Published online 6 June 2012 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* ***W521–W524***
doi:10.1093/nar/gks480

Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure

Vered Kunik, Shaul Ashkenazi and Yanay Ofran*

The Goodman Faculty of Life Sciences, Nanotechnology Building, Bar-Ilan University, Ramat Gan 52900, Israel

Received February 2, 2012; Revised April 26, 2012; Accepted May 8, 2012

ABSTRACT

Antibodies are capable of specifically recognizing and binding antigens. Identification of the antigen-binding site, commonly dubbed paratope, is of high importance both for medical and biological applications. To date, the identification of antigen-binding regions (ABRs) relies on tools for the identification of complementarity-determining regions (CDRs). However, we have shown that up to 22% of the residues that actually bind the antigen fall outside the traditionally defined CDRs. The Paratome web server predicts the ABRs of an antibody, given its amino acid sequence or 3D structure. It is based on a set of consensus regions derived from a structural alignment of a non-redundant set of all known antibody–antigen complexes. Given a query sequence or structure, the server identifies the regions in the query antibody that correspond to the consensus ABRs. An independent set of antibody–antigen complexes was used to test the server and it was shown to correctly identify at least 94% of the antigen-binding residues. The Paratome web server is freely available at http://www.ofranlab.org/paratome/.

INTRODUCTION

One of the most common problems in immunological research is the identification of paratopes, namely the residues within an immunoglobulin that recognize and bind the antigen (Ag). The high affinity and specificity of antibodies (Abs) to their cognate Ag, which allows them to block its activity or to mark it for destruction (1), are at the heart of immunity. They also make Abs powerful tools in numerous molecular applications in research as well as in diagnostics and therapy (2–7). Therefore, to understand immunity (and autoimmunity) and to engineer and improve Ab-based applications, one needs to first identify the molecular determinants that mediate Ag recognition and binding. However, currently there is no tool available for providing such prediction. complementarity-determining regions (CDRs) are considered a proxy for the sites that recognize and bind the Ag. CDRs are six hypervariable segments of amino acids, three on each of the light and heavy chains (8–10).

Attempts to computationally identify CDRs have been on going for >40 years (10–17). The most commonly used CDR identification methods to date are Kabat (10,15), Chothia (12,13,16) and IMGT (16). Each of these methods has devised a unique residue numbering scheme according to which it numbers the hypervariable region residues and the beginning and ending of each of the six CDRs is then determined according to certain key positions. The pressing need in this type of analysis is manifested in the citations: in 2010 alone these methods generated over 500 citations. Arguably, many of the users are not interested in the CDRs as such but rather are interested in identifying the residues that mediate Ag binding. However, in a recent analysis we have shown that CDR identification methods may miss >20% of the residues that actually bind the Ag (18). Furthermore, we have also shown that the residues that are missed by these methods include some that make crucial energetic contribution to Ag binding (18).

The Paratome web server implements an algorithm we developed for the identification of antigen-binding regions (ABRs) from the amino acid sequence or 3D structure of an Ab (18). The algorithm is based on the premise that the vast majority of antigen-binding residues lie in regions of structural consensus between Abs. These structural consensus regions form six sequence stretches along the Ab sequence, roughly corresponding to the six CDRs (18,19). The server uses the structural consensus regions within a multiple structure alignment (MSTA) of a non-redundant set of all antibody–antigen (Ab–Ag) complexes, as a reference according to which the ABRs of unannotated Abs are inferred (18). It is trained to identify binding regions for Abs that bind protein or peptide Ags. To our knowledge, Paratome is currently the only server aimed at

*To whom correspondence should be addressed. Tel: +972 3 531 9772; Fax: +972 3 7384197; Email: yanay@ofranlab.org

identifying the Ag-binding site of Abs, which can then be used as starting points for experiments, may help improve vaccine and Ab design and may serve for large scale analysis of Abs.

DESCRIPTION OF WEB SERVER

Input

The input for the Paratome web server is either an amino acid sequence or a 3D structure (or PDB id) of an Ab. 3D structures must be in PDB file format (http://www.wwpdb.org/docs.html, 23 May 2012, date last accessed). Analysis of multiple Abs is available by uploading a compressed file containing a collection of either sequences or structures. Each submission allows the analysis of up to 100 MB of sequences or structures. Processing time is typically 5–15 s per query Ab.

Output

The first analysis done by the server determines whether the input includes an Ab or a fragment thereof. If the input is not identified as such, the results page includes a link to a text files in which this result is stated and explained (e.g. no BLAST hits found, see Supplementary Data S1C). Otherwise, the results page links to two files—a text file and an HTML file. These files provide a list of the residues that make up each ABR and their location. The HTML file provides also visualization of the ABRs highlighted in the sequence of the query Ab. For sequences, ABRs location is indicated according to their sequence position within the query sequence (see Supplementary Data S1A and S1B). Figure 1 shows the HTML results file of running Paratome on the structure of anti-IL-15 (PDB id 2xqb). For 3D structures, the location of each residue within the predicted ABRs is indicated according to its residue number as it appears in the input PDB (Figure 1A). Disordered ABRs residues (i.e. residues missing from the ATOM field) are marked with an asterisk (Figure 1A). Additionally, the structure of the query Ab with its predicted ABRs highlighted is accessible using the Jmol Java applet (Figure 1B).

MATERIALS AND METHODS

ABRs identification

Paratome is based on a large MSTA of Abs, which revealed regions of structural consensus where the pattern of structural positions that bind the Ag is highly similar among all known Abs. These consensus regions, we found, cover virtually all the residues that bind the Ag. The algorithm is based on the fact that for each Ab in the MSTA we have already identified the ABRs. Thus, given a query Ab, the algorithm performs three consecutive steps: (i) A BLAST (20) search against all the sequences in the MSTA. If the query is a PDB file, the sequence is extracted from the SEQRES field (if exists). Otherwise, it is extracted from the ATOM field. (ii) Next, a pairwise alignment is performed between the query Ab and the top BLAST hit: If the input is a sequence, each framework region (FR) of the top BLAST hit Ab is locally aligned to the query Ab using the Smith–Waterman local alignment algorithm (21). For a detailed example of sequence-based FRs pairwise alignment, see Supplementary Data S2. If the input is an Ab 3D structure, the best BLAST hit Ab and the query Ab are structurally aligned using the combinatorial extension (CE) structural alignment algorithm (22). For a detailed example of Abs pairwise structural alignment, see Supplementary Data S3. (iii) ABRs identification—the ABRs of the query Ab are identified according to the boundaries of the FRs of the top BLAST hit Ab with which it was aligned in the previous step. Figure 2 summarizes the process of ABRs identification.

PERFORMANCE

The performance of Paratome has been tested on all Ab–Ag complexes that were added to the PDB after the

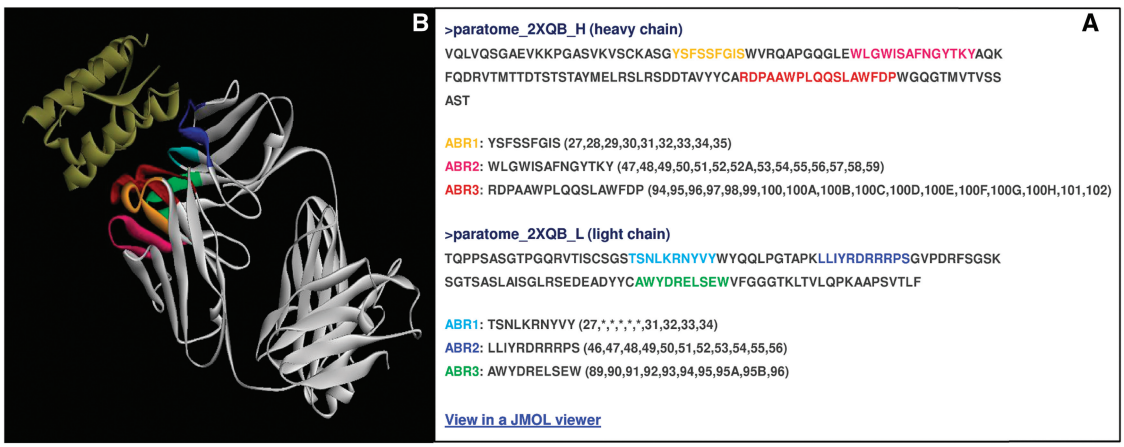

Figure 1. An example HTML results file of running Paratome on anti IL-15 (PDB id 2xqb). (**A**) The sequences of Ab chains with ABRs highlighted. The location of ABRs residues is indicated according to the numbering in the ATOM field within the input PDB file. Note that for ABR L1, the location of five residues (S, N, L, K and R) is indicated with an asterisk as they do not appear in the ATOM field within the PDB file. A link to the visualization of the analyzed complex is located below the list of identified ABRs. (**B**) The visualization of the analyzed 2qxb complex using the Jmol Java applet.

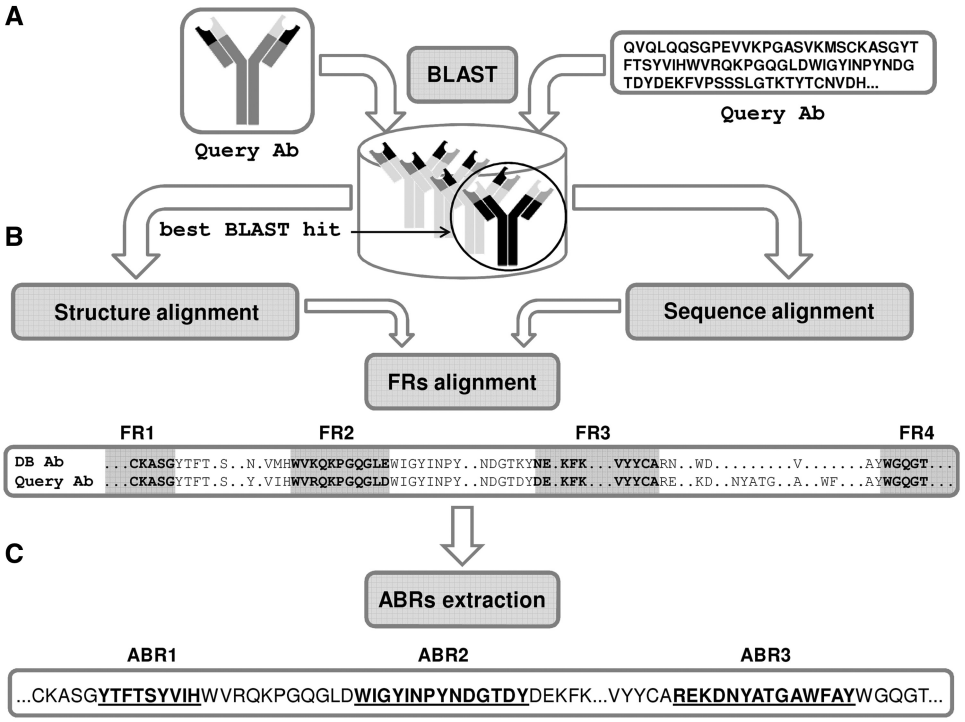

Figure 2. Antigen-binding regions identification. (**A**) The sequence of the query Ab is BLASTed against the non-redundant set of annotated Abs. (**B**) The framework regions (FRs) of the best BLAST hit Ab are aligned to the query Ab (sequences—BLAST, structures—CE). (**C**) The ABRs of the query Ab are inferred according to the ABRs of the best BLAST hit Ab.

Table 1. Comparative performance of Paratome and the most commonly used CDR identification methods

Measure/Method	Paratome (%)	Kabat (%)	Chothia (%)	IMGT (%)
Recall	94	85	79	81
Precision	42	41	44	48
Specificity	83	85	89	91

Average precision, recall and specificity were calculated for the Abs in the test set for Paratome, Kabat, Chothia and IMGT methods.

construction of the training set. For the full list of train and test sets is available at http://ofranservices.biu.ac.il/site/services/paratome/ABRs_train_pdb.txt, 23 May 2012, date last accessed http://ofranservices.biu.ac.il/site/services/paratome/ABRs_test_pdb.txt, 23 May 2012, date last accessed). A true positive (TP) prediction is a residue that is included in the ABRs and is in contact with the Ag in the experimentally determined structure of the Ab–Ag complex. A true negative (TN) is a residue that is not in the ABR and is not in contact with the Ag. A false positive (FP) is a residue that is in the ABR and not in contact with the Ag in the experimental structure and a false negative (FN) is a residue that is not in the ABR and is in contact with the Ag. The performance was assessed in terms of precision (TP/TP + FP), specificity (TN/TN + FP) and recall (TP/TP + FN). Additionally, performance was compared to that of CDRs, as identified by Kabat, Chothia and IMGT. CDRs according to Kabat and Chothia were obtained by coupling the Abnum tool (http://www.bioinf.org.uk/abs/abnum/, 23 May 2012, date last accessed) with a table of CDRs definitions (http://www.bioinf.org.uk/abs/#cdrdef, 23 May 2012, date last accessed). To obtain the CDRs according to IMGT, we applied the IMGT-gap tool (http://www.imgt.org/3Dstructure-DB/cgi/DomainGapAlign.cgi, 23 May 2012, date last accessed). An Ab amino acid and an Ag amino acid were considered as interacting if at least one of their respective atoms were ≤ 6 Å of each other (23). Table 1 summarizes the performance obtained by all methods on the test set. While the precision of all methods is roughly the same, the results show that virtually all Ag-binding residues fall within the ABRs predicted by Paratome.

IMPLEMENTATION

The server was designed and implemented using Perl, bash, Python, HTML, XML and XSL. The front end of the server is designed in HTML, XML and XSL. Structure visualization is enabled using the Jmol Java applet (http://jmol.sourceforge.net/, 23 May 2012, date last accessed). The web server and computations run on a Red-hat Enterprise Linux 5.7 (Tikanga) with 3.00 GHz 8 CPUs and 16 GB primary memory.

CONCLUSION

The specificity and affinity of Ab–Ag interactions are fundamental for understanding the biological activity of these molecules. Correct identification of the residues that mediate these interactions is crucial for immunological

research and for applications aimed at modifying and manipulating such interactions. Paratome provides a simple interface for the identification of the Ag-binding regions from the amino acid sequence or 3D structure of an Ab and has been shown to do so efficiently. Considering that the 3D structure of most known Abs is yet unknown, the ability to accurately and reliably identify the ABRs directly from sequence, is highly valuable and increases the accessibility of this essential knowledge to the entire scientific community in general and to immunity researchers in particular.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Data 1–3.

ACKNOWLEDGEMENTS

Thanks to Inbal Sela, Anat Burkovitz and Guy Nimrod for their help in testing and feedback on the server and manuscript.

FUNDING

Funding for open access charge: Research grant from the Israel Science Foundation [511].

Conflict of interest statement. None declared.

REFERENCES

1. Chan,A.C. and Carter,P.J. (2010) Therapeutic antibodies for autoimmunity and inflammation. *Nat. Rev.*, **10**, 301–316.
2. Verhoeyen,M., Milstein,C. and Winter,G. (1988) Reshaping human antibodies: grafting an antilysozyme activity. *Science*, **239**, 1534–1536.
3. Almagro,J.C. (2004) Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size: implications for the rational design of antibody repertoires. *J. Mol. Recognit.*, **17**, 132–143.
4. Lou,J. and Marks,J.D. (2010) Affinity maturation by chain shuffling and site directed mutagenesis. In: Kontermann,F. and Dübel,S. (eds), In: *Antibody Engineering,* Part IV. Springer Berlin Heidelberg, pp. 377–396.
5. Schrama,D., Reisfeld,R.A. and Becker,J.C. (2006) Antibody targeted drugs as cancer therapeutics. *Nat. Rev. Drug Discov.*, **5**, 147–159.
6. Hawkins,R.E., Russell,S.J. and Winter,G. (1992) Selection of phage antibodies by binding affinity. Mimicking affinity maturation. *J. Mol. Biol.*, **226**, 889–896.
7. Jones,P.T., Dear,P.H., Foote,J., Neuberger,M.S. and Winter,G. (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature*, **321**, 522–525.
8. MacCallum,R.M., Martin,A.C. and Thornton,J.M. (1996) Antibody-antigen interactions: contact analysis and binding site topography. *J. Mol. Biol.*, **262**, 732–745.
9. Padlan,E.A., Abergel,C. and Tipper,J.P. (1995) Identification of specificity-determining residues in antibodies. *FASEB J.*, **9**, 133–139.
10. Wu,T.T. and Kabat,E.A. (1970) An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.*, **132**, 211–250.
11. Al-Lazikani,B., Lesk,A.M. and Chothia,C. (1997) Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.*, **273**, 927–948.
12. Chothia,C. and Lesk,A.M. (1987) Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.*, **196**, 901–917.
13. Chothia,C., Lesk,A.M., Tramontano,A., Levitt,M., Smith-Gill,J.S., Air,G., Sheriff,S., Padlan,E.A., Davies,D., Tulip,W.R. *et al.* (1989) Conformations of immunoglobulin hypervariable regions. *Nature*, **342**, 877–883.
14. Kabat,E.A., Wu,T.T. and Bilofsky,H. (1977) Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *J. Biol. Chem.*, **252**, 6609–6616.
15. Kabat,E.A., Wu,T.T., Bilofsky,H., Reid-Miller,M. and Perry,H. (1983) *Sequence of Proteins of Immunological Interest*. National Institute of Health, Bathesda.
16. Lefranc,M.P., Pommie,C., Ruiz,M., Giudicelli,V., Foulquier,E., Truong,L., Thouvenin-Contet,V. and Lefranc,G. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.*, **27**, 55–77.
17. Padlan,E.A. (1977) Structural implications of sequence variability in immunoglobulins. *Proc. Natl Acad. Sci. USA*, **74**, 2551–2555.
18. Kunik,V., Peters,B. and Ofran,Y. (2012) Structural Consensus among Antibodies Defines the Antigen Binding Site. *PLoS Comput. Biol.*, **8**, e1002388.
19. Ofran,Y., Schlessinger,A. and Rost,B. (2008) Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes. *J. Immunol.*, **181**, 6230–6235.
20. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
21. Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
22. Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.
23. Ofran,Y. and Rost,B. (2003) Analysing six types of protein-protein interfaces. *J. Mol. Biol.*, **325**, 377–387.