Structural models of the catalytic domain of the yeast β-(1,3)-glucan transferase Gas1 by combined threading and secondary structure prediction methods

Elena Papaleo[a], Marina Vai[a], Laura Popolo[b], Piercarlo Fantucci[a], Luca De Gioia[a]

[a]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milano (Italy)
[b]Department of Biomolecular Sciences and Biotechnology, University of the Study of Milan, Via Celoria 26, 20126 Milano (Italy)

Keywords: threading methods, protein structure prediction, computational methods, glycosidase, TIM barrel

Introduction

Gas1p is an exocellular glycoprotein of *Saccharomyces cerevisiae* and plays a crucial role in cell wall assembly, due to its β-(1,3)-glucan transferase activity [1]. The identification of Gas1p homologues in other yeast species and fungi allowed the definition of a new family of glycosyl hydrolases, family GH72, on the basis of sequence similarity [2].
Hydrophobic cluster analysis [2] of the catalytic domain (C-domain) of some GH72 members suggests a $(\beta/\alpha)_8$ barrel fold, also supported by our recent study on the structural and functional characteristics of the C-domain of Gas1p [3].
Standard homology modelling approaches cannot be used to infer the structure of C-domain of Gas1p and related proteins, due to the lackness of suitable homologues of known 3D structures. Threading and fold recognition approaches have been shown to predict fold of novel proteins with relatively high accuracy. However it should be noted that the detection of possible remote homologues is only the first step of successful modelling. In fact alignment to the same scaffold produced by different threading methods can be significantly dissimilar and affected by local errors, making difficult the derivation of a good structural model [4].
With the aim of unraveling the key molecular characteristics of the C-domain of Gas1p and related proteins, in the present work, a procedure has been worked in which data derived from threading methods, multiple sequence alignments and secondary structure predictions were merged and compared to experimental results in order to obtain reliable and detailed three dimensional models.

Results and discussion

1. Sequence analysis and secondary structure informations

The C-domain sequences of the 29 members of family GH72 were aligned, in order to disclose strictly conserved and similar amino acids. Analysis of the alignment and use of PRATT program [http://www.expasy.org/tools/pratt/] allowed also to define 2 functional fingerprints, specific for family GH72 and comprising the two catalytic residues.
Therefore, to obtain the most reliable secondary structure, we submitted the sequence of all C-domain of the GH72 members to JPRED [http://www.compbio.dundee.ac.uk/~www-jpred/] and PSI-PRED [http://www.psipred.net] prediction servers, obtaining a consensus prediction using a 75% stringency threshold. This allowed to highlight possible structural characteristics: even if the general $(\beta/\alpha)_8$ architecture was predicted with high confidence, some irregularities characterizing the C-domain of Gas1p and cogeners cannot be excluded.

2. Model construction

The sequences of the C-domain of 3 proteins of GH72 family, extensively investigated experimentally and significantly similar (about 55% identity), were selected. These sequences were submitted to 5 threading servers that use different approaches, with the aim of producing consensus predictions: 3D-PSSM [http://www.sbg.bio.ic.ac.uk/~3dpssm/], mGenTHREADER [http://bioinf.cs.ucl.ac.uk/psipred/], SAM-T02 [http://www.cse.ucsc.edu/research/compbio/HMM-apps/], 123D+ [http://123d.ncifcrf.gov/123D+.html] and Fugue [http://www-cryst.bioc.cam.ac.uk/~fugue/].
In order to refine the search of suitable scaffolds for structure prediction and to obtain reliable alignments, a precise selection protocol was applied. In the first step the scaffold set was pruned preserving only protein characterized by the $(\beta/\alpha)_8$ fold and belonging to the same superfamily of family GH72. This subset was further pruned keeping only scaffolds that were chosen by more than half of the servers containing that particular protein in their database.
It should be noted that despite many threading servers converge on the same scaffolds, the corresponding alignments can be quite different [4]. The first pruning rule to evaluate the best sequence-structure alignments was based on the well estabilished observation that some amino acids, suggested to be crucial for catalysis, are strictly conserved in all the superfamily [1,2] and therefore were expected to be proper aligned by the threading methods. According to this rule all the alignments in which these conserved residues were misaligned between sequence and templates were eliminated.
Moreover, the 2D structure prediction data, which had been useful to define the boundaries of the elements forming the $(\beta/\alpha)_8$ fold, were used to evaluate the best corresponding matches between the known 2D structure of the templates and the predicted 2D structure for GH72 members. Alignments whit the 2D elements of the templates that were not correctly aligned or extremely different from those predicted for the C-domain of GH72 enzymes, were discarded.

On the basis of these criteria only 4 templates and 10 alignments survived the pruning procedure and were used to build 3D models of the C-domain of Gas1p by Nest program [http://trantor.bioc.columbia.edu/~xiang/jackal], which generates a 3D model on the basis of a given alignment. Finally the models were submitted to standard geometry optimization by steepest descent and conjugate gradients algorithms.

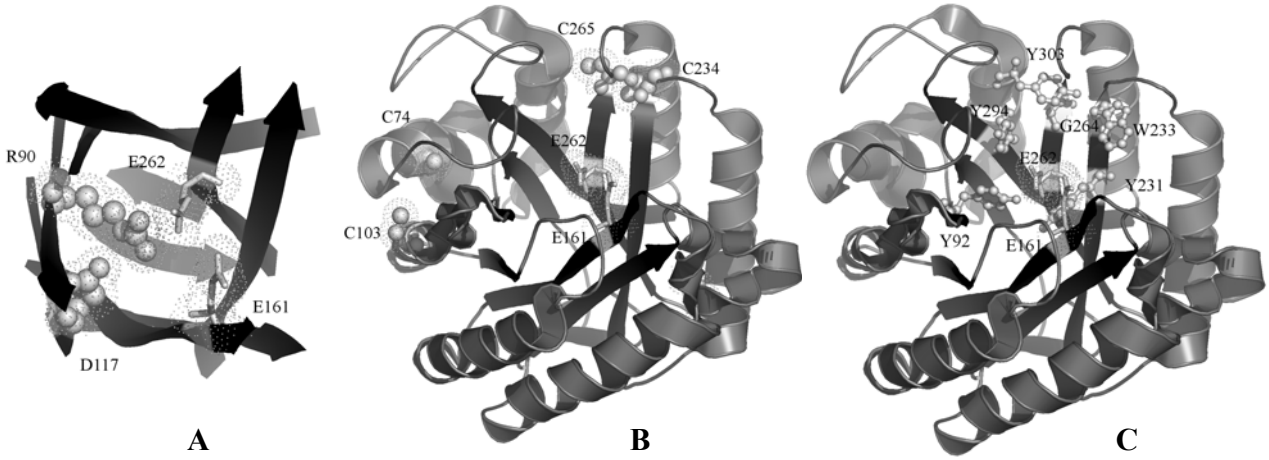

Fig 1. 3D structures of the Gas1p C-domain as predicted using as templates 7A3H. A) The β-strands forming the barrel and the side chains of conserved Arg, Asp and catalytic Glu residues are explicitly shown. **B)** The side chains of the conserved Cys and catalytic Glu residues are shown. **C)** The side chains of the conserved Tyr, Trp, Gly and catalytic Glu residues are shown.

3. Model analysis

The 3D models for Gas1p were all analyzed indipendently and then compared to generate a consensus relative to the general structural elements and to the location of some strictly conserved residues.
In particular, an Arg residue and an Asp were located in all the models in proximity of the active site and they are suggested to stabilize one of the two catalytic residues [Fig.1A]. Instead two pair of Cys residues were always located in similar position in the models, suggesting a high confidence prediction for the localization of the disulphide bridges [Fig.1B]. Moreover the presence of conserved Tyr and Trp residues could be relevant because they are often involved in substrate recognition in glycosyl hydrolases. The analysis of these residues in our models allows to highlight a possible subset of amino acids involved in substrate recognition [Fig.1C].
In conclusion, the careful merging of the results obtained by different computational methods and biochemical data, has made possible the prediction of the 3D structure of the C-domain of Gas1p and related proteins. In this context, it is relevant to note that, even if the matches obtained from automated servers are generally of high quality, it is the careful pruning of results which allows to make considerations at a highly detailed level, leading to reliable sequence-structure alignments.
Since, in some fungal pathogens, enzymes of GH72 family play a crucial role in cell wall assembly and cell wall is involved in interaction with the host cells and in virulence, the investigation of their structures and conserved residues assumes a great importance [1]. Finally, the inferred structural properties of the C-domain could be used to generate working hypothesis about the structural and functional role of key residues. This could open the possibility for targeted mutagenesis experiments and could help to increase the research about new anti-fungal agents.

Acknowledgements

The authors thank E.Stocchetti, R.Gonella Diaz and P.Mereghetti for fruitful discussion and comments.

References

[1] L. Popolo, M. Vai, *Biochim. Biophys. Acta*, 1426: 385-400, 1999
[2] I. Mouyna, M. Monod, T. Fontaine, B. Henrissat, B. Lechenne, J.P. Latgé, *Biochem. J.*, 347 :741-747, 2000
[3] C. Carotti, E. Papaleo, M. Vai, E. Ragni, T. Fontane, G. Tedeschi, J.P. Latgé, L. De Gioia, L. Popolo, *Eur. J. Biochem.*, *submitted*
[4] C.M. Lemer, M.J. Rooman, S.J. Wodak, *Proteins*, 23: 337-355, 1995