

PIGS: a resource for immunoglobulin structure prediction

Paolo Marcatili¹, Alessandra Rosi¹ and Anna Tramontano^{1,2}

¹Department of Biochemical Sciences, Sapienza University, P.le A. Moro, 5 00185 Rome (I).

²Istituto Pasteur Fondazione Cenci Bolognetti, Sapienza University, P.le A. Moro, 5 00185 Rome (I).

Abstract

Pigs (<http://arianna.bio.uniroma1.it/pigs>) is a web server for the automatic modeling of immunoglobulin variable domains based on the canonical structure method. It has a user-friendly and flexible interface, that allows the user to choose templates (for the frameworks and the loops) and modeling strategies in an automatic or manual fashion. Its final output is a complete three-dimensional model of the target antibody that can be downloaded or displayed on-line. The server is freely accessible to academic users, with no restriction on the number of submitted sequences.

Introduction

Immunoglobulins play a central role in the immune response. Even if their structure is reasonably well conserved, the construction of reliable three-dimensional models of antibodies is still a critical point in a large number of immunological activities and studies.

Immunoglobulins are composed of two heavy and two light chains, containing respectively four and two homologous domains (Chothia et al., 1989). Antibodies recognize with a high specificity their antigens using the tip of the first domains of each chain (VL and VH). This antigen binding site is mainly composed by six loops, three from the light (L1, L2, L3) and three from the heavy chain (H1, H2, H3) named according to their order of appearance in the amino acid sequence.

It is possible to model with a good accuracy the backbone of five of these loops (L1, L2, L3, H1 and H2). It has been shown that their structure is strongly dependent on the position and identity of a few specific key amino acids (Chothia & Lesk, 1987; Chothia et al., 1989; Tramontano, Chothia, & Lesk, 1990; Webster & Rees, 1995). By their means it is often possible to identify for each of these loops a set of structures that are most similar to it. Such sets are called canonical structures, and loops belonging to the same canonical structure typically have very similar backbones (between 0.2 and 0.7 Å). Table 1 reports known canonical structures together with the set of rules used to identify them.

For the third loop of the heavy chain (H3) such precise rules do not exist. However, a careful analysis of the known H3 structures has been performed demonstrating that the conformation of the ten residues closer to the framework (called torso) can be predicted in a similar fashion and with equivalent accuracy to that of the other loops, while for the remaining no simple sequence rule has been found (Morea, Tramontano, Rustici, Chothia, & Lesk, 1998; Shirai, Kidera, & Nakamura, 1996).

Such advanced insight into sequence to structure relationship for this fundamental protein family makes it possible to predict their structure quite accurately and automatically. The crucial steps in the prediction are the correct alignment of the target sequence with those of immunoglobulins of known structure, the identification of the limits of the hypervariable loops, (where insertions and deletions occur) and of the key residues determining their conformation. Classical dynamic programming methods may return an inaccurate alignment, since insertions and deletions can only occur at very specific

positions in the sequence and some important conserved residues, for example two bonded cysteines and a tryptophan, need to be aligned. Nevertheless, rule based techniques for the alignment, the identification of the canonical structures and the detection of the appropriate templates for the loops can be implemented and automated. A review of the most common antibody modeling techniques and issues can be found in (Morea, Lesk, & Tramontano, 2000).

Loop	Canonical Structure	Loop Length	Constraints
L1	1	6	29 VIL
	2	7	29 VIL
	3	13	29 VIL
	4	12	29 VIL
	5	11	29 VIL
	6	8	29 VIL
L2	1	3	
L3	1	6	95 P, 90 HNQ
	2	6	94 P, 90 Q
	3	5	96 P, 90 Q
	4	4	90 Q
	5	7	95A P, 90 Q
	6	5	90 Q, 94 L
H1	1	7	
	2	8	
	3	9	
	4	6	
H2	1	3	71 AVL
	2	3	71 RK
	3	4	71 AVL
	4	4	71 RK
	5	6	71 AVL
	6	6	71 RK
H3	Bulged	10	94 RK
	Non-bulged	>10	94 not RK

Table 1 – Canonical structures of antibody hypervariable loops. For each loop (leftmost column), the table reports every known canonical structure, the length of loops belonging to such class and the constraints that a sequence must fulfill in order to present a loop with the corresponding canonical structure. Such constraints are represented as a list of positions (reported in the antibody-specific Kabat-Chothia numbering scheme) and residues which are allowed in such position (e.g. 29 VIL means that in position 29 only valine, isoleucine or leucine are permitted)

The Pigs web server

We have developed pigs (Prediction of ImmunoGlobulin Structure), a web-server which implements and automatise such procedures to produce a full three-dimensional model of a given target immunoglobulin. Pigs is freely accessible to all the academic users, with no registration needed and no limit to the number of submitted sequences. Pigs interface is extremely simple and flexible, in order to be easily accessed to all the users and, at the same time, it offers several tools to solve and refine even complex jobs.

The user only needs to input the sequence of the variable chains of the target antibody and the server will automatically build the correct alignment using an expert set of rules. The intermediate output of the server is a list of putative templates for both the loops and the framework for each chain, together with other useful information.

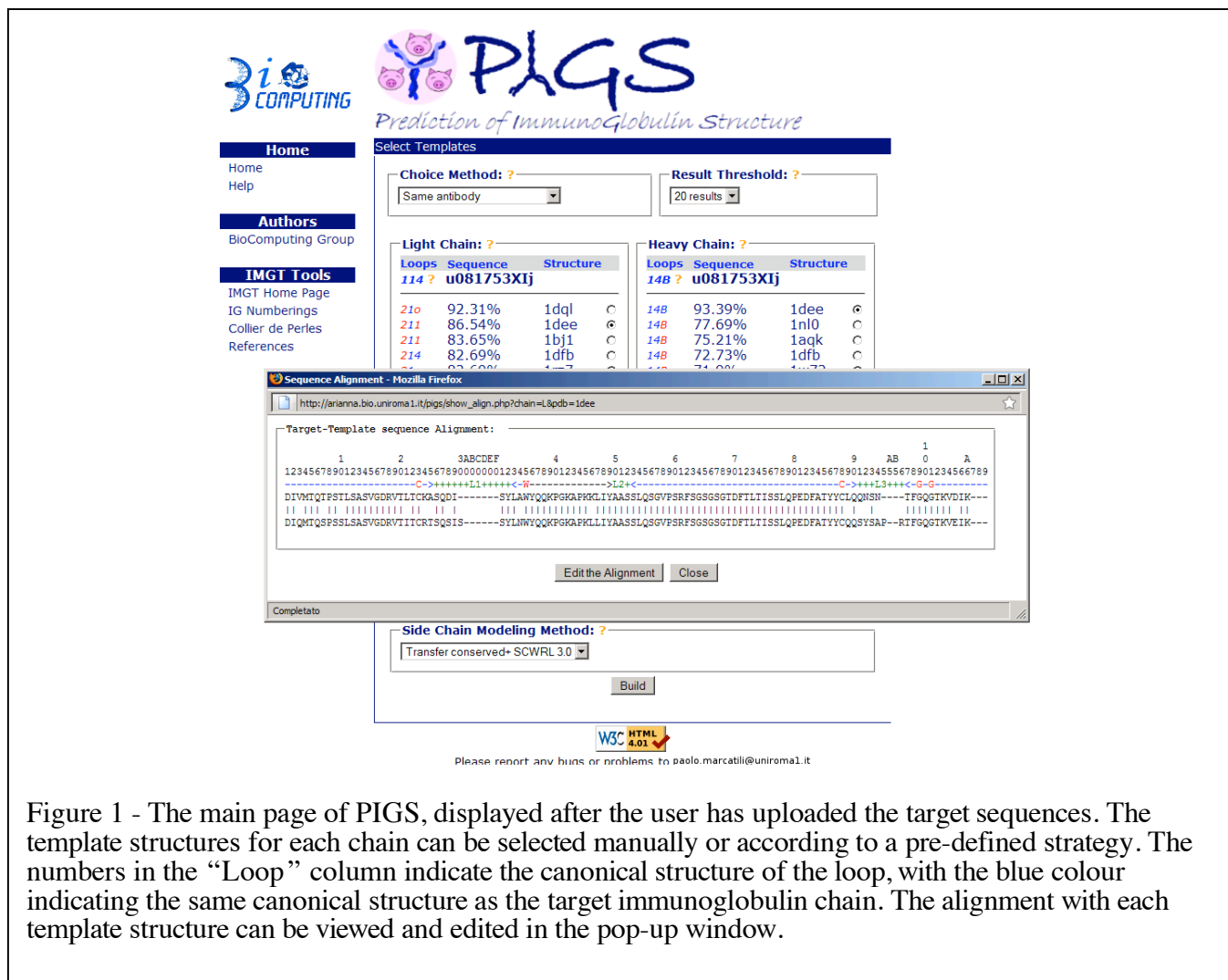


Figure 1 - The main page of PIGS, displayed after the user has uploaded the target sequences. The template structures for each chain can be selected manually or according to a pre-defined strategy. The numbers in the “Loop” column indicate the canonical structure of the loop, with the blue colour indicating the same canonical structure as the target immunoglobulin chain. The alignment with each template structure can be viewed and edited in the pop-up window.

The user manually select the templates, or choose among four different automatic modeling strategies:

- **Same antibody**: the server selects the known structure, if any, that can provide a template for both

the heavy and light chain frameworks. If some of the loops of such structure do not belong to the same canonical structure of the corresponding target ones, the loops in the database displaying a similar canonical structure and with the highest sequence identity are grafted on the template scaffold.

- **Same canonical structure:** the server select the templates, if any, whose loops display exactly the same canonical structures than the corresponding target loops. If the templates for the heavy and light chains belong to different structures, the whole complex is rearranged by means of an optimal superimposition of the most conserved residues in the interface region.
- **Same antibody and canonical structure:** the server selects the antibody structure, if any, that can be used as a template for both VL and VH and where the canonical structures of the loops are the same as those of the target. In this case no further adjustment is necessary.
- **Best L and H chain:** the server selects the two chains with the highest sequence identity to the corresponding target chains of the target and, if needed, pack the two chains together and take the loops from different structures.

It is possible to select whether the output structure should contain just the main chain coordinates, the coordinates of the main chain and of all the side chains conserved with respect to the template, or the complete structure (Figure 1). In the latter case, pigs retains the conformation of the conserved side chains and reconstructs the remaining ones using the SCWRL software (Bower, Cohen, & Dunbrack, 1997).

Once the user has selected the templates and the side chain modeling strategy, to obtain the final model he just has to click on the "build" button. In a few seconds pigs will output the model in PDB format, allowing the user to download it, to receive it by email or to display it online via Jmol (Figure 2).

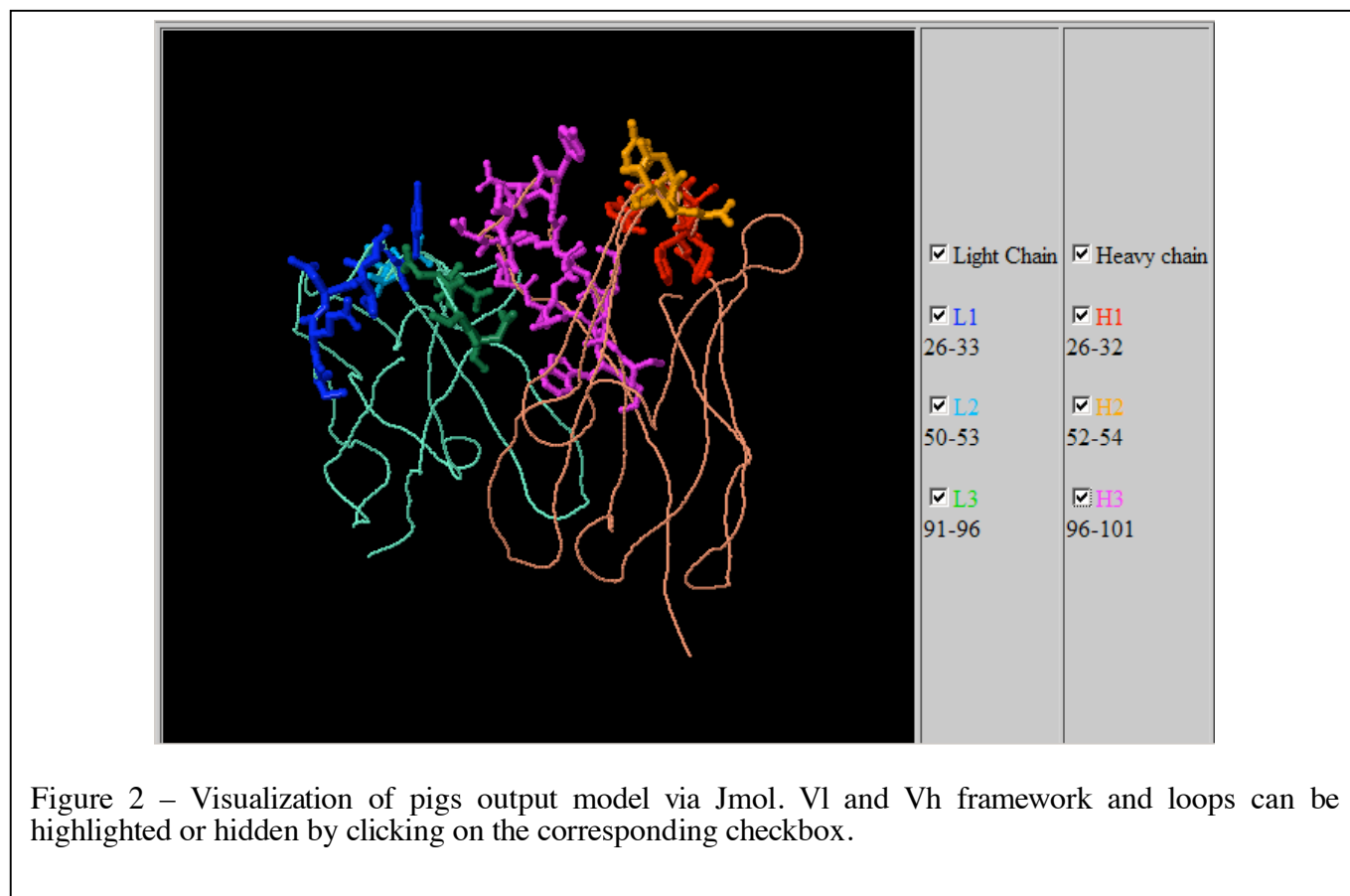


Figure 2 – Visualization of pigs output model via Jmol. VL and Vh framework and loops can be highlighted or hidden by clicking on the corresponding checkbox.

We are currently performing a set of complete tests in order to determine which parameters and modeling strategies produce the most accurate immunoglobulin models. Such tests, together with the functionalities of the pigs server described here, will hopefully be a further step in the understanding of this class of molecules, helping scientists to tackle the open problem of the molecular basis of the specificity of antibodies at large and/or to obtain data useful in the context of specific biological problems.

Conclusions

The accurate prediction of immunoglobulin structures is a fundamental task in immunology. The deep level of understanding of the sequence structure relationship for this protein family has permitted to develop automatic and user-friendly methods. The aim of the pigs web server described here is to enable all scientists (even without a specific bioinformatics background) to tackle the open issue of the specificity of antibodies from a molecular point of view and/or to collect and analyse even large amount of biologically significant data in the complex scenario of immunological systems.

References

- Bower, M. J., Cohen, F. E., & Dunbrack, R. L., Jr. (1997). Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J Mol Biol*, 267(5), 1268-1282.
- Chothia, C., & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol*, 196(4), 901-917.
- Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., et al. (1989). Conformations of immunoglobulin hypervariable regions. *Nature*, 342(6252), 877-883.
- Morea, V., Lesk, A. M., & Tramontano, A. (2000). Antibody modeling: implications for engineering and design. *Methods*, 20(3), 267-279.
- Morea, V., Tramontano, A., Rustici, M., Chothia, C., & Lesk, A. M. (1998). Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J Mol Biol*, 275(2), 269-294.
- Shirai, H., Kidera, A., & Nakamura, H. (1996). Structural classification of CDR-H3 in antibodies. *FEBS Lett*, 399(1-2), 1-8.
- Tramontano, A., Chothia, C., & Lesk, A. M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J Mol Biol*, 215(1), 175-182.
- Webster, D. M., & Rees, A. R. (1995). Molecular modeling of antibody-combining sites. *Methods Mol Biol*, 51, 17-49.