Magda E. Alvarado[1], Camila A. González[2], Moisés Wasserman[1], Claudia C. Rubiano[1,*]

[1] Laboratorio de Investigaciones Básicas en Bioquímica (LIBBIQ). Universidad Nacional de Colombia.
[2] Universidad Distrital Francisco José de Caldas.
* Corresponding author: ccrubianoc@unal.edu.co.

Recibido: 13 de Junio de 2014. Aceptado: 19 de Julio de 2014.

Preliminary study of the enzyme ubiquitin carboxyl-terminal hydrolase 14 (UBP6) in *Giardia intestinalis*: structural bioinformatic analysis and transcriptional profile during encystation

Abstract

This paper presents a combined approach with two aims. The first is to analyze the reported sequence of the enzyme ubiquitin carboxyl-terminal hydrolase 14 of *Giardia intestinalis* (UBP6) through computational methods to find components related with its hypothetical function. The second is to determine if the protein-coding gene is expressed in *G. intestinalis* and, if such is the case, also determine its transcription pattern along the life cycle of the parasite. It was established that the protein belongs to the family of Cys-dependent deubiquitinases and more specifically to ubiquitin specific proteases (USPs). Moreover, the catalytic center with the complete triad as well as typical features of the USP motif were also identified. Since the computational findings suggest that the enzyme could be functional, reverse transcription coupled to PCR was used as a first approach to establish if in fact the coding gene is expressed in the parasite. Interestingly, it was found not only that the gene is expressed, but also that there is a transcription variation along the life cycle of the parasite. These two findings are the starting point for further studies since they tentatively suggest that this enzyme could be involved in the protein turnover that occurs during parasite encystation. Although preliminary, this study is the first report concerning the study of a specific deubiquitinating enzyme in the parasite *G. intestinalis*.

Key words: deubiquitinating enzymes, ubiquitin, *Giardia intestinalis*, encystation.

Estudio preliminar de la enzima ubiquitina carboxi-terminal hidrolasa 14 (UBP6) en *Giardia intestinalis*: análisis bioinformático de su estructura y perfil transcripcional durante la enquistación

Resumen

En este trabajo se presenta una estrategia combinada que buscaba, primero, analizar por métodos computacionales la secuencia de la enzima ubiquitina carboxilo-terminal hidrolasa 14 de *Giardia intestinalis* (UBP6) reportada para buscar componentes relacionados con su función hipotética y segundo, determinar si el gen que codifica para la proteína se expresa en *G. intestinalis* y si lo hace, cómo es su patrón de transcripción a lo largo del ciclo de vida del parásito. Se encontró que la proteína pertenece a la familia de deubiquitinasas dependientes de cisteína y más específicamente a las proteasas específicas para ubiquitina (USPs por *ubiquitin specific proteases*). También se identificaron el centro catalítico con la triada completa así como características típicas del motivo USP. Teniendo en cuenta que los resultados computacionales sugieren que la enzima puede ser funcional, se usó la técnica de transcripción reversa acoplada a PCR como un primer acercamiento para establecer si el gen codificante se expresa en el parásito. De manera interesante, se determinó no solo que el gen se expresa sino que existe una variación de su transcripción a lo largo del ciclo de vida del parásito. Estos hallazgos son el punto de partida para posteriores estudios ya que sugieren de manera preliminar que esta enzima podría estar involucrada en el recambio de proteínas que ocurre en el parásito durante el proceso de enquistación. Aunque preliminar, este estudio es el primer reporte acerca de una enzima deubiquitinadora específica en el parásito *G. intestinalis.*

Palabras clave: Enzimas deubiquitinadoras, ubiquitina, *Giardia intestinalis*, enquistación.

Estudo preliminar da enzima ubiquitina carboxi-terminal hidrolasa 14 (UBP6) em *Giardia intestinalis*: análise bioinformática estrutural e perfil transcricional durante encistamento

Resumo

Este artigo apresenta uma abordagem combinada com dois objetivos. A primeira é analisar a sequência informou da enzima ubiquitina carboxil-terminal hidrolase 14 de *Giardia intestinalis* (UBP6) através de métodos computacionais para encontrar os componentes relacionados com a sua função hipotética. A segunda é para determinar se o gene de codificação da proteína é expressa em *G. intestinalis* e, se for o caso, também determinar o seu padrão de transcrição ao longo do ciclo de vida do parasita. Foi estabelecido que a proteína pertence à família de deubiquitinases Cys-dependentes e mais especificamente para proteases específicas de ubiquitina (USPs por *ubiquitin specific proteases*). Além disso, o centro catalítico com a tríade completo, bem como as características típicas do motivo USP também foram identificados. Uma vez que os resultados computacionais sugerem que a enzima poderia ser funcional, a transcrição reversa acoplada a PCR foi utilizado como uma primeira abordagem para determinar se, de facto, o gene codificante é expressa no parasita. Curiosamente, verificou-se não só que o gene é expresso, mas também que há uma variação de transcrição ao longo do ciclo de vida do parasita. Estes dois elementos são o ponto de partida para estudos posteriores, uma vez que tentativas sugerem que esta enzima pode estar envolvida no refill de proteínas que ocorre durante o parasita encistamento. Embora preliminares, este estudo é o primeiro relatório relativo ao estudo de uma enzima deubiquitinadora específica no parasita *G. intestinalis.*

Palavras-chave: Deubiquitinases, ubiquitina, *Giardia intestinalis*, encistamento.

Introduction

Characterization of ubiquitin conjugation pathway and identification of the 26S proteasome as a multi-protein complex with protease activity that degrades poly-ubiquitinated proteins, generated a whole new view of proteolysis. Previously considered as a nonspecific and therefore less important process, it became a relevant field and today, the malfunction of this system is associated with serious diseases such as cancer and neurodegeneration (1, 2). Ubiquitination is a post-translational modification that occurs when an ubiquitin molecule protein (76 amino acids long) is added to a target protein across three successive enzymatic reactions in an ATP-dependent way.

Ubiquitin modification can be of different types: monoubiquitination, multi-monoubiquitination (attachment of an ubiquitin on different residues) and polyubiquitination (assembly of a chain of ubiquitins on a single residue). Each type of modification generates various structural changes that define the fate of the target protein (3). For example, histone monoubiquitination involves transcription regulation, while polyubiquitination formed on the residue lysine 48 of the ubiquitin sends the target protein to degradation by the 26S proteasome (4).

The 26S proteasome is a multiprotein complex formed by a catalytic particle (20S) and a regulatory particle (19S). When a protein is flagged for degradation with ubiquitin chains, it is recognized by proteins that are part of the 19S particle and that carry ubiquitin-binding domains. Then, the target protein must be deubiquitinated and unfolded to be degraded before passing into the catalytic particle of the proteasome.

Like phosphorylation ubiquitination is a post-translational modification that regulates the stability, localization or activity of the modified protein (5) and like other regulatory changes, it is also reversible. Deubiquitination is performed by specific proteases, called deubiquitinating enzymes (DUBs). DUBs remove ubiquitin from the target protein and disassemble the polyubiquitin chains by hydrolyzing the isopeptidic bond that links ubiquitin to the substrate protein or to another ubiquitin unit (6).

Similar to other proteases, DUBs are normally inactive or inhibited until they are recruited to a specific site by an adapter protein or else bind to the appropriate substrate. This implies that DUB activity is regulated. Mechanisms such as substrate induced conformational changes, activity induction by scaffold or adapter binding, transcriptional regulation, and post-translational modifications have been described and reviewed for some DUBs mainly from humans and yeast (7).

Many of the DUBs described so far have a papain-like catalytic mechanism and based on the conservation of their catalytic domain, are usually grouped into five families. The first four are thiol proteases which differ by the presence of characteristic domains: the Ubiquitin C-terminal Hydrolase (UCH) domain, the Ubiquitin Specific Protease (USP) domain (UBP in yeast), the Ovarian Tumor (OTU) domain and the Josephin Domain (MJD). The fifth family is formed by a small group of metalloproteases that bind zinc and present the JAB1/MPN/Mov34 metalloenzyme (JAMM) domain (8).

Our research group has carried out several pioneering studies of the ubiquitin-proteasome system in the parasite *Giardia intestinalis,* which is a very ancient protozoan from an evolutionary point of view (9), that infects humans and other mammals and is one of the most frequent diarrhea-causing parasites worldwide. This microorganism alternates between two cell forms that allow it to survive in extreme conditions during its life cycle: the cyst, which is the infective form, is resistant to conditions in the environment, as it has a dense cellular wall; and the trophozoite, which is the replicative and mobile form that colonizes the host's small intestine. After the parasite has settled and multiplied in the small intestine, certain trophozoites are induced to differentiate into cysts through a process known as encystation (10). During encystation there are changes in the transcription of genes, proteins and specific structures to transport the material to form the cyst wall and there is a rapid and specific protein turnover (11). Surprisingly, all the components of the ubiquitin-proteasome system described in superior organisms have been identified in *G. intestinalis* by combining computational and experimental approaches (12, 13, 14). In one of those studies, 12 potential DUBs were rescued from the genome sequence data of this parasite (15, 16), a low number compared to the 86 found in the human genome (8), but not surprising considering that the genome of this parasite is considered quite simple. With the aim to explore more about the presence and function of these proteins in the parasite, a preliminary study of one of them is presented in this paper. The selected protein is the ubiquitin carboxy-terminal hydrolase 14, which is similar to the yeast protein UBP6 and is currently annotated as hypothetical in the Giardia database (*http://giardiadb.org/)* (17), that protein is named as GiUBP6 in this paper. This protein was chosen because its putative homologues in human and yeast are well documented making it appropriate to compare the results obtained to those published in other organisms.

Materials and methods

Sequence computational analysis and structure modeling

The predicted protein sequence from the GL50803_8189 gene available in the database Giardia DB and annotated as the hypothetical protein Ubiquitin carboxyl-terminal hydrolase 14 was analyzed using InterPro (*http://www.ebi.ac.uk/interpro)* (18). This resource is useful for classifying proteins into families and for predicting the presence of domains and important sites. Since InterPro combines signatures provided by several different databases (such as Prosite and Pfam), redundancy in results is greatly reduced. Multiple alignments of the USP domains were calculated by MUSCLE (19), using the default parameters, and visualized with Boxshade (*http://www.ch.embnet.org/software/BOX_form.html*).

Prediction of secondary structure as well as modeling by threading of the three-dimensional structure of the protein were carried out on the I-TASSER (Iterative threading assembly refinement algorithm) server (*http://zhanglab.ccmb.med.umich.edu/I-TASSER)* (20). I-TASSER generates full-length model of proteins by excising continuous fragments from threading alignments and then reassembling them using replica-exchanged Monte Carlo simulations. The modeling process starts from the structure templates identified by LOMETS from the PDB (Protein Data Bank) library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments (21). I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments. For each target, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and reports up to five models, which correspond to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of -5 to 2 where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these qualities (20, 22, 23).

Moreover, in order to externally assess the reliability of the selected GiUBP model, the QMEAN score function (24) was used, which is available at the QMEAN Sever for Model Quality Estimation (http://swissmodel.expasy.org/qmean/cgi/index.cgi) (25). Finally, visualization and various structure manipulations were performed using PyMol (*http://www.pymol.org*) (26) and UCSF Chimera (*https://www.cgl.ucsf.edu/chimera/*) (27).

Cell culture and in-vitro encystation

G. intestinalis trophozoites (WB, clone C6) were cultured in Diamond TYI-S-33 pH 7.0 medium supplemented with 10% bovine serum and 0.5 mg/mL bovine bile (bb) at 37 °C in borosilicate tubes (28). For *in vitro* encystation, trophozoites were cultured at 37 °C in TYI-S-33 pH 7.8 medium, supplemented with 10 mg/ml bb (29). Samples were harvested at 0, 6, 12, and 24 hours after stimulus.

mRNA detection and quantification

Total RNA was isolated from *G. intestinalis* at the indicated time of encystation using the commercial kit GeneJET RNA Purification (Thermo Scientific). Semi-quantitative reverse transcription nested polymerase chain reactions (RT-PCR) were done twice for each point. RT-PCR reactions contained 0.5 M of oligonucleotides, 0.5 mM of dNTP's, 3.0 mM of $MgCl_2$, 1U of reverse transcriptase M-MLV-RT (Promega) and 0.5 U of Taq polymerase in a buffer containing 50 mM Tris-HCl at pH 8.3, 75 mM KCl, 3 mM $MgCl_2$ and 10 mM DTT. Oligonucleotides sequences used were as follows (f and r correspond to the forward and reverse primer respectively, all the sequences are given from 5' to 3' end): gene ubp6 (GL5083_8189), ubp-f: CTTTGGAGCAGCTCTACA, ubp-r: AGCTGTGG CTTAATCAGA; control gene ubiquitin (GL50803_7110), ubi-f: ATGCAGAT CTTCGTCAA, ubi-r: CCTTCTGGATGGAGTAG; and control gene cwp1 (GL50803_5638), cwp-f: GCCTTACTTACCTCAAGAC, cwp-r: GTTGTCACTCATGTACCAG. Reverse transcription was performed according to manufacturer's instructions. cDNA was amplified with the following program: 94 °C/ 5 min, 30 cycles at 94 °C/ 45 s, 48 °C/ 45 s, 72 °C/ 1 min and 72 °C/ 7 min. DNA absence verification for each RNA preparation was assessed using RNA as a template instead of cDNA. All the gene transcripts were amplified on 40 ng of RNA. It was verified that for each amplified gene, that amount of RNA was in the lineal range of response, therefore the changes in the amount of product are proportional to the template concentration. Amplification products were analyzed by agarose electrophoresis and quantified by densitometry with imageJ (*http://imagej.nih.gov/ij/*) (30).

Results and discussion

The first approach in studying the GiUBP6 protein was to use the sequence information available in the GiardiaDB database. If not only sequence but also structure homology to other DUBs already known could be established, it would be a good indication that in fact that protein has a chance of also being a DUB; if it is then expressed in the parasite it would be worth studying it further.

Identification of conserved USP catalytic domain in GiUBP6 sequence and structure

The analysis of the coding sequence for the GL50803_8189 gene (protein GiUBP6, 459 amino acids) using InterPro showed two domains, one ubiquitin related domain also called ubiquitin-like domain (UBL domain, 78 aa length, positions 2-80, IPR029071) and one ubiquitin-specific protease domain (USP domain, 344 aa length, positions 111-455, IPR028889 and IPR001394). The presence of the last one indicates that GiUBP6 belongs to the USP/UBP family of DUBs, which is the largest and most diverse family of those proteins with around 16 UBPs in yeast and more than 50 USPs in humans (31).

The USP catalytic domain found in the GiUBP6 protein is 344 amino acids long, which agrees with the usual length for that domain (around 350 amino acids long). It is already known that the USP family contains two well-conserved sequences, the Cys box and His box, which contain the active site residues that form the catalytic core (32, 33). In order to test this the identified USP domain of GiUBP6 was aligned with USP domains of proteins of different organisms, which have already been characterized. The regions of multiple alignment containing the Cys and His boxes are presented in Figure 1A, where the conservation of the C120, H407, and D424 residues (marked with asterisks in Figure 1A) that would form the catalytic triad can be seen. Similarly, the product of the GL50803_8189 gene is annotated in the GiardiaDB database as the "Ubiquitin carboxyl-terminal hydrolase 14" hypothetical protein although it shows high similarity to the UBP6 protein of yeast, and thus, one question arose: is that protein the number 6 or the number 14 of DUBs? To answer it, the domain distribution of 9 DUBs containing USP domains was used for comparison (Figure 1B). Regarding the size of the protein and presence of domains, GiUBP6 is close to USP14 and UBP14 from human and mouse respectively, whose homologues in yeast and in Arabidopsis are called UBP6. Also, it can be seen that in these last two organisms there are UBP14 proteins, but they are not like those of humans and mice. Although there is certain ambiguity regarding the nomenclature of proteins containing USP domains, GiUBP6 can be considered as similar to USP14 in humans as well as to UBP6 in yeast. What is interesting is that both proteins (USP14 and UBP6) are localized at the proteasome, where they play a special role in rescuing proteins from degradation by removing ubiquitin (34-37). This release of ubiquitin spares it from degradation, minimizing fluctuations in the free ubiquitin pool. It remains to be tested if GiUBP6 could have similar functions.

In addition to the USP domain, GiUBP6 presents an Ubiquitin-related domain known also as Ubiquitin-Like domain (UBL) at the N-terminal region (Figure 1B and Figure 3A). These UBL domains are inserted at different sites relative to the catalytic domain (Figure 1B), N-terminal to, inserted into or C-terminal to the catalytic domain. Although there is no sequence conservation across UBLs, several different families have been described and it was found that they could play widely different roles in the regulation of the DUB activity (38). In the case of USP14, biochemical analysis has shown that the UBL domain is important for its recruitment to the proteasome (34), hence, it could be interesting to assay this possibility in GiUBP6.

Due to the lack of accurate templates in the PDB (those with a percentage of sequence identity of at least 25% compared to the target) it was not possible to model the three-dimensional structure of GiUBP6 by homology, therefore, it was modeled by threading. The modeling was carried out at the I-TASSER server, which is a well-established platform for protein analysis including modeling. As described before, I-TASSER reports up to five models for a target protein, for GiUBP6 the model with the best C-score was selected among the five reported (-0.79 for the selected model and -2.32, -2.90, -2.92, and -3.20 for the other four not selected). In addition to the C-score two other quality parameters are reported, namely TM-score and RMSD. These are known standards for measuring structural similarity between two structures; however, they are usually used to measure the accuracy of structure modeling when the native structure is known. In case where the native structure is not known, it becomes necessary to predict the quality of the modeling prediction, i.e. what is the distance between the predicted model and the native structures? In this case, I-TASSER predicts the TM-score and RMSD of the predicted models relative the native structures based on

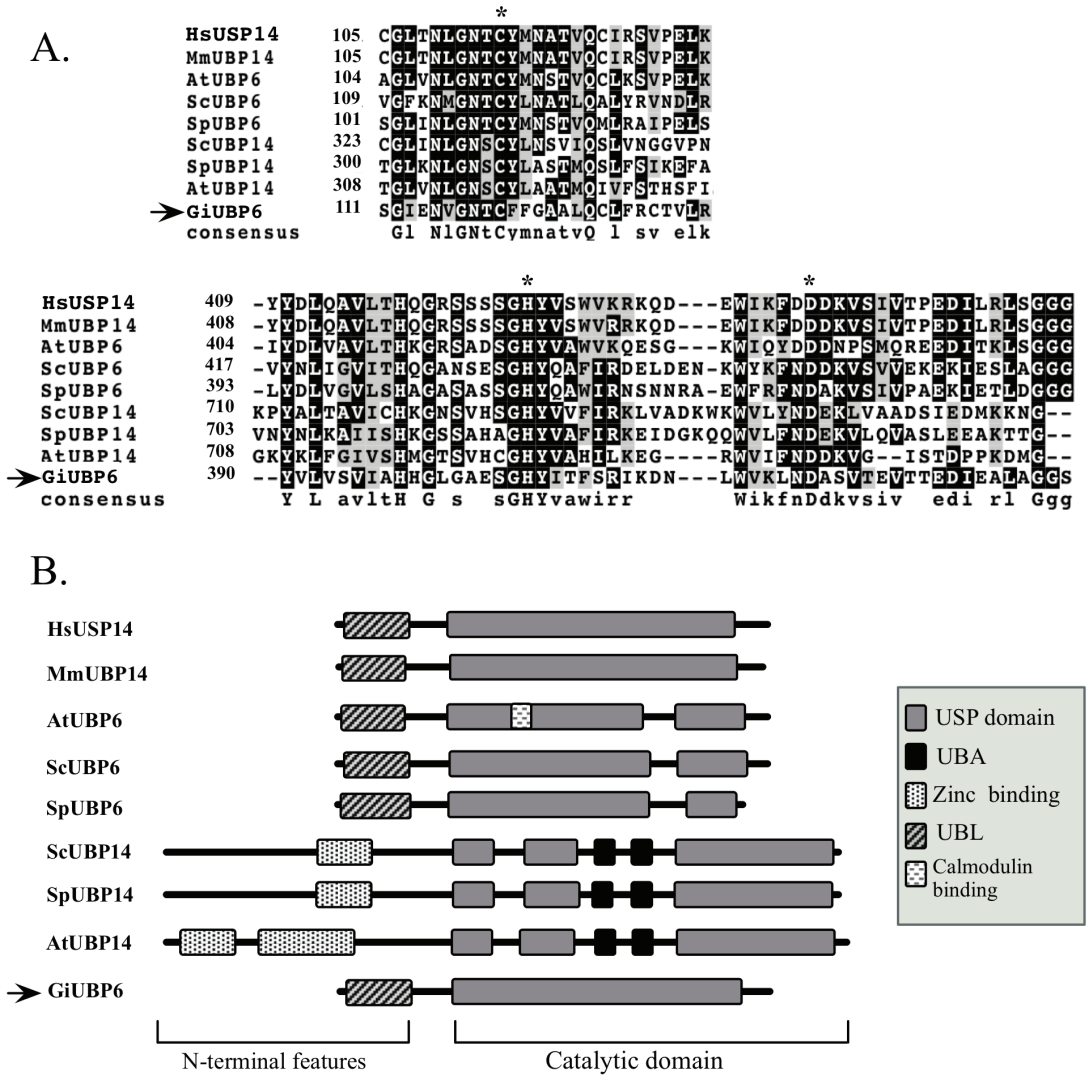

Figure 1. Alignment of USP domains and domain structure of ubiquitin-specific proteases. **A.** Multiple sequence alignment of the conserved motifs surrounding catalytically active amino acid residues (marked by asterisks) in nine known proteins belonging to the USP/UBP family of DUBs, including the protein GiUPB6 from Giardia (signaled with an arrow). The catalytic triad is signaled in the sequence context of the Cys (top) and His boxes (bottom). The catalytic domains of USP domains were extracted from ProSite (http://expasy.org/cgi-bin/prosite/). **B.** Domain architecture of the same ubiquitin-specific proteases (USPs) whose USP domains were aligned in A. UBA corresponds to the ubiquitin-associated domain and UBL to the ubiquitin-like domain also called ubiquitin-related domain. The organisms and sequences used correspond to: Hs (*Homo sapiens*), Mm (*Mus musculus*), At (*Arabidopsis thaliana*), Sc (*Saccharomyces cerevisiae*), Sp (*Schizosaccharomyces pombe*), Gi (*Giardia intestinalis*); the sequences were extracted from UniProt database (http://www.uniprot.org). Access numbers „sp:" correspond to the Swiss-Prot database (it contains manually annoted and reviewed sequences), whereas „tr:" corresponds to TrEMBL database (it contains automatically annotated and not reviewed entries, many of them are annotated as „putative proteins". Such is the case of the sequence of Giardia. HsUSP14 (sp:P54578), MmUBP14 (sp:Q9JMA1), AtUBP6 (sp:Q949Y0), ScUBP6 (sp:P43593), SpUBP6 (sp:Q92353), ScUBP14 (sp:Q11119), ScUBP14 (sp:P38237), AtUBP14 (sp:Q8L6Y1) and GiUBP6 (tr:A8BLG8).

the C-score since it has been found that this parameter is highly correlated with TM-score and RMSD (22, 23). The purpose of using TM-score is to solve the problem of RMSD, which is sensitive to the local error. Because RMSD is an average distance of all residue pairs in two structures, a local error will arise a big RMSD value although the global topology is correct. In TM-score, however, the small distance is weighted stronger than the big distance, which makes the score insensitive to the local modeling error. A TM-score > 0.5 indicates a model of correct topology and a TM-score < 0.17 means a random similarity. This cutoff does not depend on the protein length (39). For the selected model of GiUBP6, the values for the TM-score and RMSD were 0.61±0.14 and 8.9±4.6 respectively. Regarding these values, one can say that the global topology of the model is correct, however the high value of RMSD could reflect some kind of local error, for example a possible misorientation of the N-end region which is bound to the central core by a coil segment that can be very flexible (Figure 3B).

Protein structure models are computational predictions that may contain errors. Since the criteria used to choose the best model for GiUBP6 are not an indication of the global quality of the model, the well-known QMEAN and QMEAN Z-score parameters were used (24, 40). These score functions are widely used to validate experimental structures as well as theoretical models. The used QMEAN scoring function gives a normalized global score of the whole model reflecting the predicted model reliability ranging from 0 to 1 with higher values for more reliable candidates. The normalization reduces the dependence of the quality score on the size of the model. Besides, the QMEAN Z-score

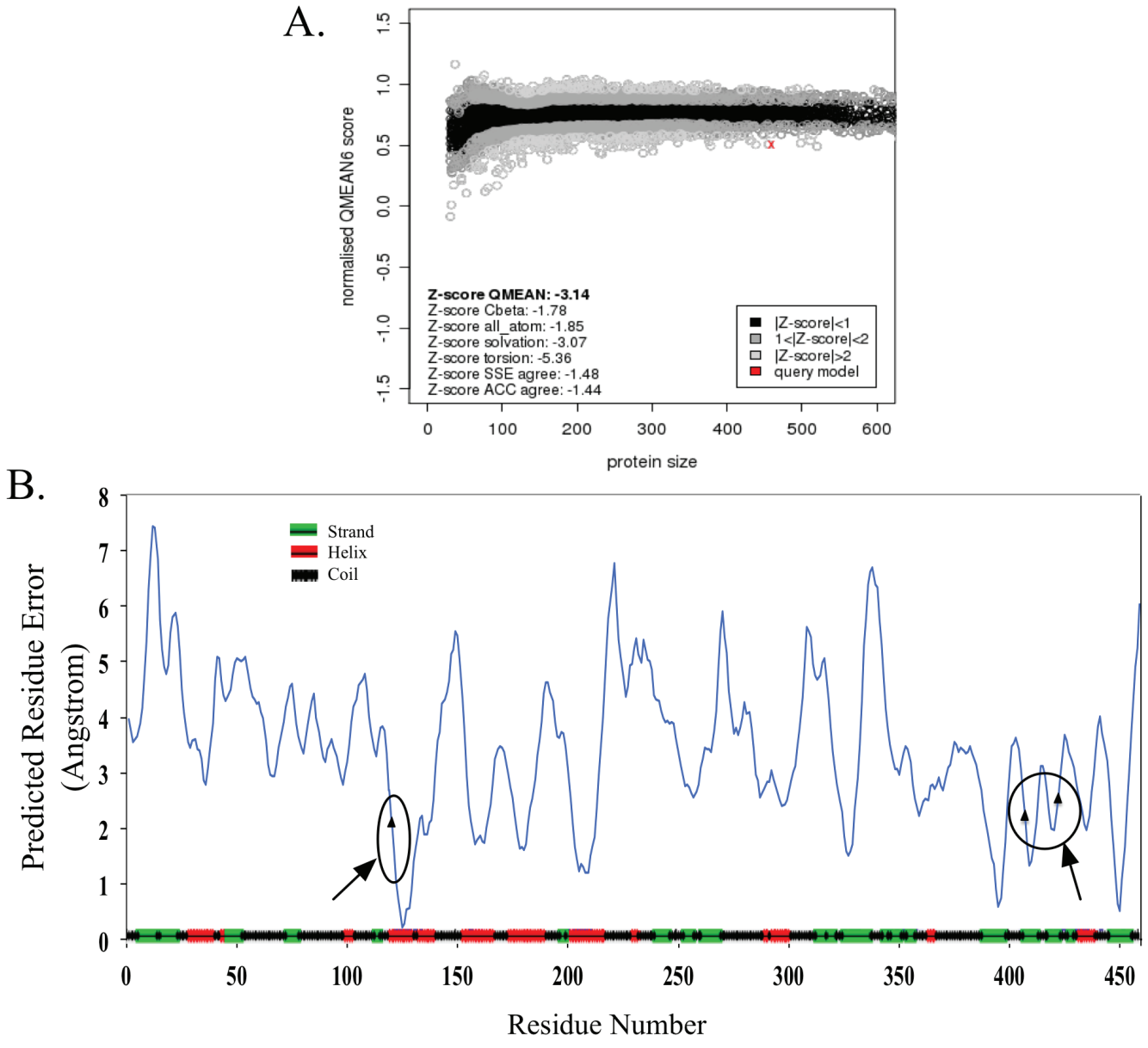

Figure 2. Model quality estimation plots calculated using QMEAN (http://swissmodel.expasy.org/qmean/cgi) (REFS QMEAN). **A.** The global estimated energy of the GiUBP6 model (red cross) is compared to the QMEAN energy estimates (REFs QMEAN) for a nonredundant set of high-quality experimental protein crystal structures of similar length, and their deviation from the expected distributions is represented as Z-scores. The QMEAN quality estimate for GiUBP6 lies within the expected range for models generated by threading and is comparable to a low-medium resolution experimental structure. **B.** Local (per residue) plot of the QMEAN predicted errors for GiUBP6. QMEAN scores for the regions containing the catalytical residues are depicted as circles and arrows, indicating that the local environment of these regions is not located in problematic segments of the predicted structure. The corresponding secondary structure of each region is shown on the horizontal axe.

provides an estimate of the absolute quality of a model by relating it to reference structures solved by X-ray crystallography. This is an estimate of the "degree of nativeness" of the structural features observed in a model by describing the likelihood that a model is of comparable quality to high-resolution experimental structures. The QMEAN and QMEAN Z-score values obtained for the GiUBP6 model were 0.506 y -3.14 respectively. Both values lie within the expected range for a model obtained by threading since this technique is usually more susceptible to errors compared to homology modeling. According to those results, the reliability of the model obtained for GiUBP6 is comparable to a low-medium resolution experimental structure (Figure 2A).

Together with the QMEAN values, it is also possible to obtain an estimated residue error that can be useful to identify problematic regions in the model. Since GiUBP6 was modeled by threading, it is not surprising that the expected quality of some regions of the model is not high. However, it was verified that the important functional sites of the protein, e.g., the catalytic residues, were better modeled than other loop regions of the protein that correspond mainly to coils (Figure 2B). Those regions of the model containing high error values could be further inspected and refined in a future work if the model were intended for specific applications that require a higher quality structure model. For the moment, the GiUBP6 model obtained is useful to recognize some typical features of this kind of proteins that could give an insight into the possible actual function of the protein.

The USP and UBL domains identified on the sequence are also well recognized and spatially separated on the structure (Figure 3A). Addi-

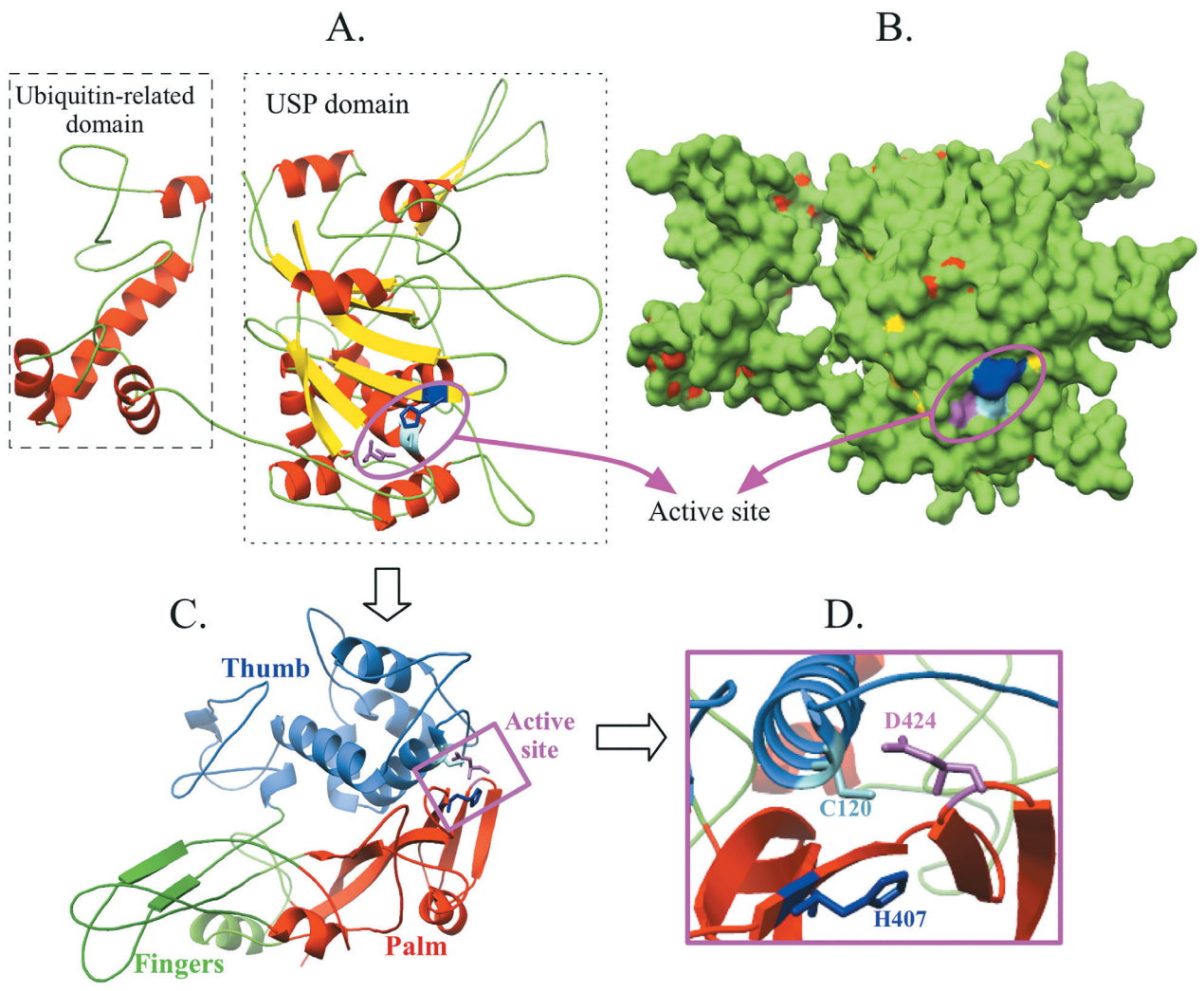

Figure 3. Modeled structure of GiUBP6 protein obtained by threading. **A** and **B.** Ribbons and solid representation respectively of the best model obtained on the I-TASSER server. Helices are presented in red, sheets in yellow and coils in green. Localization of the three amino acids that would form the catalytic center, together with their side chains is pointed out. In B it can be seen that the side chains of these amino acids are located on a cleft that is exposed on the surface of the protein. **C.** Detailed view of the USP-domain topology, the Thumb, Palm and Fingers sub-domains are shown in blue, red and green colors respectively. **D.** Detailed view of the catalytic triad Cys120, His407 and Asp424.

tionally, it can be seen that the amino acids that were postulated as the catalytic core in the sequence are spatially close and exposed on the surface of the protein (Figures 3A and 3B). It is known that the USP domain fold is highly conserved. In Figure 3C a detailed view of USP domain structure is shown and the three well-defined sub-domains (Thumb, Finger, and Palm), which form a structure that resembles a right hand, are recognized. In Figures 3C and 3D, it can be seen that in agreement with previous reports (32, 33, 41), the thumb is predominantly alpha helical and contains the Cys Box with the active site cysteine (C120), whereas the palm is composed mainly of beta strands supported by alpha helices and contains the remaining active site residues that form the catalytic triad: His (H407) and Asp (D424). It is worth to remember that these residues as well as the beta strands and alpha helices that form the core of the USP domain, are located in areas of the model with low error (Figure 2B), thus their spatial orientation is quite reliable.

Taking these results together one could say that it is highly probable that the GiUBP6 protein is an actual DUB. Therefore, in order to start an experimental approach that could determine if the coding gene for the protein is expressed, the transcription of its mRNA was evaluated. Although the expression of mRNA is not always correlated with that of the protein, it is a quite suggestive indication of its presence. Moreover, since the detection of the protein implies a lot of experimental work, it is worth to have a positive result for the expression of mRNA in advance.

mRNA Expression for GiUBP6 in *G. intestinalis*

In order to assess if the coding gene for GiUBP6 is transcribed during the transition from trophozoite to cyst in the parasite Reverse Transcription coupled to PCR (RT-PCR) was chosen. This method was chosen because it is simple, sensitive, and can be adapted to be semi-quantitative by taking care of the amount of template used. Furthermore, since the first goal was to prove the transcription of the gene and to establish an expression pattern along the life cycle of the parasite, the method used was robust enough rendering more complex techniques unnecessary.

To validate the results, two proteins were used as controls, namely Ubiquitin and CWP1 (cyst wall protein 1). On the one hand, Ubiquitin acts as a loading control since its mRNA expression does not change significantly during encystation (12); on the other hand, CWP1 is a marker protein whose gene is only transcriptionally active during encystation (42), and hence, acts as a positive control of that process.

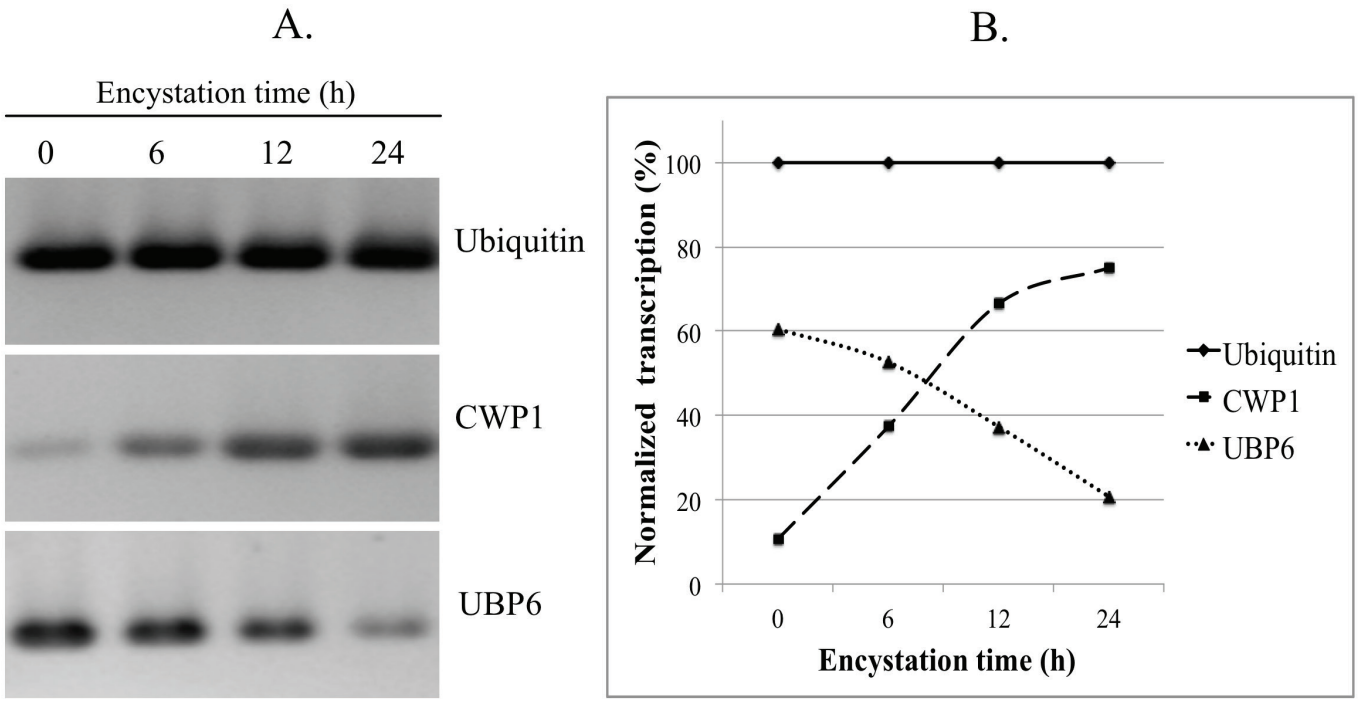

Figure 4. Expression of mRNA for proteins Ubiquitin, CWP1 and GiUBP6 during encystation of *G. intestinalis*. **A.** Agarose gel stained with ethidium bromide showing the RT-PCR amplification fragments for the three mentioned mRNAs using 40 ng of template at each point (0, 6, 12, 24 hours) of encystation. **B.** Relative expression of CWP1 and GiUBP6 transcripts. The signals were normalized against the load control ubiquitin. Data were obtained from densitometry analysis of the bands with the program ImageJ.

It was not only found that the gene coding for GiUBP6 is transcribed during the whole process of encystation, but also that its expression is a regulated process since it changes during encystation as can be seen in Figures 4A and 4B. The expression pattern observed is opposed to the one exhibited in the case of CWP1, it is higher in trophozoites (0 h of encystation), and diminishes progressively as encystation progresses.

As mentioned before, the level of mRNA does not always represent the level of the protein itself, however is a strong indicative of it. Based on the results (Figure 4), one may assume that the parasite needs to inactivate this deubiquitinating enzyme when passing from the metabolically active trophozoite to the inactive cyst.

The UBP6 protein has been well characterized in yeast, and as mentioned before, it is associated with the proteosome DUB particle 19S, which delays the degradation of poly-ubiquitinated substrates by the proteasome. In yeast, UBP6 progressively removes ubiquitin from the end of the chain and generates a time window, which allows the substrate to "escape" from degradation by proteasome (36). It has also been found that UBP6 is indispensable for the correct assembly of the 19S proteasome particle (37). Ubiquitin is then removed from the substrate and it is reused in other ubiquitin dependent processes. In this way, an ubiquitin pool, essential for cellular homeostasis, is always maintained (36). Since USP14/UBP6 can be considered degradation inhibitors, their inhibition would enhance the proteosome activity and thus, the protein degradation. Previous results from our group showed that chemical inhibition of the proteasome significantly decreases the quality and viability of the cysts generated (43). That implies that the proteasome must be very active during encystation to remove proteins with a short half-life, or misfolded due to the stress generated by the stimulus. This agrees with the observed inactivation of the transcription of the gene coding for the GiUBP6 protein showed in this paper. Although this work was only carried out at the transcription level, we are confident that the GiUBP6 protein is expressed in *G. intestinalis* because some peptides possibly coming from this protein have been detected by mass spectrometry in the trophozoite stage (44, 45).

Conclusions

This study is the beginning of the characterization of GiUBP6, a potential de-ubiquitinating enzyme in *G. intestinalis*. By using a combination of computational and experimental approaches, it was possible to establish not only that the protein has sequence and structure features that are typical of de-ubiquitinating enzymes, but also that its gene is transcriptionally active in the parasite. It was determined that GiUBP6 belongs to the USP family of DUBs and would have a functional USP domain. In addition, it was found that the mRNA for GiUBP6 is present in both trophozoites and cysts as well as that it diminishes during encystation.

Being similar in domain distribution and structure to USP14/UBP6 of humans and yeast respectively, the function of GiUBP6 in Giardia could be also related to that of proteins, a good point to test in wet experiments. This hypothesis is consistent with the decrease of the expression during the encystation process, which would have a direct effect on increasing the specific protein degradation.

Acknowledgments

This work received financial support by the Universidad Nacional de Colombia through the Research Division of the Bogotá branch. Hermes Project code 16041.

References

1. Ciechanover, A. The ubiquitin–proteasome pathway: on protein death and cell life. *The EMBO journal*.1998. **17**: 7151-7160.
2. Schmidt, M.; Finley, D. Regulation of proteasome activity in health and disease. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2014. **1843**: 13-25.
3. Navon, A.; Ciechanover A. The 26 S proteasome: from basic mechanisms to drug targeting. *J. Biol. Chem*. 2009. **284**: 33713-33718.
4. Dye, B.; Schulman, B. Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. *Annu. Rev. Biophys. Biomol. Struct*. 2007. **36**:131-150.
5. Ristic, G.; Wei, Ling, T.; Sokol V. T. An optimal ubiquitin-proteasome pathway in the nervous system: the role of deubiquitinating enzymes. *Frontiers in Molecular Neuroscience*. 2014.**7**:1-15.
6. Lee, M.; Lee, B.; Hanna, J.; King, R.; Finley, D. Trimming of ubiquitin chains by proteasome-associated deubiquitinating enzymes. *Mol Cell Proteomics*. 2011. **10(5)**: R110-003871.
7. Reyes-Turcu, F.; Ventii, K.; Wilkinson, K. Regulation and cellular roles of Ubiquitin-specific Deubiquitinating Enzymes. *Annu. Rev. Biochem*. 2009. **78**: 363-397.
8. Eletr, Z.; Wilkinson, K. Regulation of proteolysis by human deubiquitinating enzymes. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2014. **1843**: 114-128.
9. Morrison, H. G.; McArthur, A. G.; Gillin, F. D.; Aley S. B.; Adam R. D.; Olsen G. J.; Best A. A.; Cande, W. Z.; Chen, F.; Cirpriano, M. J.; Davids, B. J.; Dawson, S. C.; Elmendorf, H. G.; Hehl, A. B.; Holder. M. E.; Huse, S. M.; Kim, U. U.; Lasek-Nesselquist, E.; Manning, G.; Nigam, A.; Nixon, J. E.; Palm, D.; Passamaneck, N. E.; Prabhu, A.; Reich, C. I.; Reiner, D. S.; Samuelson, J.; Svard, S. G.; Sogin, M. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science*. 2007. **317**:1921-1926.
10. Adam, R. Biology of *Giardia lamblia*. *Clin. Microbiol. Rev*. 2001. **14**: 447–475.
11. Carranza, P.; Lujan, H. New insights regarding the biology of *Giardia lamblia*. *Microbes and Infection*. 2010. **12:**71-80.
12. Gallego, E.; Alvarado, M.; Wasserman, M.; Identification and expression of the protein ubiquitination system in *Giardia intestinalis*. *Parasitology Research*. 2007. **101**:1-7.
13. Niño, C.; Prucca, C.; Chaparro, J.; Luján, H.; Wasserman, M. The ubiquitin-activating enzyme (E1) of the early-branching eukaryote *Giardia intestinalis* shows unusual proteolytic modifications and play important roles during encystation. *Acta tropica*. 2012. **123**:39-46.
14. Chaparro, J. Estudio de los patrones de ubiquitinación y la actividad de proteólisis específica del proteosoma en *Giardia Intestinalis*. Tesis de doctorado. Departamento de Química, Universidad Nacional de Colombia. Bogotá. 2011. 1-198.
15. Castellanos, I. Implementación de una metodología in silico usando el modelo oculto de markov para la identificación de proteínas involucradas en el mecanismo de señalización con ubiquitina en el sistema ubiquitina-proteosoma en el parasito *Giardia intestinalis*. Tesis de maestría. Departamento de Química, Universidad Nacional de Colombia. Bogotá. 2009. 7-20.
16. Prada, L. F. Análisis de elementos de la maquinaría de deubiquitinación en trofozoítos de *Giardia intestinalis*. Tesis de maestría. Departamento de Química, Universidad Nacional de Colombia. Bogotá. 2012. 1-61.
17. Catic, A.; Sun, Z. Y.; Rather, D. M.; Misaghi, S.; Spooner, E.; Samuelson, J.; Wagner, G.; Ploegh, H. L. Sequence and structure evolved separately in a ribosomal ubiquitin variant. *The EMBO Journal*. 2007. **26**: 3474-3483.
18. Mitchell, A.; Chang, H-Yu.; Daugherty, L.; Fraser, M.; Hunter, S.; Lopez, R.; McAnulla, C.; McMenamin, C.; Nuka, G.; Pesseat, S.; Sangrador-Vegas, A.; Scheremetjew, M.; Rato, C.; Yong, S.; Bateman, A.; Punta, M.; Attwood, T.; Sigrist, C.; Redaschi, N.; Rivoire, C.; Xenarios, I.; Kahn, D.; Guyot, D.; Bork, P.; Letunic, I.; Gough, J.; Oates, M.; Haft, D.; Huang, H.; Natale, D.; Wu, C.; Orengo, C.; Sillitoe, I.; Mi, H.; Thomas, P.; Finn, R. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acid Research*. 2015. **43**: Database issue D213–D221.
19. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004. **32**: 1792–1797.
20. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008. **9**:40.
21. Wu, S.; Zhang, Y. LOMETS: A local meta-threading server for protein structure prediction. *Nucleic Acids Res*. 2007. **35**: 3375-3382.
22. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*. 2015. **12**: 7-8.
23. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. 2010. **5**: 725-738.
24. Benkert, P.; Tosatto, S. C. E.; Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*. 2008. **71 (1)**: 261-277.
25. Benkert, P.; Kuenzli, M.; Schwede T. QMEAN Server for Protein Model Quality Estimation. *Nucleic Acids Res*. 2009. **37** (Web Server issue): W510-4.
26. Schödinger, L. L. C. The PyMol Molecular Graphics System. Version 1.5.0.4.
27. Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera: a visualization system for exploratory research and analysis. 2004. *J. Comput. Chem*. **25**: 1605–1612.
28. Keister, D. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Transactions of the Royal Society of Tropical Medicine and Hygiene*.1983. **77**: 487-488.
29. Kane, A.; Ward, H.; Keusch, G.; Pereira, M. In vitro encystation of *Giardia lamblia*: large-scale production of in vitro cysts and strain and clone differences in encystation efficiency. *The Journal of parasitology*. 1991. **77**: 974-981.
30. Bioimaging Systems. Application Note FP-123. Quantitation using agarose electrophoresis densitometry. 2002.
31. Reyes-Turcu, F.; Wilkinson, K. Polyubiquitin binding and disassembly by deubiquitinating enzymes. *Chem. Rev*. 2009. **109(4):** 1495-1508.
32. Komander, D.; Lord, C. J.; Scheel, H.; Swift, S.; Hofmann, K.; Ashworth, A.; Barford, D. The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. *Mol Cell*. 2008. **29(4):** 451-464.
33. Ye, Y.; Scheel, H.; Hofmann, K.; Komander, D. Dissection of USP catalytic domains reveals five common insertion points. *Mol. BioSyst*. 2009. **5:** 1797–1808.
34. Hu, M.; Li, P.; Song, L.; Jeffrey, P.; Chenova, T.; Wilkinson, K.; Cohen, R.; Shi, Y. Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J*. 2005. **24:** 3747–3756
35. Peth, A.; Besche, H.; Goldberg, A. Ubiquitinated proteins activate the proteasome by binding to Usp14/Ubp6, which causes 20S gate opening. *Mol. Cell*. 2009. **36:** 794–804
36. Hanna, J.; Hathaway, N.; Tone, Y.; Crosas, B.; Elsasser, S.; Kirkpatrick, D.; Leggett, D.; Gygi, S.; King, R.; Finley, D. Deubiquitinating enzyme Ubp6 functions non catalytically to delay proteasomal degradation. *Cell*. 2006. **127**: 99-111.
37. Sakata, E.; Stengel, F.; Fukunaga, K.; Zhou, M.; Saeki, Y.; Föster, F.; Baumeister, W.; Tanaka, K.; Robinson, C. V. The catalytic activity of Ubp6 enhances maturation of the proteasomal regulatory particle. *Molecular cell*. 2011. **42**: 637-649.

38. Faesen, A.; Luna-Vargas, M.; Sixma, T. The role of UBL domains in ubiquitin-specific proteases. *Biochem. Soc. Trans*. 2012. **40:** 539–545.
39. Zhang, Y.; Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins*. 2004. **57**: 702-710.
40. Benkert, P.; Biasini, M.; Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011. **27 (3):** 343-350.
41. Komander, D.; Clague, M.; Urbe, S. Breaking the chains: structure and function of the deubiquitinases. *Nature Reviews: Molecular cell Biology*. 2009. **10:** 550-563.
42. Mowatt, M.; Luján, H.; Cotten, D.; Bowers, B.; Yee, J.; Stibbs, H. Developmentally regulated expression of a *Giardia lamblia* cyst wall protein gene. *Mol. Microbiol*. 1995.**15**: 955-963.
43. Niño, C.; Chaparro, J.; Soffientini, P.; Polo, S.; Wasserman, M. Ubiquitination dynamics in the early-branching eukaryote Giardia intestinalis. *Microbiology Open*. 2013. **2**: 525-539.
44. Catic, A.; Sun, Z. Y.; Rather, D. M.; Misaghi, S.; Spooner, E.; Samuelson, J.; Wagner, G.; Ploegh, H. L. Sequence and structure evolved separately in a ribosomal ubiquitin variant. *The EMBO Journal*. 2007. **26**: 3474-3483.
45. Jerlström-Hultqvist, J.; Stadelmann, B.; Birkestedt, S.; Hellman, U.; Svärd, S. G. Plasmid vectors for proteomic analyses in Giardia: purification of virulence factors and analysis of the proteasome. *Eukaryotic Cell*. 2012. **11**:864-873.

Article citation:

Alvarado, M. E.; González, C. A.; Wasserman, M.; Rubiano, C. C. Preliminary study of the enzyme ubiquitin carboxyl-terminal hydrolase 14 (UBP6) in *Giardia intestinalis*: structural bioinformatic analysis and transcriptional profile during encystation. *Rev Colomb Quim*. 2014. **43**(2): 32-40. DOI: http://dx.doi.org/10.15446/rev.colomb.quim.v43n2.53122