Research, Society and Development, v. 11, n. 10, e427111032765, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i10.32765

Sequence diversity and catalytic properties of phytases

Diversidade de sequências e propriedades catalíticas de fitases

Diversidad en secuencias y propiedades catalíticas en fitasas

Received: 07/11/2022 | Reviewed: 07/26/2022 | Accept: 07/28/2022 | Published: 08/06/2022

Elizabeth Bárbara Epalanga Pires
ORCID: https://orcid.org/0000-0002-9312-0591
Universidade Federal de Viçosa, Brazil
E-mail: elizabeth.pires@ufv.br
Marcelo Depólo Polêto
ORCID: https://orcid.org/0000-0001-9210-690X
Virginia Tech, Blacksburg, United States
E-mail: mdpoleto@vt.edu
Pedro Marcus Pereira Vidigal
ORCID: https://orcid.org/0000-0002-5116-9856
Universidade Federal de Viçosa, Brazil
E-mail: pedro.vidigal@ufv.br
Matheus Italo Bomfim Aragão
ORCID: https://orcid.org/0000-0002-3828-5350
Universidade Federal de Viçosa, Brazil
E-mail: matheus.aragao@ufv.br
Tarley Araujo Barros
ORCID: https://orcid.org/0000-0001-5186-3947
Agroceres Multimix Nutrição Animal Ltda, Brazil
E-mail: tarley.barros@agroceres.com
Rafael Locatelli Salgado
ORCID: https://orcid.org/0000-0001-5059-4714
Universidade Federal de Viçosa, Brazil
E-mail: rafael.salgado@ufv.br
Valéria Monteze Guimarães
ORCID: https://orcid.org/0000-0002-0735-6144
Universidade Federal de Viçosa, Brazil
E-mail: vmonteze@ufv.br
Monique Renon Eller
ORCID: https://orcid.org/0000-0002-5412-897X
Universidade Federal de Viçosa, Brazil
E-mail: monique.eller@ufv.br

Abstract
Phytic acid is an antinutritional factor in cereal feeds, and the use of phytases increases the bioavailability of nutrients bound to this molecule. However, the application of these enzymes depends on their thermal stability and activity at acidic pH. Therefore, in this study we created a database composed of 59 phytase sequences and analyzed the interactions that stabilize their structures in order to understand whether they contribute to the biochemical properties observed. The sequences were aligned and grouped at 30 % similarity, generating 5 clusters, which highlights the high variability among them. A comparative structural analysis of the cluster 3 phytases revealed conserved catalytic domains, as well as eight cysteine residues along the primary sequence, forming disulfide bonds for stabilizing the three-dimensional structure. However, the number of Van der Waals, ionic, and hydrogen interactions, and disulfide bonds was not determinant for the biochemical characteristics presented by these enzymes. The phytase KM873028, from cluster 3, was selected for characterization studies, but its expression in Pichia pastoris generated a protein with properties distinct from those derived from the same sequence expressed in a prokaryotic system. It is likely that the differences observed are associated with the location of the interactions in the structures, non-conserved amino acid residues found around the catalytic site, and post-translational modifications inherent to the expression systems. These possibilities highlight the relevance of strategic choices related to enzyme expression aiming at its production and industrial feasibility.
Keywords: Enzyme sequence conservation; Phytic acid; Animal feed; Enzyme stability; Enzyme structure.

Resumo

O ácido fítico é um fator antinutricional de cereais que compõem rações, e o uso de fitases aumenta a biodisponibilidade de nutrientes ligados a esta molécula. No entanto, a aplicação dessas enzimas depende de sua estabilidade térmica e atividade em pH ácido. Por isso, nesse estudo criamos um banco de dados com 59 sequências de fitases e analisamos as interações que estabilizam suas estruturas a fim de compreender se estas contribuem para as propriedades bioquímicas observadas. As sequências foram alinhadas e agrupadas a 30% de similaridade, gerando 5 *clusters*, o que evidencia a alta variabilidade entre elas. Uma análise comparativa estrutural das fitases do *cluster* 3 revelou domínios catalíticos conservados, assim como oito resíduos de cisteína ao longo da sequência primária, formando pontes dissulfeto para estabilização da estrutura tridimensional proteica. Entretanto, o número de interações de hidrogênio, Van der Waals e iônicas e pontes dissulfeto não foi determinante para as características bioquímicas apresentadas por essas enzimas. A fitase KM873028, do *cluster* 3, foi selecionada para estudos de caracterização, mas a sua expressão em *Pichia pastoris* gerou uma proteína com propriedades distintas daquelas derivadas da mesma sequência expressa em sistema procarioto. É provável que as diferenças observadas estejam associadas à localização das interações nas estruturas, resíduos de aminoácidos não conservados encontrados ao redor do sítio catalítico, além de modificações pós-traducionais inerentes aos sistemas de expressão escolhidos. Essas possibilidades evidenciam a relevância das escolhas estratégicas relacionadas à expressão da enzima visando sua produção e viabilização industrial.

Palavras-chave: Conservação de sequências de enzimas; Ácido fítico; Alimentação animal; Estabilidade de enzimas; Estrutura de enzimas.

Resumen

El ácido fítico es un factor antinutricional en los cereales que componen las raciones, y el uso de fitasas aumenta la biodisponibilidad de los nutrientes en esta molécula. La aplicación de estas enzimas depende de su estabilidad térmica y actividad a pH ácido. Por lo tanto, creamos una base de datos con 59 secuencias de fitasa y analizamos las interacciones que estabilizan sus estructuras para comprender si contribuyen a las propiedades bioquímicas observadas. Las secuencias fueron agrupadas al 30% de similitud, generando 5 clusters, lo que demuestra la alta variabilidad. Un análisis estructural comparativo de las fitasas del grupo 3 reveló dominios catalíticos conservados, así como ocho residuos de cisteína a lo largo de la secuencia primaria, formando enlaces disulfuro para estabilizar la estrutura. Sin embargo, el número de interacciones de hidrógeno, Van der Waals, iónico y disulfuro no es determinante para las características bioquímicas que presentan estas enzimas. La fitasa KM873028, del grupo 3, fue seleccionada para estudios de caracterización, pero su expresión en *Pichia pastoris* generó una proteína con propiedades diferentes a las derivadas de la misma secuencia expresada en un sistema procariótico. Es probable que las diferencias observadas estén asociadas con la ubicación de las interacciones en las estructuras, los residuos de aminoácidos no conservados que se encuentran alrededor del sitio catalítico, y las modificaciones postraduccionales inherentes a los sistemas de expresión elegidos. Estas posibilidades muestran la relevancia de elecciones estratégicas relacionadas con la expresión de la enzima, con el objetivo de su producción y viabilidad industrial.

Palabras clave: Conservación de secuencias enzimáticas; Ácido fítico; Alimentación animal; Estabilidad enzimática; Estructura enzimática.

1. Introduction

Phytic acid is a natural constituent of cereals, such as soybeans and corn, used to produce feed for monogastric animals, such as poultry and swine. These animals do not efficiently metabolize phosphorus from those foods, thus supplementation is necessary to obtain better performance (Haefner *et al*., 2005), which generates an excess of phosphorus in the excreta of the animals and, consequently, in the environment. In addition, phytic acid forms complexes with proteins and other minerals, interfering with the absorption of these nutrients. The addition of exogenous phytases has been considered the most efficient strategy to reduce the need for phosphorus supplementation in pig diets (Huber and Zeller, 2015; McCormick *et al*., 2017; Pramanik *et al*., 2018; Arredondo *et al*., 2019; Kryukov *et al*., 2021; Mulvenna *et al*., 2022; Selim *et al*., 2022), in addition to reducing phosphorus released into the environment through excretion (Korgenay *et al*., 1996). Some phytase-based products available on the market include: PHYZYME (DuPont); QUANTUM and FINASE (AB Vista, AB Enzymes); NATUPHOS (BASF); RONOZYME and ROXAZYNE (Novozyme-Novo Nordisk); AVIZYME (Finffeds International); ALLZYME SSF (Alltech); Bio-Feed phytase (DSM); AMAFERM (Biozyme); OPTIPHOS (Enzyvia, Phytex, Cornell); and ROVABIO (Genecour International). Each of these products has different production yields, stability, specific activity and dosage requirements.

Microbial phytases are the most widely phytases used in the industry main due to the high production yield. However, the biochemical characteristics of most of the phytases that have already been characterized limit their application (Pires *et al*., 2019), since they must maintain their activity after pelleting process (80 °C/2 min) and be active during the passage through the gastrointestinal system of monogastric animals (Bohn *et al*, 2008; Zhang *et al*., 2010; Cowieson *et al*., 2011). Due to this, commercially available enzymes are expensive, especially for Brazilian producers, where these products are imported. Therefore, there is a great demand for enzymes with high thermal stability and ability to act in acidic environments, either by prospecting new enzymes or by improving enzymes already characterized.

The association between the structure and the biochemical properties of enzymes can provide relevant information about factors that affect stability and catalytic activity. Several methods have been used for this purpose, including comparative sequence, structure analysis, molecular dynamics studies, or even the evaluation of random or rational mutations on enzyme activity (Zhang 2007; Liao, *et al*., 2013; Fei *et al*., 2013; Hesampour *et al*., 2015; Kumar *et al*., 2015; Tan *et al*., 2016; Han *et al*., 2018; Li *et al*., 2019; Zhang *et al*., 2020; Acquistapace *et al*., 2022). In this study, we report a comparative approach between sequences, three-dimensional structures and biochemical characteristics of phytases already characterized and described in the literature, adopting factors such as: hydrogen, ionic and Van der Waals (VdW) interactions and disulfide bonds. The study of these factors aims to search for markers responsible for characteristics such as thermal stability and performance in a wide pH spectrum, and then propose strategies for improving the enzyme application in industry.

2. Methodology

2.1 Generating a phytase sequence database

Phytases which sequences were available in the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) by April 2020 were selected. From there, research in scientific articles for information about their biochemical characteristics was performed, especially those related to their ideal temperature, pH and thermostability. From the collected data, a database was generated relating these characteristics, as well as their respective access codes in Genbank (Supplementary Material S1). These sequences were then used as a basis for further comparative analyses.

2.2 Sequence grouping based on similarity

Amino acid sequences were generated and used for clustering based on the sequence identity of the respective proteins. For this, the phytase sequences were aligned using the MAFFT software (https://mafft.cbrc.jp/alignment/server/) (Katoh *et al*., 2019) using the "auto" alignment strategy settings. A pairwise distance matrix (p-distance) was calculated from the alignment of the proteins, which were then grouped using the Neighbor joining (NJ) method and the Molecular Evolutionary Genetics Analysis (MEGA) version 11 software (https://www.megasoftware.net/) (Tamura *et al*., 2021). The consistency of the clusters obtained was verified by means of 1000 bootstrap repetitions. Subsequently, the proteins were regrouped using the algorithm implemented in the CD-Hit software (http://weizhong-lab.ucsd.edu/cdhit-web-server/) (Huang *et al*., 2010), from sequences with identity ≥ 30%. The representative sequence of each group (clusters 1-5) was then identified and compared to a reference phytase (4arv), considering the high degree of dissimilarity between the phytase sequences present in the database. This stage was performed with the help of the BLASTp tool of the BLAST software version 2.13 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul, *et al*., 1990) and aimed to identify the cluster representative among the generated groups, being selected for structural study of the cluster that presented greater

similarity to the structure of the reference phytase. A pairwise matrix of identity and similarity between the proteins was also obtained to identify the level of variation between the sequences.

2.3 Generation of three-dimensional models, quantification and analysis of intramolecular interactions

The comparison of the three-dimensional structures was performed only among the cluster 3 phytase sequences, which presented greater similarity with the phytase reference (4arv). For this, the models were initially built with the help of the AlphaFold server (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb#scrollTo=VzJ5iMjTtoZw) (Jumper *et al*., 2021) and these structures were aligned using the PyMOL tool (https://pymol.org/2/). The model structure was extracted from the RCSB software (https://www.rcsb.org/structure/4ARV) (Ariza *et al*., 2013). Each model generated in the previous step was submitted to the server RING - Residue Interaction Network Generator (http://old.protein.bio.unipd.it/ring/) (Piovesan *et al*., 2016) to identify the number and type of interactions that compose protein structures. The objective of this step was to verify the existence of a connection between the interactions present in the generated structures and their respective biological characteristics in terms of enzymatic activity, supporting a possible rational design of the modifications, aiming at the generation of a more stable phytase.

2.4 Phytase selection and expression.

A phytase sequence was selected based on the biochemical characteristics of the phytases in cluster 3, the group considered as the most representative of the sequences in the database. Therefore, the selected enzyme sequence KM873028 was shown to be expressed with high productivity in prokaryotic system and presented characteristics of heat resistance and performance at acidic pH. It was identified from a metagenomics study, being derived from a garden fungus (*Dendroctonus frontalis*), but originally expressed in *Escherichia coli* (BL21) by Tan *et al*., 2016). The gene encoding this enzyme was optimized for expression in *P. pastoris* and synthesized by GenOne Biotech (Rio de Janeiro, Brazil), inserted into the pPICZαB vector (Invitrogen) under the AOX promoter. The vector was multiplied in DH5alpha cells and the extracted plasmids were linearized and used for transformation of competent *P. pastoris* GS115 cells by electroporation, according to the recommendations of the manufacturer (Invitrogen). Recombinant cells were selected from plates with YPD agar (Yeast Extract - Peptone - Dextrose) supplemented with 100 µg/mL zeocin. The induction took place in BMMHY medium (Buffered Methanol-Complex Medium) containing 0.5% (v/v) of methanol for up to 72 h and the phytase activity was evaluated from the extract obtained from the supernatant culture.

Enzyme activity was determined by using sodium phytate (Sigma-Aldrich) as substrate, as described by Pires *et al*. (2019). The ideal pH of activity was evaluated in medium containing 200 mM of glycine-HCl (pH 2.0–3.5); sodium acetate (pH 4.0-6.0); or Tris-HCl (6.5-7.5). For this, an aliquot of 275 µL of a 1.5 mM substrate solution was added to 100 µL of the extract. The mixture was incubated at 37 °C for 15 min and the reaction interrupted with the addition of 125 µL of 10 % (m/v) trichloroacetic acid. Then, 500 µL of the colorimetric reagent containing 10 % (v/v) of ammonium molybdate in 5 M sulfuric acid and 5 % (m/v) of ferrous sulfate was added to the reaction (Taussky and Skorr 1953). The blank consisted of a mixture under the same conditions, stopping the reaction before starting the incubation.

The absorbance of the solution was measured at 700 nm and the values compared to a standard curve of sodium phosphate. An activity unit was defined as the amount of enzyme capable of releasing 1 µmol of inorganic phosphate per minute of reaction.

3. Results and Discussion

The search for phytase sequences in the Genbank database revealed the existence of 62 enzyme sequences from different organisms, 59 of which could be associated with information about their biochemical characteristics (Supplementary Material S1). The described phytases present a great diversity of properties, being able to act in temperatures between 37 ºC and 60 ºC, and environments with pH of 2.5 to 11 (Supplementary material S1).

3.1 Phylogenetic analysis and sequence similarity

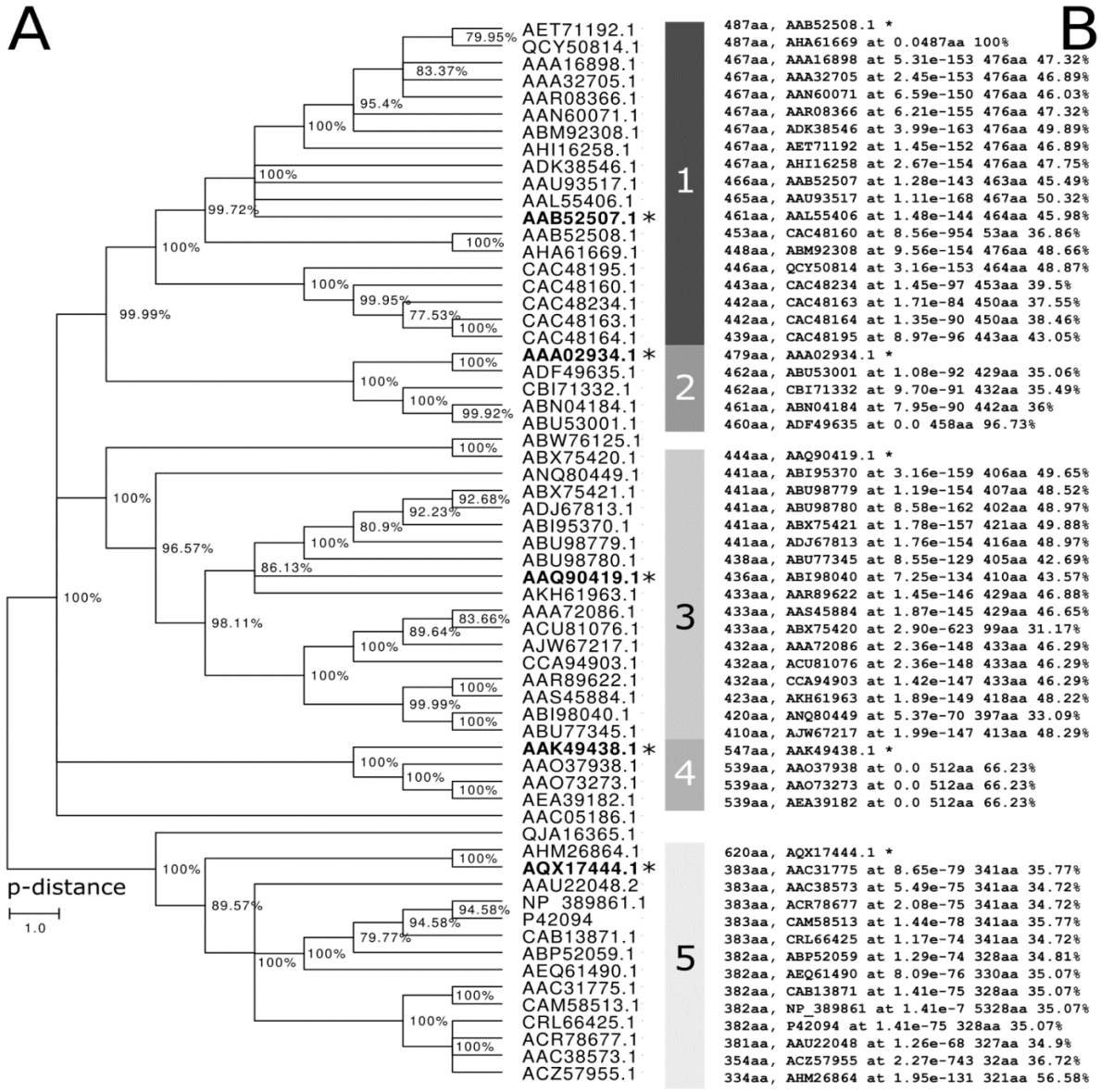
The grouping of the primary amino acid sequences of the phytases present in the database showed a great diversity in the existing phytase sequences, with few conserved regions between the enzymes, and no conserved sequence among all the enzymes. The alignment led to the clustering of the sequences in 5 clusters (Figure 1), which was also confirmed by clustering on CD-Hit using an identity threshold of 30% (Figure 1). Despite the low index adopted, the clustering in CD-Hit showed the same pattern as the clustering in the distance matrix, proving the high diversity between them.

To determine the level of variation in identity and similarity between the sequences, a pairwise matrix was generated from the obtained alignments using MAFFT.

For clusters 1, 2, 4 and 5, the identity between the sequences of each cluster ranged from (33 to 99)%, (17 to 99)%, (64 to 100)%, and (20 to 99)%, respectively. The similarity ranged from (47 to 99)%, (29 to 99)%, (77 to 100)%, and (29 to 99)%. In cluster 3, the identity ranged from (26 to 99)% and the similarity from (43 to 99)% (Supplementary Material S2).

In the analyzes carried out between the representative sequences of each group, it was observed that among the models representing the five clusters, the phytase belonging to cluster 3 (AAQ90419) has greater similarity to the phytase 4arv, from *Yersinia kristeensenii*, identified as the reference sequence. In this case, 49.6% of identity and 65.7% of similarity were observed, while the sequences of the other clusters did not show significant similarity and identity (Supplementary Material S2). This result indicates that the sequences which belong to cluster 3 are the most representative among all the phytase sequences contained in the database. Because of this, this cluster was selected for associative analysis of the biochemical and structural characteristics of the 17 sequences that compose it.

Figure 1: Clustering of phytase amino acid sequences using the Neighbor-Joining method. The bootstrap consensus tree was inferred from 10,000 replicates from a p-distance matrix. The percentages of bootstrap is shown beside each node and branches reproduced in less than 75% bootstrap replicates were collapsed. (B) Clustering of phytase sequences using CD-HIT. The representative sequence of each cluster is highlighted in bold and marked with an asterisk.



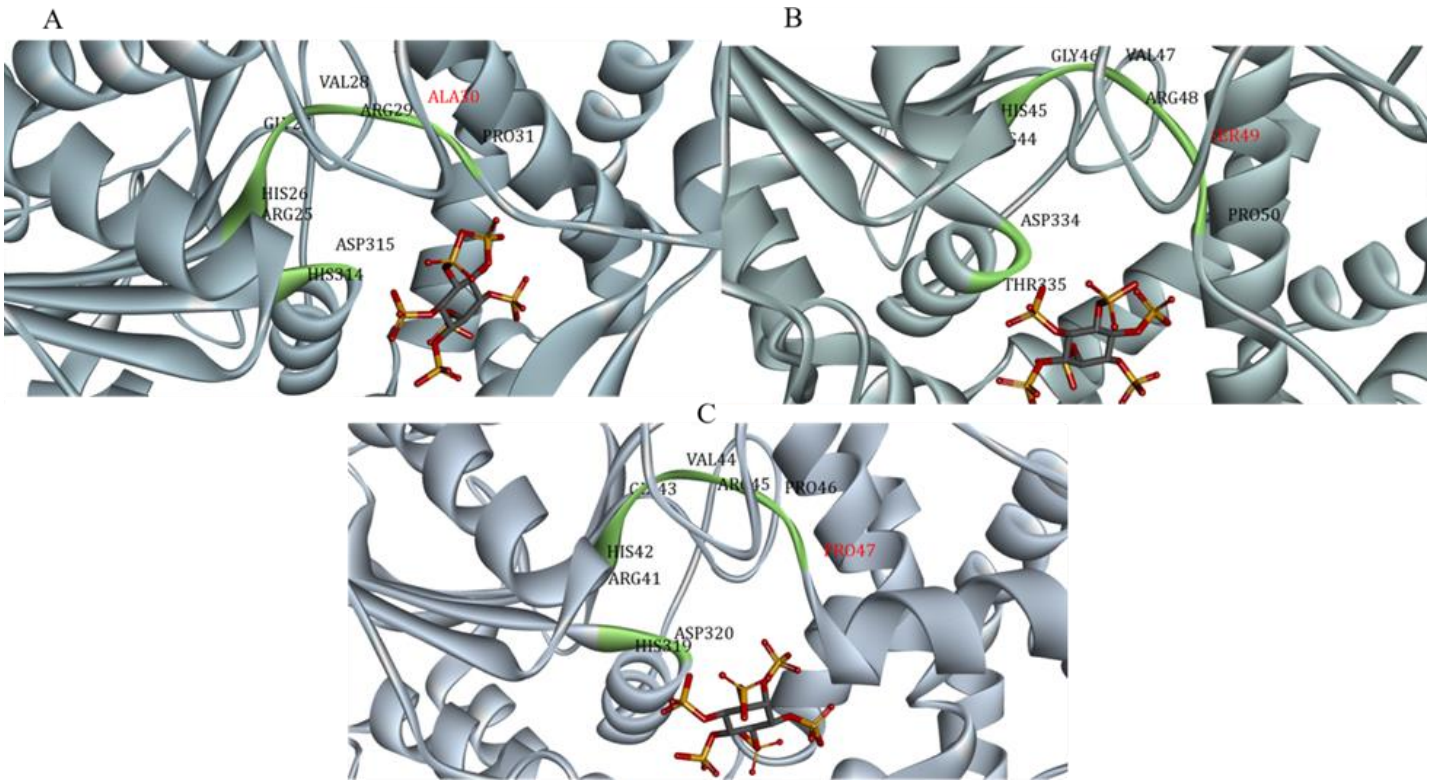
Source: Authors.

3.2 Multiple Sequence Alignment

Concerning the 17 sequences that encode phytases grouped and aligned as the cluster 3, three are identical (AAA72086, ACU81076 and AJW67217) (Figure 3), while the others present distinct residues with few conserved regions with an average of 46% identity. Among the conserved regions, there is the catalytic motif RHGVRAP-HD, except for some A49S mutations, and one A46P mutation (Figure 2). This sequence comprises the amino acids that form the ligand binding site and has been pointed out in

several studies as characteristic of the histidine phosphatase superfamily (HAPs), that share a conserved catalytic core centered on a histidine that is phosphorylated during the course of the reaction (Zinin *et al*., 2004; Wada *et al*., 2019; Isabella *et al*., 2020).

Figure 2: Structure of modeled phytases illustrating the conserved RHGVRAP-HD active site. The active site loop is shown in green and the substrate (phytic acid) in red. A: AKH61963; B: ABX75421 (A49S); and C: ABX75420 (A46P).



Source: Authors.

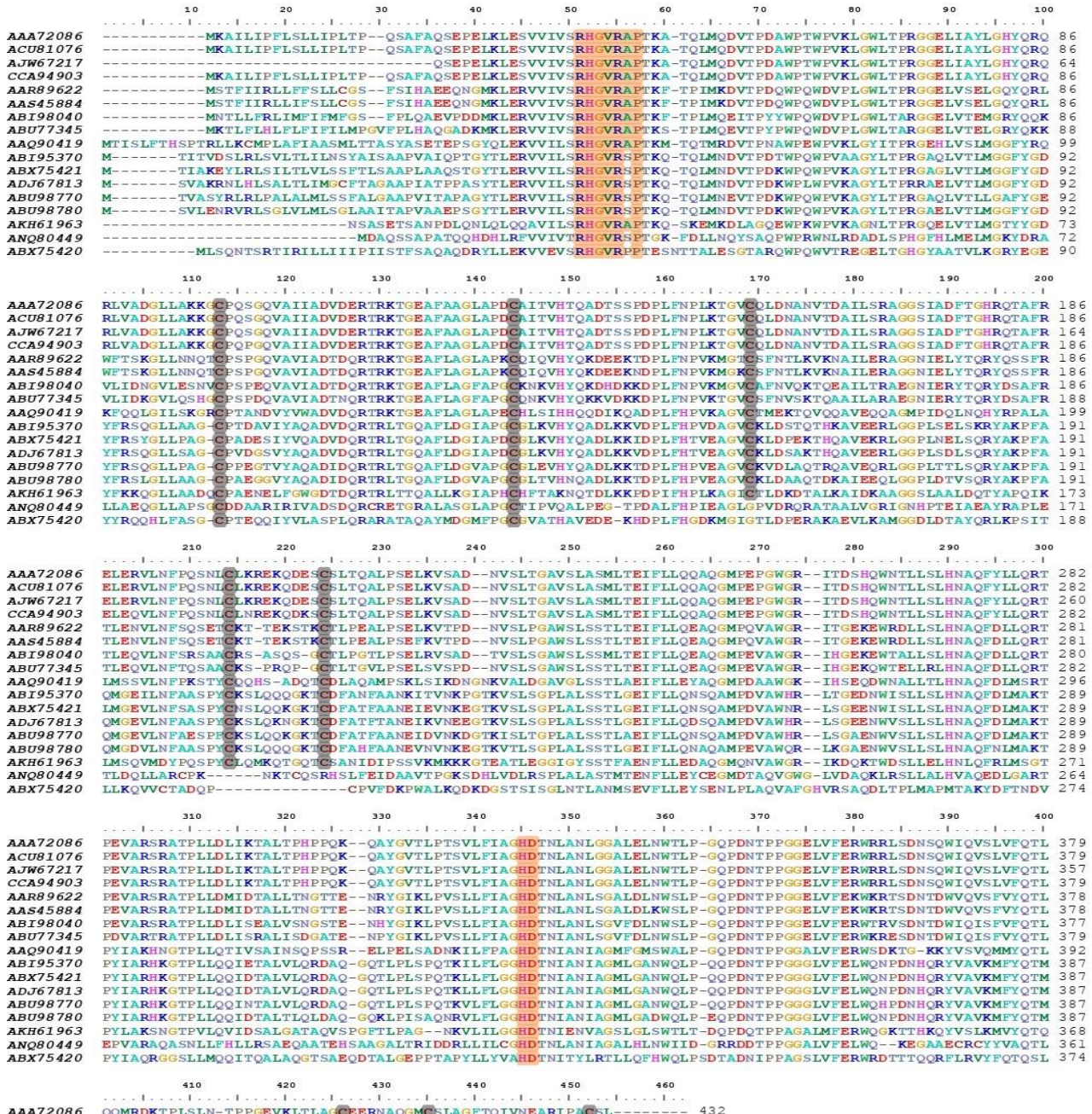
Wada *et al* (2019), evaluated the contribution of each amino acid from the active site in the AppA activity, and observed a drastic change in activity when the mutation involved charged residues within the conserved motif. Non-polar residues such as Ala and Gly did no influence the catalytic characteristics of the enzyme. In another study, when a residue near or within the conserved site of a phosphatase was replaced by bulky residues (Tyr and Phe), the specific activity, thermal stability and tolerance were significantly changed (Fu *et al*., 2011). In our study, variations within the active site involved smaller residues (Ser and Pro) substituting Ala and do not seem to have influenced the enzymatic characteristics, since this substitution was observed for both, a thermostable (ABX75421) as well as a less stable (ABX75420) phytase (Table 1 and Supplementary material S3).

Although the phytases presented a conserved active motif, the residues located near to the active site showed a low degree of conservation. The influence of these changes needs to be deeply explored to effectively evaluate their impact on the interaction with the substrate (activity) and stability of these enzymes. Moreover, eight cysteine residues were also found in conserved regions along the primary protein sequence, which formed disulfide bonds in the predicted tertiary structure for these enzymes. There were

also observed other additional cysteine residues, but which do not maintain the same conservation pattern (Figure 3). Conserved residues normally have functions related to protein activity (Shivange *et al*., 2016) and alterations of residues close to or within the active site of an enzyme can alter substrate binding capacity and enzyme activity depending on the charge and size of the residues (Nayeem *et al*.2009).

Pearson (2013) points out that sequences with more than 40% identity tend to be homologous and have functional similarity, although changes in a small number of residues can lead to significant changes in enzyme activity. Consequently, associating functional characteristics based only on significant local similarity is not as reliable as inferences based on global similarity and conserved active site residues.

Figure 3: Multiple alignment of the 17 phytase sequences from cluster 3: The orange lines highlight the amino acid residues forming the catalytic site; The grey lines show the eight conserved cysteine residues along the primary sequences.



Source: Authors.

Table 1. Biochemical characteristics available in the literature about the 17 phytases from cluster 3.

Acession number	Reference	Original source	Ideal pH or pH range	Theoretical pI	Ideal temperature	Thermostability
ABI95370	Vieira *et al*., 2019	*Yersinia* sp.	8.0	7.60	40 °C	-
ABU98779	Huang *et al*., 2008	*Yersinia rohdei*	1.5 – 6.0	6.79	55 °C	50 % of residual activity after 80 °C for 4 min
ABU98780	Huang *et al*., 2008	*Yersinia intermeda*	1.5 – 6.0	6.01	55 °C	70 % of residual activity after 80 °C for 4 min.
ABX75421	Fu *et al*., 2008	*Yersinia kristeensenii*	4.5 1.5 – 11	6.32	55 °C	46 % of residual activity after 80 °C for 10 min.
ADJ67813	Niu *et al*., 2017	*Yersinia enterocolitica*	2.0 – 9.0	6.60	60 °C	-
ABI98040	Luo *et al*., 2007	*Citrobacter amalonaticus* CGMCC 1696	2.5 - 10.0	6.27	55 °C	-
AAR89622	Zhao *et al*., 2010	*Citrobacter freundii*	2.5 – 4.5 At least 80 % of residual activity between 1.5 and 8.0	6.29	50 °C	
AAS45884	Kim *et al*., 2006	*Citrobacter braakii*	4.0	6.96	37 °C	66 % of residual activity after 70 °C for 10 min.
ABX75420	Shao *et al*., 2008	*Pectobacterium wasabiae* DSMZ 18074	4.0 - 5.5	5.72	50 °C	
AAA72086	Dassa *et al*., 1990	*Escherichia coli*	2.5	6.26		

ACU81076	Yao *et al*., 2013	Synthetic construct	4,5	-	50 °C	75 % of residual activity after 85 °C for 5 min.
CCA94903	Paul Roy *et al*., 2016	*Shigella sp. CD2*	5.5	6.26	60 °C	More than 50 % of residual activity at 70 °C
AKH61963	Tan *et al*., 2016a	*Dendroctonus frontalis*	3.9 3.5 - 5.5	-	52.5 ºC	93 % of residual activity after 15 min for 100 °C
ANQ80449	Tan *et al*., 2016b	Synthetic construc	2.5	-	55 °C	-
AJW67217	Akbarzadeh *et al*., 2015	Synthetic construc	4.5	-	60 °C	70 % of residual activity at 70 °C
AAB52508	Mitchell *et al*., 1997	*Myceliophthora thermophila*	3.5 – 6.0	5.05	-	-
AAQ90419	Zinin *et al*., 2004	*Obesumbacterium proteus*	1.5 - 6.5	6.67	50 °C	-
ABW76125	Gu *et al*., 2009	*Dickeya paradisiaca*	4.5 - 7.5	6.16	55 °C	No activity at 60 °C after 5 min.

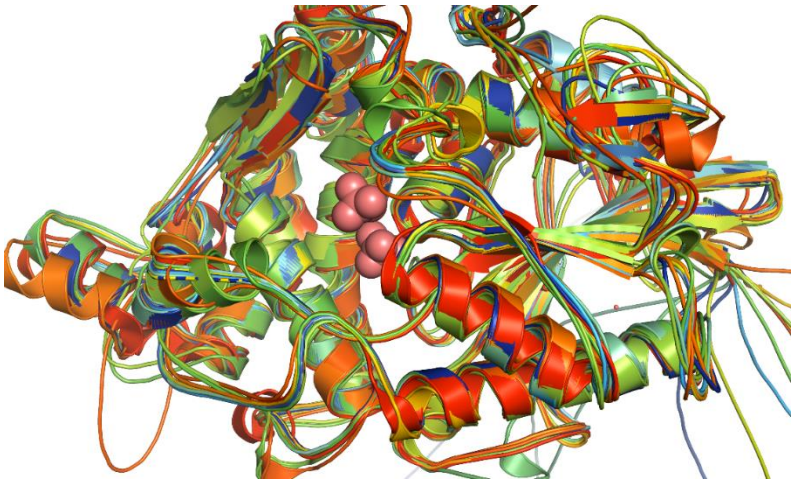
Source: Authors' elaboration.

3.3 Quantification and analysis of interactions of the structural models

The 17 sequences comprising the cluster 3 were submitted to the AlphaFold server (Jumper *et al*., 2021) to predict their three-dimensional structures. The predicted structures of this group are composed of two domains: a major α/β domain and an α smaller domain. The catalytic motif is found in a cavity between the two domains, where they form a loop that integrates the active site. In general, the phytases presents 8 longer α-helices and 4 short α-helices, beyond 7 large β-sheets and 3 short β-sheets. These structural features are well conserved in many bacterial and fungal phytases that have already been studied (Ariza *et al*., 2013; Bohm *et al*., 2010). From these structures, the interactions that form and stabilize each structure were quantified using of the RING server, which allows the identification and enumeration of intra and intermolecular covalent and non-covalent bonds: i) number of hydrogen bond interactions, ii) salt bridges, iii) van der Waals interactions (VdW), and iv) disulfide bridges (Piovesan *et al*., 2016) (Supplementary material S3). Then, the biochemical characteristics of the phytases that constitute this group were correlated with these structural characteristics.

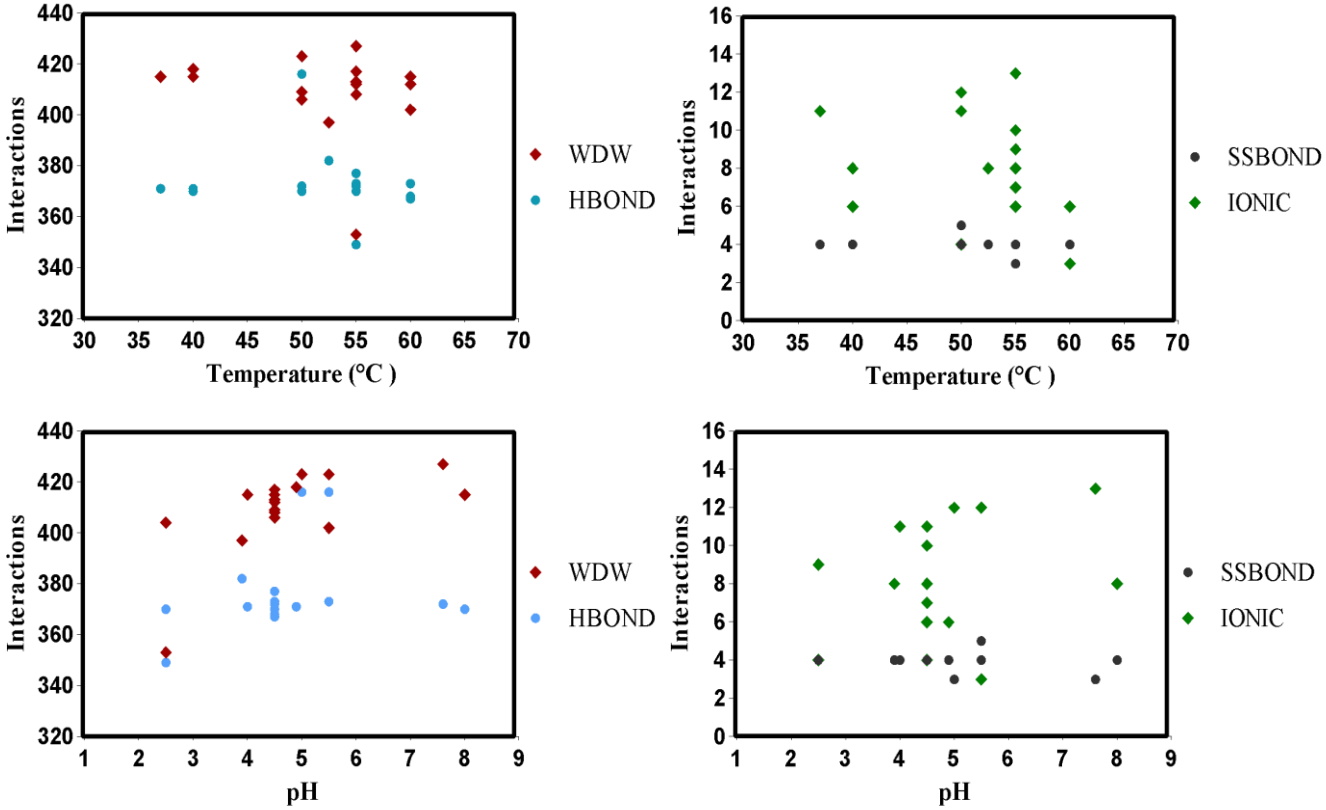
Despite the low identity between them, the three-dimensional structures of cluster 3 phytases presented great structural similarity to each other (Figure 4), which resulted in a low variation in the number of predicted hydrogen (349-416) and VdW interactions (353-427) and disulfide bonds (4-5) (Figure 5). This constancy evidences the non-relevance of the number of these interactions for the biochemical characteristics of these enzymes. On the other hand, the number of predict ionic interactions varied among the evaluated structures (Figure 5), but has not yet been determinant for the different characteristics observed for each enzyme. In other words, the number of those interactions did not influence the catalytic behavior of these phytases.

Figure 4: Structural alignment between cluster 3 sequences, highlighting the structural similarity between them. The spheres in the center represent the substrate (phytic acid).



Source: Authors.

Figure 5: Number of predict interactions and the ideal temperature and pH of the phytases from cluster 3. The blue, red, yellow and black dots represent hydrogen, Van der Waals and ionic interactions and disulfide bonds, respectively.



Source: Authors.

The number of hydrogen, ionic and disulfide bond interactions is believed to have a strong correlation with thermostability (Zhang 2007; Liao, *et al*., 2013; Fei *et al*., 2013; Hesampour *et al*., 2015; Kumar *et al*., 2015; Tan *et al*., 2016; Han *et al*., 2018; Li et al., 2019; Zhang *et al*., 2020). However, here we show that, for these enzymes, this is not true.

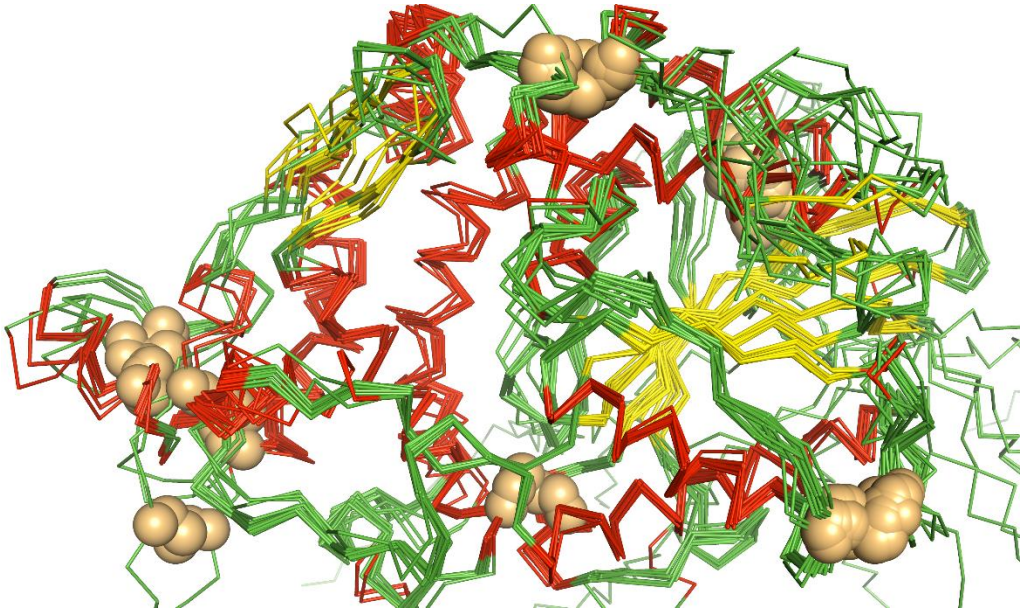
Studies involving one specific enzyme indicates that substitutions of specific amino acid residues are relevant for the catalytic behavior of phytases, depending on the modifications they cause on the interactions that stabilyze the enzyme structure. For example, Fei *et al*., (2013) found a relation between thermostability and ionic interactions. They showed that the Q307D mutation constituted an additional salt bridge and led to a 9.15% increase in phytase thermostability, whereas the loss of this interaction caused (E31Q) reduced its thermostability in 13.96%. The same behavior was described by Tan *et al*., (2016), who observed that the thermal resistance of a phytase was lower with the reduction of the number of ionic interactions. Other studies also describe a relation between the number of hydrogen bonds and the thermostability of individual phytases (Zhang 2007; Hesampour *et al*., 2015). The thermostability of a E. coli phytase increased from 20 to 75% due to increased electrostatic interactions related to the formation of hydrogen bonds resulting from the replacement of charged residues by hydrophobic and aromatic ones (Li *et al*., 2019). However, those studies describe isolated analysis and could not be used for general considerations. Our study considered not only one phytase sequence, but the more representative group of phytases among those described, and is according to Noorbatcha *et al*., (2013) and

Kumar *et al*., 2016, who affirm that the number of interactions does not influence the phytase thermostability, but their location in the three-dimensional structure.

The disulfide bond, due to its well-known stabilization potential, also plays a key role in the conformational stability of the enzyme (Cheng *et al*., 2004; Yang *et al*., 2017). It promotes higher binding energy when compared to the other interactions in enzyme structure, leading to a more rigid and stable enzyme (Wang *et al*., 2022). For this reason, inducing the formation of new disulfide bonds has been a mechanism to improve the thermostability of many enzymes, including phytases (Ebrahimi *et al*., 2016; Tan *et al*., 2016; Yang *et al*., 2017; Navone *et al*., 2021; Wang *et al*., 2022).

On the other hand, although the disulfide bond has a stabilizing effect, in many cases the effect is not significant and may even produce the opposite effect. Cheng *et al*., (2015), successfully introduced a disulfide bond, but without any change in the biochemical properties of the enzyme. This can be attributed to negative effects on flexibility or the already rigidly packed hydrophobic core (Zhang *et al*., 2020). As previously mentioned, the position of the bonds seems to be more related to stability efficiency. This strategy has been proposed by several authors as a strategy to be considered in order to obtain stability, with the exposed loop regions on the surface of the protein normally prioritized, due to its high mobility and flexibility in the structure (Noorbatcha *et al*., 2013; Zhang *et al*., 2020). Some studies have revealed greater thermostability in most mutations found in the loop or surface regions (Kim and Lei 2008; Shivange 2016).

Figure 6: Overlay of predicted structures highlighting the location of disulfide bonds. Lines highlighted in green, red and yellow indicate loop, helix and beta sheets, respectively. Disulfide bonds are shown in spheres and can be seen mainly in the loop region.



Source: Authors.

In the structural comparison between the phytases studied, it was observed that they mostly contain four disulfide bonds, but with very different thermal stability between them (Table 1). However, the disulfide bonds were mostly located in the loop regions or at the point of connection with the helix for some of them (Figure 6). Therefore, although satisfactory results have been observed in the thermal stability of phytases in the cited studies, this method alone does not seem to be decisive in the stability of

phytases. This result is confirmed by Zhang *et al*., (2020) that observed that when evaluating the introduction of disulfide bond in the stability of phytase of *Bacillus licheniformis* WHU, of all the variants analyzed, only one variant had a positive effect on the stability of phytases and all the others had a positive effect on phytase stability or negative effect despite all of them being located in the loop region.

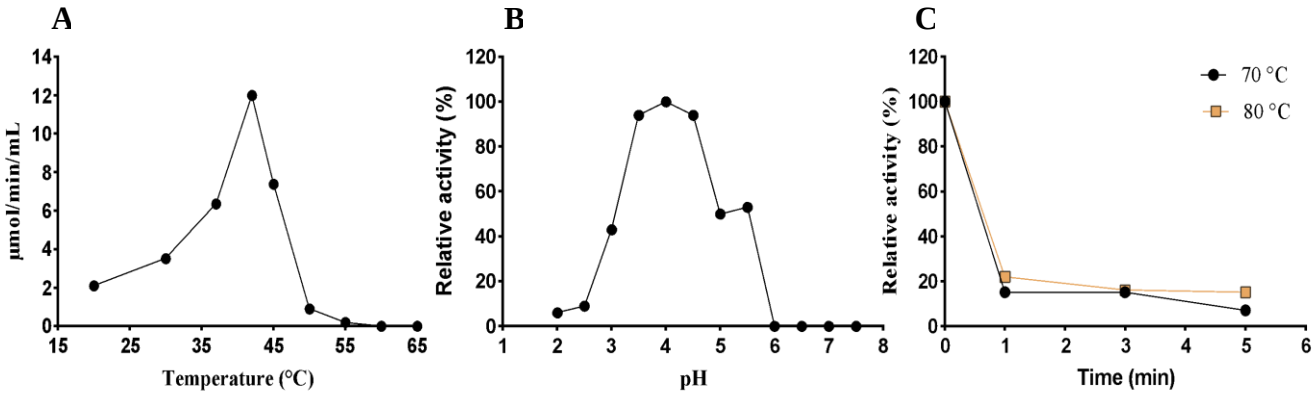
The relationship between the three-dimensional structure of proteins and their biological function is very complex and several factors may be involved to be carefully studied individually or in combination (Shivange *et al*., 2016). Although some studies have been successful in evaluating single or multiple factors, the final effect of residue substitutions on enzyme stability is varied and sometimes undesirable (Navone *et al*., 2021).

3.4 Selection and expression of the phytase KM873028, from Dendroctonus frontalis

The rPhyXT52 gene, isolated from the metagenome of *Dendroctonus frontalis* (Tan *et al*., 2016), was selected among the component sequences of cluster 3 due to the high productivity achieved in a heterologous system, its thermostability (93 % at 100 °C for 15 min) and ability to act in a medium with acidic pHs (3.5 - 5.5). In the present research, the gene was cloned and the protein was expressed in *P. pastoris*, originating a phytase with an ideal pH of activity between 3.5 and 4.5 and an optimal temperature of 42 °C.

However, temperatures above 45 ºC drastically reduced the activity of this enzyme, maintaining a residual activity of only 22 % after just one minute of exposure at 70 and 80 ºC (Figure 7). The phytase selected from the database also lacked stability in a wide pH range to be ideal for application in the animal industry (1.2 to 7.8) (Hamed, 2018). The significant differences observed for this enzyme expressed in a prokaryotic (Tan *et al*., 2016) and a eukaryotic system highlight the need for a deeper understanding about the synthesis and post-translational processing of these enzymes, beyond the influences of the composition of the induction medium, in order to allow rational choices for better production and catalysis.

Figure 7: Activity of the KM873028 phytase expressed in P. pastoris. A: temperature; B: pH and C: thermostability at 70 and 80 °C.



Source: Authors.

4. Conclusion

The design of new, improved phytases demands a deep knowledge on the relation between their structure and function. For this, one strategy relays on the analysis of sequence homology and the development of 3D models in order to find critical points of mutations and interactions. In this work, we showed that phytases are very heterogeneous in relation to their primary sequence, but share similar structures. They were grouped in 5 clusters, where cluster 3 comprises HAPs as the most representative sequences. They share a conserved catalytic motif, beyond eight cysteine residues, which are predict to form disulfide bonds. However, amino acids around the catalytic site were not conserved. On the other hand, contrary to what was believed, the number of interactions was not determinant for the catalytic behavior of these enzymes. Even the primary structure could not be directly related to their function, since phytase activity differed when expressed in prokaryotic or eukaryotic systems.

Future studies on sequence functions and 3D structure, associated with post-translational processing, and the extracellular environment in which they are inserted, are essential for better understanding the determinants of phytase activity. A deeper understanding about the correlation between the structure and function of these enzymes will allow the development of new, improved catalysts with properties ideal for industrial processes.

Declaration of interest

This research was supported by the company Agroceres Multimix Nutrição Animal Ltda and results from a technical cooperation agreement for the development of technology of interest to the company. The author Tarley Araujo Barros is an employee of the company and actively participates in the strategic decisions related to this study.

Acknowledgements

The authors would like to thank the company Agroceres Multimix Nutrição Animal Ltda for funding the research and granting a scholarship to Elizabeth B. E. Pires and Matheus I. B. Aragão. The experiments were carried out in laboratories with equipment acquired through the financing of projects by the entities CAPES, CNPq and FAPEMIG, to which the authors acknowledge.

We are grateful to the Núcleo de Análise de Biomoléculas (NuBioMol) of the Universidade Federal de Viçosa (UFV) for providing the facilities for the conduction of the experiments and data analysis. NuBioMol is financially supported by the following Brazilian agencies: Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (Finep) and Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

Supplementary Information

All supplementary files are temporally available here: https://figshare.com/s/0b9714bd022b8c65a5c8; and will be publish under the following: https://doi.org/10.6084/m9.figshare.20371791

References

Abdulla, J. M., Rose, S. P., Mackenzie, A. M., & Pirgozliev, V. R. (2017). Feeding value of field beans (Vicia faba L. var. minor) with and without enzyme containing tannase, pectinase and xylanase activities for broilers. *Archives of Animal Nutrition*, 71(2), 150-164.

Acquistapace, I. M., Thamopson, E. J., Kuhn, I., Bedford, M. R., Brearley, C. A., & Hemmings, A. M. (2022) Insights to the structural basis for the stereospecificity of the *Escherichia coli* phytase, AppA. International journal of molecular science. 23(11):6346.

Akbarzadeh, A., Dehnavi, E., Aghaeepoor, M., & Amani, J. (2015). Optimization of recombinant expression of synthetic bacterial phytase in *Pichia pastoris* using response surface methodology. *Jundishapur journal of microbiology*, 8(12), e27553.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.

Arredondo, M. A., Casas, G. A., & Stein, H. H. (2019). Increasing levels of microbial phytase increases the digestibility of energy and minerals in diets fed to pigs. *Animal Feed Science and Technology*, 248, 27-36.

Azeke, M. A., Egielewa, S. J., Eigbogbo, M. U., & Ihimire, I. G. (2011). Effect of germination on the phytase activity, phytate and total phosphorus contents of rice (*Oryza sativa*), maize (*Zea mays*), millet (*Panicum miliaceum*), sorghum (*Sorghum bicolor*) and wheat (*Triticum aestivum). Journal of food science and technology*, 48(6), 724-729.

Bohn, L., Meyer, A. S., & Rasmussen, S. (2008). Phytate: impact on environment and human nutrition. A challenge for molecular breeding. *Journal of Zhejiang University Science B*, 9(3), 165-191.

Böhm, K., Herter, T., Müller, J. J., Borriss, R., & Heinemann, U. (2010). Crystal structure of *Klebsiella* sp. ASR1 phytase suggests substrate binding to a preformed active site that meets the requirements of a plant rhizosphere enzyme. *The FEBS Journal*, 277(5), 1284-1296.

Cowieson, A. J., Wilcock, P., & Bedford, M. R. (2011). Super-dosing effects of phytase in poultry and other monogastrics. *World's Poultry Science Journal*, 67(2), 225-236.

Dassa, J., Marck, C., & Boquet, P. L. (1990). The complete nucleotide sequence of the *Escherichia coli* gene appA reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase. *Journal of Bacteriology*, 172(9), 5497-5500.

Fu, D., Huang, H., Luo, H., Wang, Y., Yang, P., Meng, K., ... & Yao, B. (2008). A highly pH-stable phytase from *Yersinia kristeensenii*: cloning, expression, and characterization. *Enzyme and Microbial Technology*, 42(6), 499-505.

Fu, D., Li, Z., Huang, H., Yuan, T., Shi, P., Luo, H., Meng, K., Yang, P & Yao, B. (2011). Catalytic efficiency of HAP phytases is determined by a key residue in close proximity to the active site. *Applied microbiology and biotechnology*, 90(4), 1295-1302.

Gu, W., Huang, H., Meng, K., Yang, P., Fu, D., Luo, H., & Zhan, Z. (2009). Gene cloning, expression, and characterization of a novel phytase from *Dickeya paradisiaca*. *Applied biochemistry and biotechnology*, 157(2), 113-123.

Haefner, S., Knietsch, A., Scholten, E., Braun, J., Lohscheidt, M., & Zelder, O. (2005). Biotechnological production and applications of phytases. *Applied microbiology and biotechnology*, 68(5), 588-597.

Hamed, R. (2018). Physiological parameters of the gastrointestinal fluid impact the dissolution behavior of the BCS class IIa drug valsartan. *Pharmaceutical Development and Technology*, 23(10), 1168-1176.

Han, N., Miao, H., Yu, T., Xu, B., Yang, Y., Wu, Q., Zhang, R & Huang, Z. (2018). Enhancing thermal tolerance of *Aspergillus niger* PhyA phytase directed by structural comparison and computational simulation. *BMC Biotechnology*, 18(1), 1-8.

Hesampour, A., Siadat, S. E. R., Malboobi, M. A., Mohandesi, N., Arab, S. S., & Ghahremanpour, M. M. (2015). Enhancement of thermostability and kinetic efficiency of *Aspergillus niger* PhyA phytase by site-directed mutagenesis. *Applied biochemistry and biotechnology*, 175(5), 2528-2541.

Huber, K., Zeller, E., & Rodehutscord, M. (2015). Modulation of small intestinal phosphate transporter by dietary supplements of mineral phosphorus and phytase in broilers. *Poultry Science*, 94(5), 1009-1017.

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, 26(5), 680-682.

Huang, H., Luo, H., Wang, Y., Fu, D., Shao, N., Wang, G., ... & Yao, B. (2008). A novel phytase from *Yersinia rohdei* with high phytate hydrolysis activity under low pH and strong pepsin conditions. *Applied microbiology and biotechnology*, 80(3), 417-426.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596 (7873), 583-589.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, 20(4), 1160-1166.

Kim, Y. O., Kim, H. W., Lee, J. H., Kim, K. K., & Lee, S. J. (2006). Molecular cloning of the phytase gene from *Citrobacter braakii* and its expression in *Saccharomyces cerevisiae*. *Biotechnology letters*, 28(1), 33-38.

Kryukov, V. S., Glebova, I. V., & Zinoviev, S. V. (2021) Reevaluation of phytase action mechanism in animal nutrition. Biochemistry *(Mosc.)*, 86(1), S152-S165.

Kumar, K., Patel, K., Agrawal, D. C., & Khire, J. M. (2015). Insights into the unfolding pathway and identification of thermally sensitive regions of phytase from *Aspergillus niger* by molecular dynamics simulations. *Journal of Molecular Modeling*, 21(6), 1-13.

Li, J., Li, X., Gai, Y., Sun, Y., & Zhang, D. (2019). Evolution of *E. coli* phytase for increased thermostability guided by rational parameters. Journal of *Microbiology and Biotechnology*, 29(3): 419-428.

Liao, Y., Li, C. M., Chen, H., Wu, Q., Shan, Z., & Han, X. Y. (2013). Site-directed mutagenesis improves the thermostability and catalytic efficiency of *Aspergillus niger* N25 phytase mutated by I44E and T252R. *Applied biochemistry and biotechnology*, 171(4), 900-915.

Luo, H., Huang, H., Yang, P., Wang, Y., Yuan, T., Wu, N., ... & Fan, Y. (2007). A novel phytase appA from *Citrobacter amalonaticus* CGMCC 1696: gene cloning and overexpression in *Pichia pastoris. Current Microbiology*, 55(3), 185-192.

McCormick, K., Walk, C. L., Wyatt, C. L., & Adeola, O. (2017). Phosphorus utilization response of pigs and broiler chickens to diets supplemented with antimicrobials and phytase. *Animal Nutrition*, 3(1), 77-84.

Mitchell, D. B., Vogel, K., Weimann, B. J., Pasamontes, L., & van Loon, A. P. (1997). The phytase subfamily of histidine acid phosphatases: isolation of genes for two novel phytases from the fungi *Aspergillus terreus* and Myceliophthora thermophila. *Microbiology*, 143(1), 245-252.

Mulvenna, C. C., McCormack, U. M., Magowan, E., McKillen, J., Bedford, M. R., Walk, C. L., Oster, M., Reyer, H., Wimmers, K., Fornara, D. A., & Ball, M. E. E. (2022) The growth performance, nutrient digestibility, gut bacteria and bone strength of broilers offered alternative, sustainable diets varying in nutrient specification and phytase dose. *Animals (Basel.)*, 12(13):1669.

Noorbatcha, I. A., Sultan, A. M., Salleh, H. M., & Amid, A. (2013). Understanding thermostability factors of *Aspergillus niger* PhyA phytase: a molecular dynamics study. *The protein journal*, 32(4), 309-316.

Navone, L., Vogl, T., Luangthongkam, P., Blinco, J. A., Luna-Flores, C. H., Chen, X., Hellens, J. V., Mahler, S & Speight, R. (2021). Disulfide bond engineering of AppA phytase for increased thermostability requires co-expression of protein disulfide isomerase in *Pichia pastoris*. *Biotechnology for biofuels*, 14(1), 1-14.

Nayeem, A., Chiang, S. J., Liu, S. W., Sun, Y., You, L., & Basch, J. (2009). Engineering enzymes for improved catalytic efficiency: a computational study of site mutagenesis in epothilone-B hydroxylase. *Protein Engineering, Design & Selection*, 22(4), 257-266.

Niu, C., Yang, P., Luo, H., Huang, H., Wang, Y., & Yao, B. (2017). Engineering of *Yersinia* phytases to improve pepsin and trypsin resistance and thermostability and application potential in the food and feed industry. *Journal of agricultural and food chemistry*, 65(34), 7337-7344.

Pal Roy, M., Mazumdar, D., Dutta, S., Saha, S. P., & Ghosh, S. (2016). Cloning and expression of phytase appA gene from *Shigella* sp. CD2 in *Pichia pastoris* and comparison of properties with recombinant enzyme expressed in *E. coli*. *PloS one*, 11(1), e0145745.

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, Cap3, Unid 3.1.

Piovesan, D., Minervini, G., & Tosatto, S. C. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic acids research*, 44(1), 367-374.

Pires, E. B. E., de Freitas, A. J., Salgado, R. L., Guimarães, V. M., Pereira, F. A., & Eller, M. R. (2019). Production of fungal phytases from agroindustrial byproducts for pig diets. *Scientific Reports*, 9(1), 1-9.

Pramanik, K., Kundu, S., Banerjee, S., Ghosh, P. K., & Maiti, T. K. (2018). Computational-based structural, functional and phylogenetic analysis of Enterobacter phytases. *3 Biotech*, 8(6), 1-12.

Selim, S., Abdel-Megeid, N. S., Khalifa, H. K., Fakiha, K. G., Majrashi, K. A., & Hussein, E. (2022) Efficacy of various feed additives on performance, nutrient digestibility, bone quality, blood constituents, and phosphorus absorption and utilization of broiler chickens fed low phosphorus diet. *Animals (Basel.)*, 12(14):1742.

Shao, N., Huang, H., Meng, K., Luo, H., Wang, Y., Yang, P., & Yao, B. (2008). Cloning, expression, and characterization of a new phytase from the phytopathogenic bacterium *Pectobacterium wasabiae* DSMZ 18074. *Journal of microbiology and biotechnology*, 18(7), 1221-1226.

Shivange, A. V., Hoeffken, H. W., Haefner, S., & Schwaneberg, U. (2016). Protein consensus-based surface engineering (ProCoS): a computer-assisted method for directed protein evolution. *BioTechniques*, 61(6), 305-314.

Shivange, A. V., Roccatano, D., & Schwaneberg, U. (2016). Iterative key-residues interrogation of a phytase with thermostability increasing substitutions identified in directed evolution. *Applied microbiology and biotechnology*, 100(1), 227-242.

Singh, B., Sharma, K. K., Kumari, A., Kumar, A., & Gakhar, S. K. (2018). Molecular modeling and docking of recombinant HAP-phytase of a thermophilic mould Sporotrichum thermophile reveals insights into molecular catalysis and biochemical properties. *International journal of biological macromolecules*, 115, 501-508.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, 38(7), 3022-3027.

Tan, H., Wu, X., Xie, L., Huang, Z., Peng, W., & Gan, B. (2016a). Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Applied microbiology and biotechnology*, 100(5), 2225-2241.

Tan, H., Miao, R., Liu, T., Cao, X., Wu, X., Xie, L., & Gan, B. (2016b). Enhancing the Thermal Resistance of a Novel Acidobacteria-Derived Phytase by Engineering of Disulfide Bridges S. *Journal of Microbiology and Biotechnology*, 26(10), 1717-1722.

Vieira, M. S., Pereira, V. V., da Cunha Morales Álvares, A., Nogueira, L. M., Lima, W. J., Granjeiro, P. A., & Galdino, A. S. (2019). Expression and Biochemical Characterization of a Yersinia intermedia Phytase Expressed in Escherichia coli. *Recent Patents on Food, Nutrition & Agriculture*, 10(2), 131-139.

Taussky, H. H., & Skorr, E. (1953). A microcolorimetric method for the determination of inorganic phosphorus. *The Journal of Biological Chemistry*, 202(2), 675–685.

Wang, H., Chen, H., Li, Q., Yu, F., Yan, Y., Liu, S., & Tan, J. (2022). Enhancing the thermostability of transglutaminase from *Streptomyces mobaraensis* based on the rational design of a disulfide bond. *Protein Expression and Purification*, 195, 106079.

Yang, W., Yang, Y., Zhang, L., Xu, H., Guo, X., Yang, X., & Cao, Y. (2017). Improved thermostability of an acidic xylanase from *Aspergillus sulphureus* by combined disulphide bridge introduction and proline residue substitution. *Scientific reports*, 7(1), 1-9.

Yao, M. Z., Wang, X., Wang, W., Fu, Y. J., & Liang, A. H. (2013). Improving the thermostability of *Escherichia coli* phytase, appA, by enhancement of glycosylation. *Biotechnology letters*, 35(10), 1669-1676.

Yi, Z., & Kornegay, E. T. (1996). Sites of phytase activity in the gastrointestinal tract of young pigs. *Animal Feed Science and Technology*, 61(1-4), 361-368.

Zhang, G. Q., Dong, X. F., Wang, Z. H., Zhang, Q., Wang, H. X., & Tong, J. M. (2010). Purification, characterization, and cloning of a novel phytase with low pH optimum and strong proteolysis resistance from *Aspergillus ficuum* NTG-23. *Bioresource technology*, 101(11), 4125-4131.

Zhang, Z., Yang, J., Xie, P., Gao, Y., Bai, J., Zhang, C.,Liu, L., Wang, Q & Gao, X. (2020). Characterization of a thermostable phytase from *Bacillus licheniformis* WHU and further stabilization of the enzyme through disulfide bond engineering. *Enzyme and Microbial Technology*, 142, 109679.

Zhao, W., Xiong, A., Fu, X., Gao, F., Tian, Y., & Peng, R. (2010). High level expression of an acid-stable phytase from *Citrobacter freundii* in *Pichia pastoris. Applied biochemistry and biotechnology*, 162(8), 2157-2165.

Zinin, N. V., Serkina, A. V., Gelfand, M. S., Shevelev, A. B., & Sineoky, S. P. (2004). Gene cloning, expression and characterization of novel phytase from *Obesumbacterium proteus*. FEMS microbiology letters, 236(2), 283-290.