Archives of Razi Institute, Vol. 73, No. 3 (2018) 229-237

Short Communication

Bioinformatics Analysis of Upstream Region and Protein Structure of Fungal Phytase Gene

Gholizadeh [1,*], M., Nassiri [2], M., Saberi [3], M.R., Haddad-Mashadrizeh [4], A.A.

1. Department of Animal Sciences, Faculty of Animal and Food Sciences, Khuzestan Ramin Agriculture and Natural Resources University, Mollasani, Ahvaz, Iran
2. Department of Genetics, Faculty of Animal Sciences, Ferdowsi University of Mashhad, Mashhad, Iran
3. Department of Medical Chemistry, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran
4. Cell and Molecular Biotechnology Research Group, Institute of Biotechnology, Ferdowsi University of Mashhad, Mashhad, Iran

Received 17 February 2017; Accepted 08 July 2017
Corresponding Author: ma.gholizade@gmail.com

ABSTRACT

Phytase increases the bioavailability of phytate phosphorus in seed-based animal feeds and reduces the phosphorus pollution of animal waste. Since most animal feeds for pellets are heated up to 65-80 °C, the production of a thermostable structure for phytase can be useful. In this study, we sought to perform bioinformatics analysis of the upstream region and protein structure of fungal phytase to improve its expression and thermostability properties. We used bioinformatics methods such as similarity search, multiple alignment, statistical analysis of physicochemical properties of amino acids, pattern recognition, and protein modeling to find out the effective factors in heat resistance of phytase. Change in Gibbs free energy (ΔG) of the best pattern promoter resulting from the interaction between RNA polymerase and the promoter sequences of modified genes of phytase was equal to -9 kcalmol-1, which is lower compared to other interactions. The evaluation of the three-dimensional structure of new phytases showed that amino acid substitutions aimed at improving thermostability did not change the form and structure of the protein. The results of Prochek, Whatcheck, and ERRAT for structural analysis and verification were 84, 72, and 70, respectively, that were satisfactory.

Keywords: Fungal phytase, Bioinformatics, Homology modelling, Molecular docking, Upstream region analysis

Analyse Bio-informatique de la Région en Amont du Gène de Phytase Fongique Ainsi que de sa Structure Protéique

Résumé: La phytase augmente la biodisponibilité du phytate phosphore dans les graines dédiées à l'alimentation animale et réduit la pollution au phosphore des déchets animaliers. Étant donné que la plupart des aliments pour animaux destinés à la production de granulés sont chauffés à 65-80 °C, la production d'une structure thermostable pour la phytase peut être utile. Notre étude avait pour objectif l'analyse bioinformatique de la région en amont du gène de la phytase fongique ainsi que la sturucture de la protéine codée, afin d'améliorer ses propriétés d'expression et de thermostabilité. A cet effet, des méthodes bio-informatiques telles que la recherche de similarité, l'alignement multiple, l'analyse statistique des propriétés physicochimiques des acides aminés, la reconnaissance des formes et la modélisation des protéines pour déterminer les facteurs de résistance à la chaleur de la phytase ont été utilisées. La modification (ΔG) dans l'énergie libre de Gibbs du meilleur promoteur de modèle résultant de l'interaction entre l'ARN polymérase et les séquences promotrices de gènes modifiés de la phytase était égal à -9 kcalmol-1, ce qui est inférieur aux autres connexions. L'évaluation de la structure tridimensionnelle de nouvelles phytases a montré que les substitutions d'acides aminés visant à

améliorer la thermostabilité n'ont pas modifié la forme et la structure de la protéine. Les résultats obtenus par Prochek, Whatcheck et ERRAT pour l'analyse structurelle étaient satisfaisants et la vérification était respectivement de 84, 72 et 70.

Mots-clés: Phytase Fongique, Bio-informatique, Modélisation Homologique, Amarrage Moléculaire, Analyse des Régions en Amont

INTRODUCTION

The phytases (myo-inositol hexakisphosphate phosphohydrolases; EC 3.1.3.8 and 3.1.3.26) are a subfamily of high-molecular-weight histidine acid phosphatases that are capable of releasing phosphates from phytic acid (myo-inositol 1, 2, 3, 4, 5, 6- hexakis phosphate) in animal feeds of plant origin (Marlida et al., 2010; Stefan et al., 2005). The exploitation of phytases in the feed industry has been hampered by high production costs, low stability, and poor specificity of phytases (Boder, 2012). One of the desirable properties of phytases in the feed industry is high thermostability, which enables them to withstand high temperatures (65–80 °C) in the animal feed pelleting process (Tengand et al., 2010). Most studies on this enzyme have been focused to improve expression rate and thermostability of several microbial phytases. Lim et al. (2001) showed a high level of fungal phytase in Saccharomyces cerevisiae; in that study, various expression vectors were constructed with different combinations of promoters, translation enhancers, signal peptides, and terminators. Three different promoters fused to the phytase gene (phyA) from Aspergillus niger were tested. Among the recombinant vectors, pGphyA-6 containing the glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter, the Ω sequence, RAmy1A, and GAL7 terminator had the highest phytase activity in a culture filtrate. Promdonkoy et al. (2008), to improve thermostability properties of phytase, two thermostable phytases were obtained from Thai isolates of Aspergillus japonicus BCC18313 (TR86) and Aspergillus niger BCC18081 (TR170) and then cloned and transferred them to Pichia pastoris. Optimal phytase activity for both recombinant phytases was at 50 °C and pH 5.5, showing that this substance is suitable for use in dietary supplements and is a viable alternative for the current commercial phytases. To increase expression and thermostability of phytase, Tran et al. (2010) isolated a thermostable phytase from Bacillus spp. (MD2), and then cloned and expressed it in *Escherichia coli*. Their results showed that recombinant phytases are highly expressed in the presence of calcium chloride and have high stability at temperatures up to 100 °C. Although the properties of phytases vary, no enzyme is ideal for field applications (Tengand et al., 2010). In this study, to obtain enzymes with improved properties, we used two strategies: promoter engineering to achieve high expression and protein engineering to enhance the thermostability of this enzyme. Many regulating parameters have been studied for gene expression, but we focused on RNA polymerase binding to promoter sequence. For increasing the thermostability of proteins, we used amino acid substitution strategies (Mullaney et al., 2010). This approach is based on the comparison of amino acid sequences of homologous proteins and the subsequent calculation of a consensus amino acid sequence using one of the available standard programs (Teng et al., 2010). Herein, we aimed to improve some of the properties of fungal phytase gene by bioinformatics analysis of upstream region and protein structure. As a popular subject in bioinformatics, we used many of web-based software and statistical methods.

MATERIALS AND METHODS

In silico analysis of upstream region. To investigate and analyze different phyA upstreams, we obtained the

nucleotide sequences of this gene from the NCBI and Gene bank database. Then, upstream sequences of all the fungal phytases (phyA) were analyzed with BLST-2, NsiteH, and CpG Plot software. BLAST tool was used to find conserved regions of the phyA that match with consensus regions of the upstream from other subspecies of fungal phyA. NsiteH software searches for statistically non-random motifs of known transcription regulatory elements in a single DNA sequence (D'haeseleer, 2006). A predicted motif was considered statistically significant if the expected number of such motifs was less than a given threshold and the total number of identified motifs was greater than the upper limit of 95% confidence interval. Various methods and tools have been proposed to predict CpG islands, and in this study, we used CpGPlot program that reports all CpG-rich regions (Deaton and Bird, 2011). Transcription is initiated when RNA polymerase recognizes the duplex promoter DNA in the closed complex, thus, we focused on the interaction between holoenzyme RNA pol II and promoter sequence (Brewster et al., 2012; Marr et al., 2004). In this model, RNA pol II was docked to the 10 proposed promoters (with different substitutions in sequence) with expert interface of DNA docking software. This software is an information-driven flexible docking tool for the modeling of biomolecular complexes (Shaikh and Jayaram, 2007; Gupta et al., 2007).

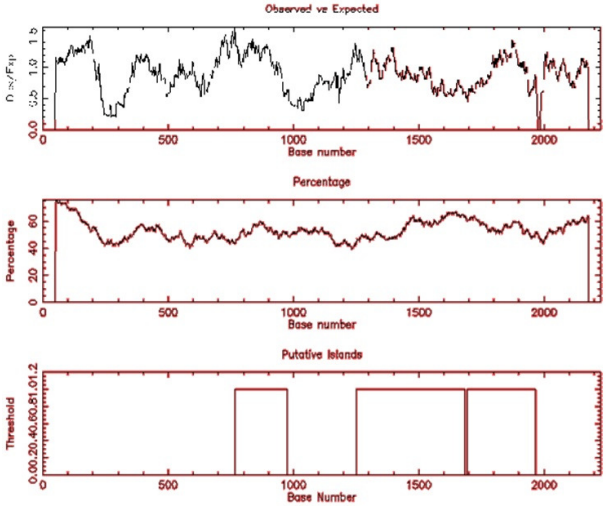
Substitution of amino acids to provide thermostability. To determine the structure of a thermostable fungal phytase, amino acid sequences of 40 thermostable enzymes, 30 random enzymes, and 30 sequences of phytase family proteins were extracted from the UniProt database. A thorough literature review was carried out to find the previously mentioned proteins. InterProScan was utilized to investigate the possible common signatures within the thermostable proteins (Jones et al., 2014; Finn et al., 2017). Biochemical and structural properties of three groups of proteins (i.e., heat-resistant, phytase, and random) were analyzed with ProtScal web service (Gasteiger et al., 2005). ProtParam was used to identify some physicochemical properties, including molecular weight, theoretical pI, amino acid composition, atomic composition, instability index, and grand average of hydropathicity (Konga et al., 2016). Statistical analysis was performed using SPSS, version 16. Motif sites of phytase with the accession number of O00092 were achieved through motif scanning (Pagni et al., 2007) and motif search, and the secondary structures were extracted from UniProt database. Protein models were finally created using MOE 2008.10 software (Deschênes and Sourial, 2007). Energy was optimized with HyperChem 7.1 software and assessed via SAVS website (Pontius et al., 1996). Signal peptide cleavage site for the deduced amino acid sequence was predicted using Signal P 4.1 server.

RESULTS AND DISCUSSION

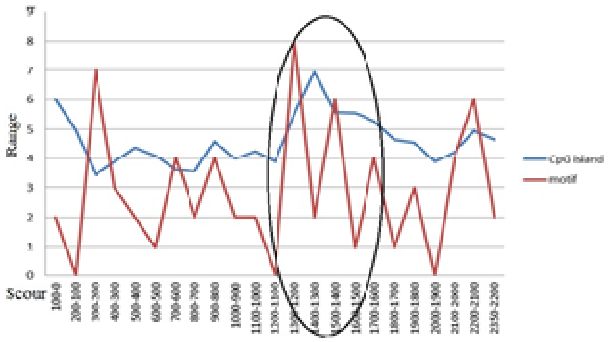
Upstream of phytase gene analysis. The identification of promoters and their regulatory elements is one of the major challenges in bioinformatics, which integrates comparative, structural, and functional genomics. Gene expression is to a large degree regulated by the specific binding site sequence motifs in gene promoter regions (Lis and Walther, 2016). Motifs are short, recurring patterns in DNA that are presumed to have a biological function. They often indicate sequence-specific binding sites for proteins such as RNA polymerase (D'haeseleer, 2006). Many different approaches have been developed to detect conserved motifs in upstream of the gene, however, attempts for unambiguous identification of regulatory elements have not been successful. In this study, we used NsiteH software for predicting motifs in the upstream region. Shahmuradov and Solovyev (2015) employed the NsiteH software for the identification of motif in promoter sequences. This tool is widely used by researchers and has been cited in 200 articles. NsiteH is accessible through the Softberry and KAUST Bioinformatics WEB servers (www.softberry.

com and www.molquest.kaust.edu.sa). This software was chosen to find sequence upstream regions with the highest density of regulatory elements. Upstream of the Aspergillus niger var. awamori species (L02421) contains the highest motif/length proportion within all fungal species (D'haeseleer, 2006). Thus, this upstream sequence was selected for the next analysis.

CpG islands analysis in upstream region: Despite the sequence diversity among promoters, genes transcribed by RNA pol II can be classified into different groups according to the distribution of CpG dinucleotides across their upstream regions, where the frequency of CpGs is approximately 10 times higher than the genome average called CpG islands (CpG islands or CGIs) (Antequera 2003; Hackenberg et al., 2010). CGIs are associated with thousands of genes that are active in all cell types of the organism, and they are expected to contain many binding sites for ubiquitous transcription factors (Deaton and Bird, 2011). Given their location in promoters, CGIs may play important roles in the regulation of gene expression (Hackenberg et al., 2010). Since CpG islands are the most reliable feature for promoter prediction and they are particularly suited for replication initiation, it is suggested that CpG islands might simultaneously serve as promoters and replication origins (Antequera, 2003). Sarda et al. (2017) through extensive analysis showed that transcription is initiated by a distal upstream CpG island located several kilobases away, which functions as an alternative promoter. Also, based on RNA pol II localization data, they found strong evidence of transcription initiation at upstream CpG island. The CpG islands of upstream fungal phytase sequences were detected by EMBOSS CpG plot software. The algorithms of these programs have been developed to search for the presence of CpG islands in the 5′ region of genes. In this analysis, only sequence of Myceliophythora thermophile (Acc: U59806) contained CpG islands (Shown in Graph 1). A demonstration of motif with high density areas and rich CpG islands in two fungal phytase proteins was showed in graph 2. Finally, 10 sequence lengths of 400 bp with regions rich in regulatory elements were suggested (Graph 2).



Graph 1. Sequence of thermostable phytase (Acc: O00092) before replacing the amino acid; the motif shown in yellow was detected by motif scan and motif search software. Secondary structure and active site information were obtained from UniProt database. Beta strand is shown in blue, helix in red, turn in green, and active sites in violet.



Graph 2. Interaction between high density motif areas and regions rich in CpG islands in U59806 sequence

New promoter for fungal phytase: The level of gene expression exhibited by a cell can be targeted at multiple levels along the path from DNA to protein (Setny et al., 2012). Key biological tuning variables include the copy number of transcription factors, strength of their binding sites, strength of RNA polymerase binding, strength of ribosomal binding sites, and degradation rates of proteins (Brewster et al.,

2012). Such binding energies are at the heart of the molecular interactions, hence precise energy control of this interaction and binding is an essential prerequisite for quantitative control of transcription (Marr et al., 2004). Bioinformatics docking methods can provide structural models of protein-DNA complex, where it is difficult or impossible to obtain an experimental complex structure. Bioinformatics docking is a predictive method based on the structures of the individual partners (Brewster et al., 2012). Several computational and experimental methods have been developed and used extensively for the prediction of the interaction between RNA pol II and promoter. Hendrix et al. (2016) studied the type of promoter elements associated with RNA pol II binding in Drosophila embryogenesis. They presented a combined computational and experimental (permanganate footprint assays) analysis of promoters to determine how they bind RNA pol II in early Drosophila embryogenesis. They found that GC-rich sequences in core promoter increase the bindings between PR promoter and RNA pol II. Minetoki et al. (2003) indicated that Aspergillus species are widely used to produce useful enzymes. They found that there are three highly conserved sequences (designated regions I, II, and III) in promoter regions of the genes encoding α-amylase, gluco amylase, and α-glucosidase of A. oryzae. They demonstrated the advantage of over-expression of genes controlled by the improved promoter, especially using region III in various Aspergillus species. Therefore, 10 new promoters for phytase gene were proposed using the addition of new regulatory sequences to upstream of Aspergillus niger var. awamori. Then, to select the best promoter, we investigated the interaction between these 10 promoters and RNA pol II using DNA docking software. The sequence of RNA pol II extracted from PDB database (PDB ID: 3FKI) and input file of software were prepared. The most stable binding between the proposed sequences was chosen as promoter. Change ΔG in Gibbs free energy of this interaction was equal to -9 kcal.mol-1, which is lower than that in other connections.

Signal peptide: Signal peptide is a short (5-30 amino acids long) peptide present at the N-terminus of the majority of newly synthesized proteins that is destined towards the secretory pathway (Lum et al., 2011). With extracellular secretory enzyme formation, product recovery is facilitated because time- and energy-consuming cell disruption and downstream processing are avoided (Hemmerich et al., 2016). As phytase is a secretary enzyme, we seek to find the best signal peptide sequences for the further production of this enzyme. Searching in FunSecKB database (Lum and Min, 2011) identified phytase signal peptides. Signal peptide sequences were analyzed with SignalP version 4 software (Liu et al., 2016). Since the graphs with higher S-, C-, and Y-scores are supposed to be optimal in showing the relationship between the quality of the signal peptide and the secretory effect (Emanuelsson et al., 2007), it was suggested that U59805 peptide (Aspergillus terreus 9A1), presenting the highest score in all the three indicators of S-, C-, and Y-scores, could be the most favorable signal peptide in this regard (Petersen et al., 2011). U59805 was selected for further procedures in phytase gene improvement.

Prediction of the thermostable phytase structure. Physicochemical properties and structural classification of protein phytase: Analysis of amino acid sequences by InterProScan software does not provide a signature shared among all proteins (Finn et al., 2017). Therefore, we searched other patterns of heat resistant proteins that were common among thermostable proteins to differentiate them from normal proteins. Thermostable proteins and random proteins were compared by ProtScale software in terms of nine biochemical and structural features including bulkiness, polarity, hydrophobicity, beta-sheet, alpha-helix, antiparallel beta-strand, percent of accessible residues, and average area buried (Liu et al., 2016). SPSS version 16 was used for the determination of significant differences between the two groups of proteins. Mann-

Whitney U test showed that bulkiness in thermostable proteins was significantly more than that in random proteins, and average area buried in thermostable proteins was less than that in random proteins. There were no significant differences in other features. Teng et al. (2011) showed that to study protein thermostability, bulkiness was the most important structural feature of proteins. Providing a standard amino acids composition for thermostable proteins: To confirm the results of the ProtScale analysis and specify amino acids that must be replaced in phytase protein, composition of amino acids for each protein was determined by the ProtParam program (Konga et al., 2016). Amino acids with bulky side chains include valine, tryptophan, leucine, isoleucine, phenylalanine, proline, and tyrosine. The results showed that leucine in thermostable proteins is significantly more than in random proteins. Teng et al. (2010) studied two thermostable proteins. They showed that both protein groups were rich in glutamine and glycine in their amino acid composition. For amino acid substitution, we first found total thermostable proteins, and then analyzed them by ProtParam and ProtScale software (Table 1).

Amino acid substitution in phytase sequence: The crystal structure of Aspergillus fumigatus phytase at 2 °5 Å resolution revealed three distinct domains, including a large a-helical domain, a β-sheet domain, and a small a-helical domain. The large a-helical and the small a-helical domains contain five a-helixes and four a-helixes, respectively, and the β-sheet domain contains eight β-sheets. The active site was located at the interface between the two domains (Yao et al., 2012). The substitution of the amino acid sequence of phytase in places other than motif, active site, beta strand, helix, and turn were tested. Motifs of phytase were found using motif search and motif scan (Acc: O00092). Secondary structure and active sites were obtained from the data on the UniProt database (Figure 1). According to the survey results and characteristics of the amino acid sequence (i.e., polarity, hydrophobic, charged amino acid, and side chain), we concluded that in phytase sequence, phenylalanine should be substituted with isoleucine. Homology modelling, structural analysis and verification of new phytase: After the substitution of amino acids in protein, a three-dimensional structure model was constructed using the MOE 2008.10 software.

Table 1. Amino acids composition for thermo-stable proteins and theromo-stable phytase (Acc: O00092) was determined by the ProtParam program.

thermo-stable proteins		thermo-stable phytase	
Ala	9.7	Ser	10.3
Leu	8.3	Leu	9.5
Glu	8	Ala	9.2
Iys	7.1	Gly	7.1
Ser	6.8	Thr	6.7
Val	6.6	Val	6.5
Gly	6.3	Phe	5.8
Thr	5.6	Lys	5.6
Ile	5.5	Asp	5.4
Asp	5.1	Pro	4.7
Asn	5	Arg	4.3
Cys	4.6	Glu	4.3
Pro	4.5	Asn	4.1
Gln	3.8	Tyr	3.9
Phe	3.6	Gln	3.4
Arg	3	Ile	3
Met	2.4	Cys	2.2
Try	2.3	His	1.7
His	1.1	Met	1.3
Trp	0.7	Trp	1.1

Energy levels of this protein were reported at 736 kcal.mol-1. The model of tertiary structure of phytase was built using the best sequence homology. Then, for checking the functionality of phytase, tertiary structure of this protein was compared with natural phytase presented in PDB website (PDB ID: 1QWO). Despite the amino acids changes in the designed phytase, the structure of the protein remained intact. The structural model was subjected to energy minimization using the HyperChem software. After minimization of energy in the software, energy was calculated at 347 kcal.mol-1. Structural analysis and verification were carried out using a variety of tools, like stereo chemical quality assessment by Prochek and Whatcheck and analysis of statistics of non-bonded interactions between different atom types by ERRAT. SAVS web service was used to evaluate the models. The results of Prochek, Whatcheck, and ERRAT (84, 72, and 70, respectively)

were satisfactory. To promote the expression of proteins in recombinant systems, putting all the parts of a gene in a molecular structure is not possible (Cao at al., 2012). It has been observed that the expression of most genes in these conditions is significantly lower than in normal condition. We were investigating the best factors that play a critical role in the least possible gene sequence.

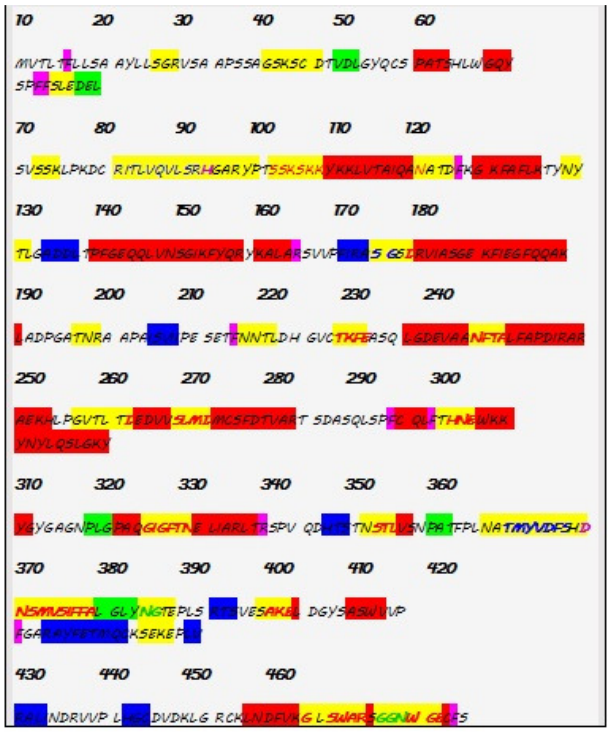


Figure 1. The output of EMBOSS CpG Plot Software with the accession number U59806; three regions containing CpG islands of 975-768, 1683-1251m, and 1966-1694 are clearly visible in the last graph

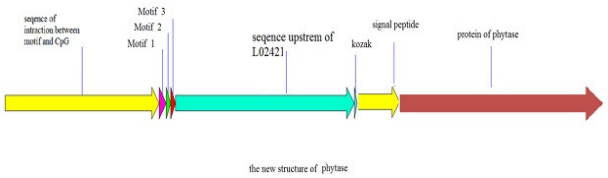


Figure 2. The proposed structure of fungal phytase gene with improved properties of expression and heat-resistance drawn using the SimVector 5.5 software

To define most regulatory factors in a gene, the non-coding segments were placed beside the coding segments of our gene of interest due to the largeness of sequences in in-vitro assays. However, detecting gene expression regulation would be very time consuming and costly (Wasserman et al., 2004). Therefore, using bioinformatics analysis, we detected the regions that have regulatory potential and proposed a protein sequence with appropriate characteristics. To overcome low expression and thermostability, bioinformatics methods were utilized to detect the area with high regulatory potential. These methods helped us to design genes that could be useful candidates for the expression of proteins with improved characteristics and perhaps functions. We also proposed a heat resistant phytase with logical and meaningful replacement of some amino acids in natural phytases. The mentioned changes did not affect the 3D structure of the phytase. Finally, to confirm our findings, laboratory procedures and clinical studies seem to be necessary. The proposed structure of the phytase gene with improved properties was drawn using the SimVector 5.5 software (Figure 2).

Ethics

I hereby declare all ethical standards have been respected in preparation of the submitted article.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Abreu, E.F., Aragão, F.J., 2007. Isolation and characterization of a myo-inositol-1-phosphate synthase gene from yellow passion fruit (Passiflora edulis f. flavicarpa) expressed during seed development and environmental stress. Ann Bot 99, 285-292.

Antequera, F., 2003. Structure, function and evolution of CpG island promoters. CMLS 60, 1647-1658.

Berka, R.M., Rey, M.W., Brown, K.M., Byun, T., Klotz, A.V., 1998. Molecular characterization and expression of a phytase gene from the thermophilic fungus Thermomyces lanuginosus. Appl Environ Microbiol 64, 4423-4427.

Bischof, J.C., He, X., 2006. Thermal stability of proteins. Ann N Y Acad Sci 1066, 12-33.

Blundell, T., Elliott, G., Gardner, S., Hubbard, T., Islam, S., Johnson, M., et al., 1989. Protein engineering and design. Philosophical Transactions of the Royal Society of London B: Biological Sciences 324, 447-460.

Boder, E.T., 2012. Protein engineering: tighter ties that bind. Nature 484, 463-464.

Bortoluzzi, S., Coppe, A., Bisognin, A., Pizzi, C., Danieli, G.A., 2005. A multistep bioinformatic approach detects putative regulatory elements in gene promoters. BMC bioinformatics 6, 121.

Brewster, R.C., Jones, D.L., Phillips, R., 2012. Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. PLoS Comput Biol 8, e1002811.

Cao, Y., Jiao, R., Xia, Y., 2012. A strong promoter, PMagpd, provides a tool for high gene expression in entomopathogenic fungus, Metarhizium acridum. J Biotechnol Lett 34, 557-562.

Corrêa, T.L.R., de Queiroz, M.V., de Araújo, E.F., 2015. Cloning, recombinant expression and characterization of a new phytase from Penicillium chrysogenum. J Microbiol Res 170, 205-212.

Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. Genes Dev 25, 1010-1022.

Deschênes, A., Sourial, E., 2007. Ligand Scaffold Replacement using MOE Pharmacophore Tools. J. Chem. Comp. Group.

D'haeseleer, P., 2006. What are DNA sequence motifs? Nature biotechnology 24, 423-425.

Emanuelsson, O., Brunak, S., Von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature protocols 2, 953-971.

Finn, RD., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y and et al., 2017. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. Vol 4;45(D1):D190-D199. PubMed PMID: 27899635

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.e., Wilkins, M.R., Appel, R.D., et al., 2005. Protein identification and analysis tools on the ExPASy server, Springer.

Gupta, A., Gandhimathi, A., Sharma, P., Jayaram, B., 2007. ParDOCK: An all atom energy based Monte Carlo docking protocol for protein-ligand complexes. Protein Pept Lett 14, 632-646.

Hackenberg, M., Barturen, G., Carpena, P., Luque-Escamilla, PL., Previti, C., Oliver, JL., 2010. Prediction of CpG-island function: CpG clustering vs. sliding-window methods. BMC Genomics. Vol 26;11:327. doi: 10.1186/1471-2164-11-327. PubMed PMID:20500903.

Haefner, S., Knietsch, A., Scholten, E., Braun, J., Lohscheidt, M., Zelder, O., 2005. Biotechnological production and applications of phytases. Appl. Microbiol. Biotechnol 68, 588-597.

Hemmerich, J., Rohe, P., Kleine, B., Jurischka, S., Wiechert, W., Freudl, R., Oldiges, M., 2016. Use of a Sec signal peptide library from Bacillus subtilis for the optimization of cutinase secretion in Corynebacterium glutamicum. Microb Cell Fact. Vol 7;15(1):208. PubMed PMID: 27927208; PubMed Central PMCID: PMC5142396.

Hendrix, DA., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., Levine, M.S., 2016. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc Natl Acad Sci U S A. Vol 3;105(22):7762-7.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236-1240.

King, D.C., Taylor, J., Zhang, Y., Cheng, Y., Lawson, H.A., Martin, J., et al., 2007. Finding cis-regulatory elements using comparative genomics: some lessons from ENCODE data. Genome Res 17, 775-786.

Konga, S.L., Abdullaha, S.N.A., Hoc, C.L., Amiruddind, M.D., 2016. Molecular cloning, gene expression profiling and in silico sequence analysis of vitamin E biosynthetic genes from the oil palm. J Plan Gene 5, 100–108.

Lim, Young-Yi., Eun-Ha, Park., Ji-Hye, Kim., Seung-Moon, Park., Hyo-Sang, Jang., Youn-Je, Park., Sewang, Yoon., Moon-Sik, Yang., Dae-Hyuk, Kim., 2001. Enhanced and Targeted Expression of Fungal Phytase in Saccharomyces cerevisiae. Journal of Microbiology and Biotechnology. 11, 915-92.

Lin, H., Chen, W., 2011. Prediction of thermophilic proteins using feature selection technique. J Microbiol Methods 84, 67-70.

Liu, F., Zhang, X., Lu, C., Zeng, X., Li, Y., Fu, D., et al., 2015. Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. J Exp Bot 313.

Liu, Wan. Jing, Lv, Hai. zhou., He, Liu., Song, Jing.yuan., Sun, Chao., Luo, Hong.mei., Chen, Shi.lin., 2016. Cloning and Bioinformatic Analysis of HMGS and HMGR Genes from Panax notoginseng. J Chinese Herb Med 8, 344-351.

Lis, M., Walther, D., 2016. The orientation of transcription factor binding site motifs in gene promoter regions: does it matter?. BMC Genomics 17, 310.

Lum, G., Min, X.J., 2011. FunSecKB: the fungal secretome knowledgebase. Database 2011, bar001.

Lum, G., Min, X.J., 2011. FunSecKB: the fungal secretome knowledgebase. Database 2011, bar001.

Marlida, Y., Delfita, R., Adnadi, P., Ciptaan, G., 2010. Isolation, characterization and production of phytase from endophytic fungus its application for feed. Pak J Nutr 9, 471-474.

Marr, M.T., Roberts, J.W., Brown, S.E., Klee, M., Gussin, G.N., 2004. Interactions among CII protein, RNA polymerase and the λ PRE promoter: contacts between RNA polymerase and the–35 region of PRE are identical in the presence and absence of CII protein. Nucleic Acids Res 32, 1083-1090.

Minetoki, T., Tsuboi, H., Koda, A., Ozeki, K., 2003. Development of high expression system with the improved promoter using the cis-acting element in Aspergillus species. J Biol Macromol 3, 89-96.

Mudge, S.R., Smith, F.W., Richardson, A.E., 2003. Root-specific and phosphate-regulated expression of phytase under the control of a phosphate transporter promoter enables Arabidopsis to grow on phytate as a sole P source. J Plant Sci 165, 871-878.

Mullaney, E.J., Locovare, H., Sethumadhavan, K., Boone, S., Lei, X.G., Ullah, A.H., 2010. Site-directed mutagenesis of disulfide bridges in Aspergillus niger NRRL 3135 phytase (PhyA), their expression in Pichia pastoris and catalytic characterization. Appl Microbiol Biotechnol 87, 1367-1372.

Ohler, U., Niemann, H., 2001. Identification and analysis of eukaryotic promoters: recent computational approaches. Trends Genet 17, 56-60.

Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C.V., Hau, J., et al., 2007. MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res 35, W433-W437.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8, 785-786.

Pontius, J., Richelle, J., Wodak, S.J., 1996. Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol 264, 121-136.

Promdonkoy, P., Tang, K., Sornlake, W., Harnpicharnchai, P., Kobayashi, RS., Ruanglek, V., Upathanpreecha, T., Vesaratchavest, M., Eurwilaichitr, L., Tanapongpipat, S., 2008. Expression and characterization of Aspergillus thermostable phytases in Pichia pastoris. FEMS Microbiol Lett 290, 18-24.

Sarda, S., Das, A., Vinson, C., Hannenhalli, S., 2017. Distal CpG islands can serve as alternative promoters to transcribe genes with silenced proximal promoters. Genome Res 27, 553-566.

Setny, P., Bahadur, R.P., Zacharias, M., 2012. Protein-DNA docking with a coarse-grained force field. BMC bioinformatics 13, 228.

Shahmuradov, I.A., Solovyev, V.V., 2015. Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. Bioinformatics 31, 3544-5.

Shaikh, S.A., Jayaram, B., 2007. A Swift All-Atom Energy-Based Computational Protocol to Predict DNA− Ligand Binding Affinity and Δ T m. J Med Chem 50, 2240-2244.

Sharif, J., Endo, T.A., Toyoda, T., Koseki, H., 2010. Divergence of CpG island promoters: a consequence or cause of evolution? Dev Growth Differ 52, 545-554.

Teng, S., Srivastava, A.K., Wang, L., 2010. Sequence feature-based prediction of protein stability changes upon amino acid substitutions. BMC Genet 11, S5.

Tran, T.T., Mamo, G., Mattiasson, B., Hatti-Kaul, R. A., 2010. Thermostable phytase from Bacillus sp. MD2: cloning, expression and high-level production in Escherichia coli. J Ind Microbiol Biotechnol 37, 279-87.

Wasserman, W.W., Sandelin, A., 2004. Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5, 276-287.

Yao, M.Z., Zhang, Y.H., Lu, W.L., Hu, M.Q., Wang, W., Liang, A.H., 2012. Phytases: crystal structures, protein engineering and potential biotechnological applications. J Appl Microbiol 112, 1-14.