Modelling the Three-Dimensional Structure of a Sugar Binding Protein from a Thermophilic Organism: Analysis on Stability and Sugar Binding Simulations.

Anna Marabotti[(1),(2)], Sabato D'Auria[(3),(4)], Mosè Rossi[(3)] and Angelo M. Facchiano[(1),(2)]

[(1)] Institute of Food Science, Lab. of Bioinformatics, Italian National Research Council, Via Roma 52A/C, 83100 Avellino, Italy
amarabotti@isa.cnr.it; angelo.facchiano@isa.cnr.it
[(2)]Research Center of Computational and Biotechnological Sciences (CRISCEB), Second University of Naples
Via S. Maria di Costantinopoli, 16, 80138 Naples, Italy
[(3)]Institute of Protein Biochemistry, Italian National Research Council, Via P. Castellino, 111, 80131 Naples, Italy
[(4)]Center for Fluorescence Spectroscopy, UMAB, Baltimore, MD, USA.

Keywords. Protein Modelling, Thermostability, Sugar Transporter

Introduction

The characterization of proteins from thermophilic organisms is becoming more and more interesting for possible biotechnological applications. Recently, the complete genome of a hyper-thermophilic archaebacterium, *P. horikoshii*, was sequenced [1] and a sugar binding protein (Ph-SBP) was identified by means of analysis of its sequence similarity. Some preliminary experimental information are available on its binding properties and on its structural features; however, the lack of information about its 3D structure impairs the complete knowledge of its conformational properties and interactions with its ligands. Here, we present the results of the homology modelling strategy we used to predict the 3D structure of Ph-SBP, and the analysis we made on the resulting model in order to assess its reliability, with particular care to its expected thermostability features and sugar binding properties.

Methods

The 3D crystallographic templates used for modelling procedure were found with a BLAST search, and in addition, a TOPITS prediction-based threading was performed, confirming the reliability of chosen templates. The alignment of the proteins was made with the program CLUSTALW and few manual refinements were added to correct gaps falling in the middle of secondary structures of Ph-SBP, predicted with PHD. The program MODELLER 4 implemented in Quanta molecular simulation package was used to create the full-atom protein models. All other molecular simulations and analysis were carried out with InsightII package. The best model among those obtained was chosen by evaluating the stereochemical quality with the program PROCHECK. Solvent accessibility of the amino acids was calculated with the program NACCESS. Evaluation of helix stability was carried out according to Facchiano et al. [2]. Docking of sugars into the binding site of Ph-SBP was performed using as starting reference the positions of the oligosaccharides in the crystallographic templates, with further optimization in order to allow a better accommodation of the ligand in the binding groove and to decrease structural conflicts due to sterical hindrance. After then, evaluation of the intermolecular energy of the optimized complexes was performed. As a comparison, intermolecular energy was also calculated for crystallographic complexes.

Results and discussion

1. Modelling procedure

We found three proteins suitable as templates to model Ph-SBP: maltodextrin-binding protein from *P. furiosus*, trehalose-maltose binding protein from *T. litoralis* and maltodextrin-binding protein from *E. coli*. They belong to the same functional and structural family as Ph-SBP, but the first two arise from thermophilic organisms, whereas the third comes from a mesophilic source. The percentage of sequence identity between Ph-SBP and all templates is lower than 30%. In order to overcome alignment problems, we performed a multiple sequence alignment with Ph-SBP, the chosen templates and other 15 proteins belonging to the same family and found significantly similar to our protein with the BLAST search. The alignment of the sequences of Ph-SBP and of templates was then extracted from the multiple alignment, and the positions of the experimentally observed secondary structure elements of templates and of the predicted secondary structure of Ph-SBP were superimposed to the aligned sequences. Manual refinement was made before using this final alignment as the starting point to predict the 3D structure of Ph-SBP.
A peculiar procedure was adopted to create the 3D model. Two sets of ten structural models were created by using MODELLER, in two distinct sessions, either excluding or including the structure of maltodextrin-binding protein from *E. coli,* the mesophilic organism. We then selected the best model created in each session, and used them as templates to create a third set of ten models. The most reliable of them in terms of stereochemical quality shows 90.6% of residues in most favoured regions with no residues in disallowed regions. A complete validation analysis on this model was also performed with ADIT Validation server at RCSB and confirmed the good quality of this model. The coordinates of Ph-SBP were deposited at PDB data bank and accepted with the PDB code 1R25.

By comparing the RMSD between the final model and the two template models, we found a mean value of 0.53 Å from the one obtained without the protein from *E. coli* and 1.16 Å from the one obtained with the mesophilic protein. We hypothesize that some structural features typical of thermophilic proteins and hidden in Ph-SBP sequence were recognized by this modelling strategy.

2. Stability of Ph-SBP

In order to further assess the reliability of our model, we verified if its structural features were in agreement with its thermophilic origin. With a comparative solvent accessibility analysis, we found in our model at least three clusters of Ile residues buried in the interior of Ph-SBP. Moreover, three Leu clusters are in close contact with Ile clusters and determine a tight disposition of hydrophobic residues in the core of the model, typical of thermostable proteins. Several protic residues are buried in the C-terminal domain of our model, forming a cluster of salt bridges sequestered by the aqueous environment that are probably co-responsible of the high thermostability of this protein. Furthermore, 6 proline residues are found at the end and 5 at the beginning of α-helices. This particular location of proline residues also seems to be typical of thermophilic proteins.
Finally, the calculation of the energetic contribution of α–helices to the global energy of the structure indicates that in the Ph-SBP model the helices are more stabilized than in the other three templates. Ph-SBP helices show a lower content of destabilizing β-branched residues (both as absolute number and in percentage) and a higher number of stabilizing positive charges at the C-end of helices, with respect to the mesophilic template, and in line with thermophilic proteins belonging to the same family.
Taken together, all these considerations allow us to consider this modelled structure well compatible with the structural features expected for a thermophilic protein, and confirm the good quality of the prediction.

3. Simulation of sugars' binding

We performed a preliminary binding simulation of different oligosaccharides (maltotriose, maltose and trehalose) to Ph-SBP. To identify residues interacting with the ligand we defined a "minimal binding site" composed by amino acids included in a distance of 5 Å starting from the centre of gravity of the sugar, for all complexes. The binding site is characterized by a relevant presence of apolar amino acids that forms favourable interactions with the apolar rings of the sugars. Polar amino acids are also present and create a network of hydrogen bonds, together with the backbone of other residues. Apparently, the presence of protic amino acids in the binding site of Ph-SBP is very limited. In order to understand how favourably these sugars interact with the Ph-SBP, the intermolecular energy of the complexes was computed and resulting values were compared with those of crystallographic complexes. These energy values for the Ph-SBP complexes are in the same range with respect to templates complexes, and their negative values give an indication of the possibility for this protein to create stable complexes with all these different sugars. From our data, it is possible to argue that the complex between Ph-SBP and maltotriose is more stable than those with other sugars, and trehalose, whose absolute value is smaller than those of any other complex, is the less preferred sugar for binding. It is interesting to note that large parts of difference in intermolecular energy arise from the difference in Van der Waals contribution, further confirming the importance of non polar interactions to stabilize and promote the binding.

Conclusions

The model of the Ph-SBP is compatible with expected features of a thermophilic sugar binding protein. The characterization of its binding site could open an interesting perspective to modify or increase the affinity of this protein for specific sugars in the view of its possible biotechnological application, e.g. as biosensor.

References

[1] Y. Kawarabayasi, M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. I. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K .I. Aoki, T. Yoshizawa, Y. Nakamura, F. T. Robb, K. Horikoshi, Y. Masuchi, H. Shizuya and H. Kikuchi, Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Research,* 5:55-76, 1998

[2] A. M. Facchiano, G. Colonna and R. Ragone, Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. *Protein Engineering***,** 11:753-760, 1998