**K *Y*T* *YK** A***K A**T* A*K** AY***

Retrieving all putative epitopes for a target or a set of target proteins

The peptide ensemble graph is processed to extract the sequences of all possible peptides able to mimic the surface of a given protein, or of a set of protein structures. For each adjacency list, we search all neighbours within a 5.5 Å distance. The search is recursively repeated for each of the identified neighbours until its length reaches the value *l* selected by the user. The peptide sequences can be inter-actively inspected and/or stored in a file in FASTA format. This allows sequence similarity searching tools such as BLAST or FASTA to be directly used on the surface ensemble peptides.

The threshold is only applied to the distance between adjacent amino acids and therefore we implicitly list all peptides shorter than the maximum length selected by the user.

Searching for a conformational epitope

The user can input a protein structure (or a set of protein structures), a peptide sequence and a minimum number of mismatching residues and retrieve the location, if it exists, of the corresponding epitope.

In this case, the system first exhaustively lists all possible patterns with mismatches (Table 1). Each pattern is searched in the graph, similarly to what is done for finding exact matches. In this case, however, the distance range between two adjacent amino acids depends upon the number of intervening mismatches.

The minimum and maximum distances between C β (C α for glycines) of amino acids separated by 1, 2, 3, etc amino acids have been obtained by a statistical analysis of well resolved structures in the protein structure databank (PDB) [12], consequently they are not only chemically possible, but also stereochemically plausible.

User interface

Figure 1 shows the entry page of the MEPS server and the page displayed upon selecting the "Search" button. The user can select the search mode, or retrieve the result of a previously run job.

If the "pattern with mismatch" option is selected, several possibilities are offered: the user can upload a local PDB file, a file from the PDB database, or a set of PDB files.

In all cases, the user is asked for the searched pattern and the maximum number of mismatches. In the example shown, the user has selected the 1ECF protein structure and the pattern "DIVMAAQ" allowing two mismatches. The protein contains two overlapping sets of residues that can mimic the input peptide (the last glutamine can mimic residue 339 or 340). The user can visualize the location of the peptide on the protein structure via Ras-mol, by clicking on the appropriate link.

If the "all fixed-size patterns" option is selected, the user only needs to select the target protein(s) and choose the length of the peptides. In the case shown in the figure, the 3MBN entry of the PDB database is selected. The list of peptides (length 5 in the example) is shown in a new window, together with the rasmol links. They can also be stored as a fasta file that can be used for further analyses. The header of each of the sequences in the fasta file lists the position in the structure of the residues of the peptides.

Discussion

The MEPS server has been designed to answer the question of which regions of a known or modelled protein structure can be mimicked by a peptide.

Although our main interest in developing the server was to enable user to map antibody epitopes, it is clear that the

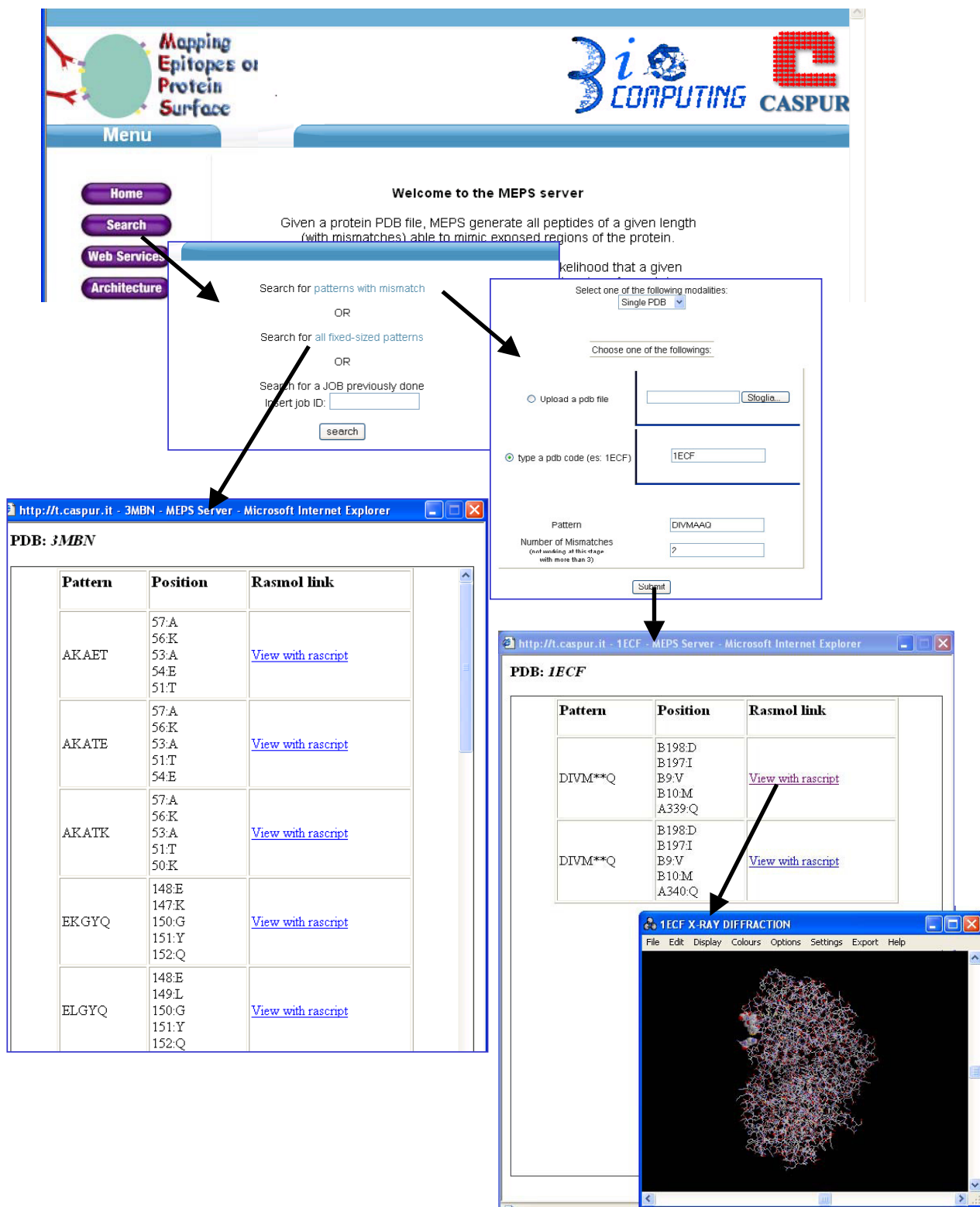


Figure 1
Using MEPS. The Figure shows a search for a pattern with mismatches (DIVMAAQ) in the 1ECF protein structure and the results of a search for all fixed size patterns of length 5 able to mimic the surface of the 3MBN protein structure.

method can be used for many other purposes. For example, for designing peptides able to mimic a portion of two interacting proteins, or for identifying common surface patches among two or more proteins.

Conclusion

New functionalities to the basic structure of the server described are already being implemented. These include the possibility of selecting the target proteins on the basis of a keyword search and of comparing the surface ensemble of two different proteins.

Methods

Implementation details

The MEPS web server is freely available [13]. It is implemented on a 4-processor Linux server (HP DL585 with SUSE SLESS 9) running the apache web server version 2.0 [14] and has been developed as a Service-Oriented Architecture (SOA).

The MEPS interface developed in PHP [15] allows the user to submit one or more protein structures in PDB format. Upon submission, PHP scripts, running in background, call three web services through the Simple Object Access Protocol (SOAP) open standard. When the job is complete, the PHP interface shows a table of results. Each pattern can be displayed in the context of the protein structure using a standalone visualization tool (for windows and linux systems) based on the 32 bit version of the open source protein visualization tool Rasmol® (1994 Roger Sayle) (available through the server). WSs have been implemented in java enterprise 1.5; all WSs details are described in [13].

Distance thresholds

The minimum and maximum distances between C β (C α for glycines) of amino acids separated by 1, 2, 3, etc amino acids have been obtained by a statistical analysis of well resolved structures in the protein structure databank (PDB). The computation was performed using a script based on BioPython classes. The data set included single chain protein structures in PDB with a resolution of 2.0 Å or better.

Authors' contributions

PDD and DC contributed to writing the computer code. MO and MF analyzed the PDB data base entries for deriving parameters and extensively tested the server. TC and AT developed the methodology. All authors contributed to writing the manuscript.

Acknowledgements

This project is partially supported by BioSapiens Network of Excellence, contract number LHSCT-203-503265, by the Istituto Pasteur- Fondazione Cenci Bolognetti and by AIRC.

The authors are grateful to Dr. Elisabetta Pizzi for very useful discussions.

This article has been published as part of *BMC Bioinformatics* Volume 8, Supplement 1, 2007: Italian Society of Bioinformatics (BITS): Annual Meeting 2006. The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2105/8?issue=S1>.

References

1. Morea V, Tramontano A, Rustici M, Chothia C, Lesk A: **Antibody structure, prediction and redesign.** *Biophys Chem* 1997, **68**:9-16.
2. Helmer_Citterich M, Rovida E, Luzzago A, Tramontano A: **Modeling antibody-antigen interactions: ferritin as a case study.** *Molecular Immunology* 1995, **32**:1001-10.
3. Luzzago A, Felici F, Tramontano A, Pessi A, Cortese R: **Mimicking of discontinuous epitopes by phage-displayed peptides, I. Epitope mapping of human H ferritin using a phage library of constrained peptides.** *Gene* 1993, **128**:51-7.
4. Mumeby BM, Bailey BV, Kirkpatrick B, Jesaitis AJ, Angel T, Dratz EA: **A New Method for Mapping Discontinuous Antibody Epitopes to Reveal Structural Features of Proteins.** *Journal of Computational Biology* 2003, **10**:555-567.
5. Moreau V, Granier C, Villard S, Laune D, Molina F: **Discontinuous epitope prediction based on mimotope analysis 10.1093/bioinformatics/btl012.** *Bioinformatics* 2006, **22**:1088-1095.
6. 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.
7. Enshell-Seijffers D, Denisov D, Groisman B, Smelyanski L, Meyuhar R, Gross G, Denisova G, Gershoni JM, U-h w s c s a B W-W F T-B a c d d d a c a: **The Mapping and Reconstitution of a Conformational Discontinuous B-cell Epitope of HIV-1.** *Journal of Molecular Biology* 2003, **334**:87-101.
8. Jayawickreme CK, Graminski GF, Quillan JM, Lerner MR: **Creation and functional screening of a multi-use peptide library.** *Proc Natl Acad Sci* 1994, **91**:1614-1618.
9. Felici F, Castagnoli L, Musacchio A, Jappelli A, Cesareni C: **Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector.** *J Mol Biol* 1991, **222**:301-310.
10. Cortese R, Monaci P, Nicosia A, Luzzago A, Felici F, Galfre G, Pessi A, Tramontano A, Sollazzo M: **Identification of biologically active peptides using random libraries displayed on phage.** *Current Opinion in Biotechnology* 1995, **6**:73-80.
11. Pizzi E, Cortese R, Tramontano A: **Mapping epitopes on protein surfaces.** *Biopolymers* 1995, **36**:675-80.
12. Berman HM, Henrick K, Nakamura H: **Announcing the worldwide Protein Data Bank.** *Nature Structural Biology* 2003, **10**:980-980.
13. [<http://www.caspur.it/meps>].
14. [<http://www.apache.org>].
15. [<http://www.php.net>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

