BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(4)

Functional analysis and structure determination of alkaline protease from *Aspergillus flavus*

Rabbani Syed[1]*, Roja Rani[2], Sabeena[4], Tariq Ahmad Masoodi[1], Gowher Shafi[3], Khalid Alharbi[1]

[1]College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia; [2]Biotechnology Department, Acharya Nagarjuna University,Guntur,AP, India; [3]Institute of Genetics and Hospital for Genetic Diseases, Hyderabad, India; [4]Jawaharlal Nehru Institute of Advanced Studies, Hyderabad, India; Rabbani Syed - Email: rabbanisyd@gmail.com; *Corresponding author

Received February 06, 2012; Accepted February 7, 2012; Published February 28, 2012

Abstract:
Proteases are one of the highest value commercial enzymes as they have broad applications in food, pharmaceutical, detergent, and dairy industries and serve as vital tools in determination of structure of proteins and polypeptides. Multiple application of these enzymes stimulated interest to discover them with novel properties and considerable advancement of basic research into these enzymes. A broad understanding of the active site of the enzyme and of the mechanism of its inactivation is essential for delineating its structure-function relationship. Primary structure analysis of alkaline protease showed 42% of its content to be alpha helix making it stable for three dimensional structure modeling. Homology model of alkaline protease has been constructed using the X-ray structure (3F7O) as a template and swiss model as the workspace. The model was validated by ProSA, SAVES, PROCHECK, PROSAII and RMSD. The results showed the final refined model is reliable. It has 53% amino acid sequence identity with the template, 0.24 Å as RMSD and has -7.53 as Z-score, the Ramachandran plot analysis showed that conformations for 83.4 % of amino acid residues are within the most favored regions and only 0.4% in the disallowed regions.

Keywords: Alkaline Proteases, Homology Model, Ramachandran plot

Background:
Proteases are proteolytic enzymes that catalyze the breakdown of proteins by hydrolysis of peptide bonds. Proteolytic enzymes are ubiquitous in existence, being found in all living organisms, and are crucial for cell growth and differentiation. Proteases represent one of the three largest groups of industrially essential enzymes **[1]**. Bacteria are the most dominant group of alkaline protease, Bacillus being the most relatively prominent and serve as an ideal source of these enzymes of biotechnological importance **[2, 3]** because of their rapid growth and small space required for their cultivation **[4].** Among different types Acidic, neutral and alkaline proteases, alkaline protease plays very significant role as most commonly used industrial enzyme in view of their activity and stability at alkaline pH. Proteases are one of the highest value commercial enzymes as they have broad applications in food, pharmaceutical, detergent, and dairy industries and serve as vital tools in determination of structure of proteins and polypeptides **[5, 6]**. The biotechnological promise of proteases makes them an ideal candidate for structure function relationship studies. Alkaline proteases hold a great potential for application in leather and detergent industries due to the increasing awareness of developing environmentally friendly technologies. Fungal alkaline proteases offer a distinct advantage over the currently used bacterial proteases in terms of the ease of preparation of microbe free enzyme as against the cost-effective filtration technology required for the isolation of a bacterial enzyme **[7]**. Multiple application of these enzymes stimulated interest to discover them with novel properties and considerable advancement of basic research into these enzymes. A broad understanding of the active site of the enzyme and of the mechanism of its inactivation is essential for delineating its structure-function relationship.

Studies of DNA and protein sequence homology are essential for a variety of purposes and have therefore become routine in computational molecular biology. The nucleotide and amino acid sequences of a number of proteases have been determined and their comparison is useful for elucidating the structure-function relationship **[8]**. The homology of proteases with respect to the nature of the catalytic site has been studied previously **[9, 10]**. Accordingly, the enzymes have been allocated to evolutionary families and clans. It has been suggested that there may be as many as 60 evolutionary lines of peptidases with distinct origin. The availability of genome sequences from several species of Aspergillus has facilitated the identification of a huge number of putative secondary metabolism genes and gene clusters which were previously unknown. While some limited prediction of function has been possible through bioinformatics, functional analysis has been necessary to confirm such predictions. Some of the putative secondary metabolism genes are not expressed at a sufficient level to detect products. This difficulty can sometimes be overcome by manipulating regulatory and structural genes to obtain expression, or by studying different strains of the same species, since expression can be strain dependent. Comparison of the genomes of Aspergillus species has revealed a surprising degree of secondary metabolic diversity within the genus, and provided some insights into how new clusters might evolve. The present study is an attempt to determine the functional and structural aspects of alkaline protease from Aspergillus flavus adopting *in-silico* analysis approach.

Methodology:

Primary and Secondary Structure Analysis

Percentages of hydrophobic and hydrophilic residues were calculated from the primary structure analysis. The physicochemical parameters were computed using Expasy's ProtParam prediction server. The tool SOPMA, in Expasy was used for the secondary structure prediction, secondary structure class identification and for the computation of percentages of α-helical, β- strand and coiled regions.

Functional site prediction in the Alkaline Protease

ScanProsite tool in Expasy was used to determine the functional regions present in alkaline protease.

Sequence Alignment

The FASTA sequence of alkaline protease was retrieved from the UNIPROTKB with ID Q71RZ0 that has 403 amino acids. Comparative modeling usually starts by searching the PDB of known protein structures using the target sequence as the query **[16].** This search is generally done by comparing the target sequence with the sequence of each of the structures in the database. The target sequence was searched for similar sequence using the BLAST (Basic Local Alignment Search Tool) **[17]** against Protein Data bank. The BLAST results yielded X-ray structure of 3F7O chain a (template) with an Eval of 2e-84 and 53% similarity to the target protein **Table 3 (see supplementary material).** Swiss- PdbViewer was used to produce a structure-based alignment and SWISS-MODEL was used in the optimized mode to minimize energy.

Comparative Modeling

The theoretical structure of alkaline protease from 3F7O is generated using Swiss model workspace. SWISS-MODEL repository is a database of annotated 3D protein structure models generated by the SWISS-MODEL homology-modelling pipeline. The resultant output model files consists of one or more 3D models accompanied by detailed information about the target protein and the model building process, functional annotation, a detailed template selection log, target-template alignment, summary of the model building and model quality assessment.

Validation of Model

ProSA

ProSA program explores the advantages of interactive web-based applications for the display of scores and energy plots that highlight potential problems spotted in the modeled protein structure. ProSA is a tool widely used to check 3D models of protein structures for potential errors. In particular, the quality scores of a protein are displayed. Its range of application includes error recognition in experimentally determined structures.

Saves:

SAVES tool is used for Errat value prediction, Verify 3d plot and Ramachandran plot determination. ERRAT is a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement. The program works by analyzing the statistics of non-bonded interactions between different atom types. Verify 3d plot determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc) and comparing the results to good structures.

Procheck

A versatile protein structure analysis program **[18]** available at the Joint Centre for Structural Genomics, Bioinformatics core, University of California, San Diego; was used in validation of protein structure and models by verifying the parameters like Ramachandran plot quality, peptide bond planarity, bad no bonded interactions, main chain hydrogen bond energy, alpha chirality and over-all G factor and the side chain parameters like standard deviations of chi gauche minus, trans and plus, pooled standard deviations of chi1 with respect to refined structures **[19]**

ProsaII

This program compares Z scores between target and template structure. The Z scores of model is a measure of compatibility between its sequence and structure. The model Z score should be comparable to the Z scores obtained from the template **[20]**

RMSD

Root Mean Squared Deviation (RMSD) is commonly used to represent the distance between two objects. In a structural sense, this value indicates the degree to which two three dimensional structures are similar. The lower the value, the more similar the structures are. The RMSD value **[21]** between the template and the model structure was calculated using SPDBV program.

Discussion:

Primary and Secondary Structure Analysis

Number of amino acids, molecular weight, and total number of negatively and positively charged residues, grand average of hydropathicity (GRAVY), theoretical pI and the individual composition of each amino acid are shown in **Table 1(see supplementary material)**. The results show that the alklaine protease seems stable **[11].** The target protein is good for 3D modeling as the alpha helix content is 42 % which will make the protein stable **(Figure 1 & 2).**

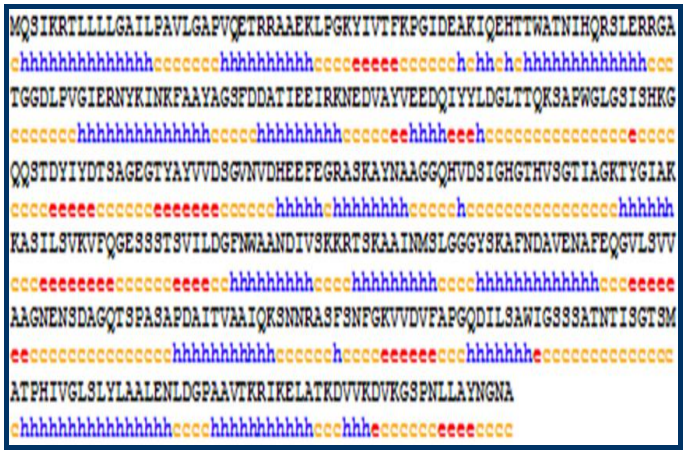

Figure 1: Secondary Structure Analysis of Alkaline Protease**.**

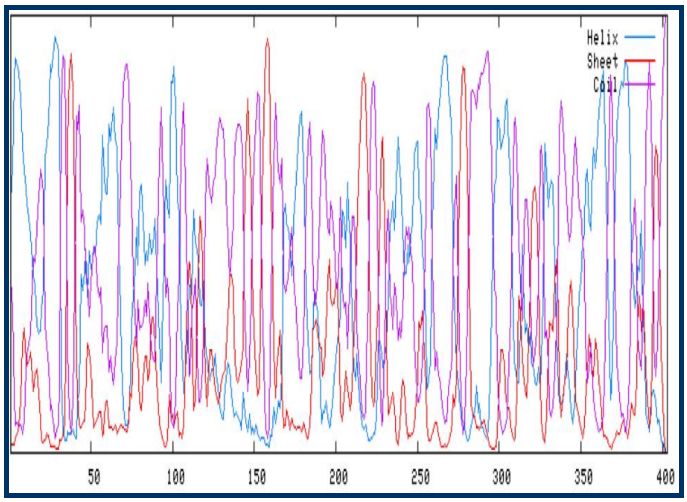

Figure 2: Distribution of secondary structure elements in Alkaline Protease.

Functional site evaluation

Three functional domains were present in the alkaline protease as a serine protease belonging to subtilase family with the given functional residues. These proteins belong to family S8 in the classification of peptidases. Subtilases are an extensive family of serine proteases whose catalytic activity is provided by a charge relay system similar to that of the trypsin family of serine proteases that evolved by independent convergent evolution. The sequence around the residues involved in the catalytic triad (aspartic acid, serine and histidine) are completely different from that of the analogous residues in the trypsin serine proteases and can be used as signatures specific to that category of proteases **Table 2 (see supplementary material).**

***Modeling of Alkaline Protease*:**

Tertiary structure of a protein is built by packing of its secondary structure elements to form discrete domains. Comparative modeling predicts the 3-D structure of alkaline protease model as a given protein sequence (target) based primarily on its alignment to template. The hypothetical protein model created is stored as PDB output file. The 3D structures of the template and the model are given in **(Figure -3).**

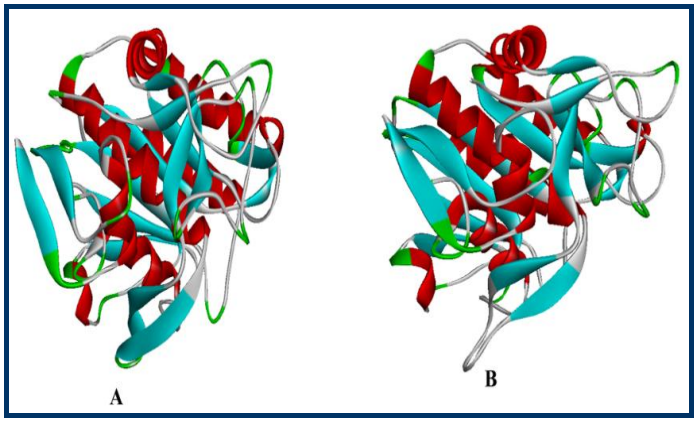

Figure 3: (A) 3D structure of template (3F70) obtained from PDB; **(B)** 3D structure of alkaline protease predicted using 3F70 as the template**.**

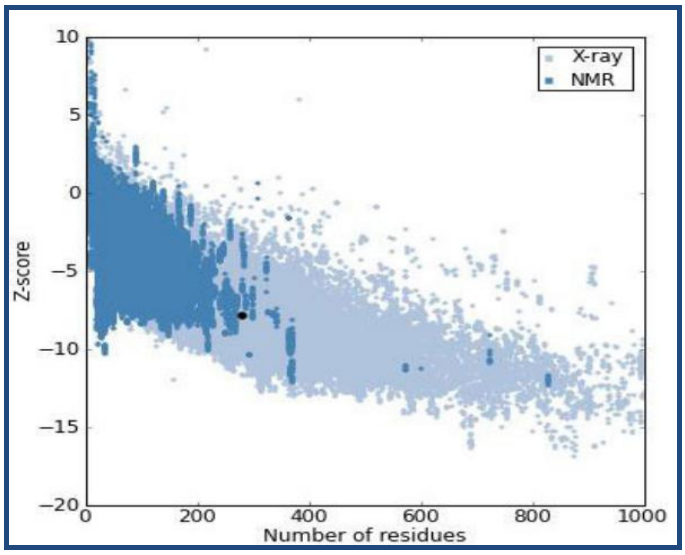

Figure 4: PROSA result showing Z-score

Evaluation and Validation of Model

The results of ProSA program used for the display of scores and energy plots that highlight potential problems spotted in the modeled protein structure are shown in **(Figure 4).** The hypothetical protein model generated was analyzed online by submitting to Joint Center for Structural Genomics (JCSG), Bioinformatics core, University of California, San Diego. Accuracy of the protein model generated was judged by validity report generated by PROCHECK. Parameter comparisons of these proteins were made with well-refined structures that have similar resolution. The main chain parameters plotted are Ramachandran plot quality, peptide bond planarity, Bad non-bonded interactions, main chain hydrogen bond energy, C- alpha chirality and over-all G factor. In the Ramamchandran plot analysis, the residues were classified according to their regions in the quadrangle. The Ramachandran map for alkaline protease is represented in **(Figure 5).**

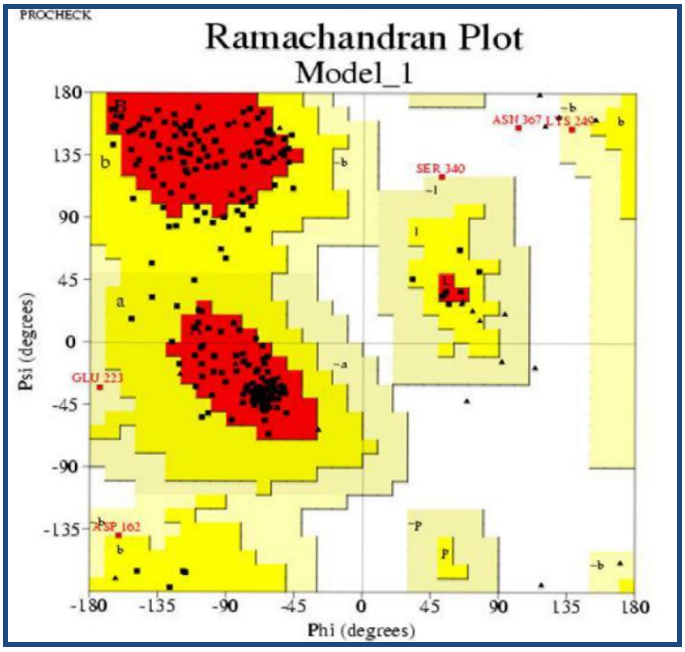

Figure 5: Ramachandran Plot

Homology protein modelling uses experimentally determined protein structures (templates) to predict the 3-D structure of another protein that has a similar amino acid sequence (the target). This approach to modelling is possible since a small change in the protein sequence usually results in a small change in its 3D structure **[12].** Homology modelling remains the only modelling method that can provide models with a root mean square error lower than 2Å. The FASTA sequence alkaline protease was obtained from Uniprot. The primary requirement for reliable homology modeling is a detectable similarity between the sequence of interest (target sequence) and a known structure (template). Due to high sequence identity between target and the template, it is good enough to use crystallographic structure of 3F7O as a template in order to obtain high quality alignment for structure prediction by homology modeling. The alignment between target and 3F7O is shown in **Table 3 (see supplementary material)**. Homology modeling is currently restricted to protein sequences (targets) that share 30% or more sequence identity to an experimentally solved protein structure template **[13].** Under this sequence identity, the reliability of the sequence alignment between target and template declines fast, resulting into significant modeling errors, low accuracy models should still be treated with attention. Medium accuracy models, obtained with a template-target sequence identity of 30–50%, tend to have nearly 85% of their C-α atom within 3.5 Å of the correct position. These models often fit a variety of applications, including the testing of ligand binding states by designing site directed mutants with altered binding capacity, and computational screening of databases listing small molecules for potential lead compounds or inhibitors. Top accuracy models, based on sequence identities more than 50%, usually have structures comparable to 3 Å resolution X-ray structures and can be used for more reliable calculations as (ligand docking, drug design), however sequence identities more than 90% can be used to facilitate a meaningful biophysical description of the active site **[14].**

The model was also tested for φ and ψ torsion angles using the Ramchandran plot. A comparison of the results show that one of the models generated is more acceptable. The molecular visualization Pymol program was used to manipulate the models based on residue interactions, energy minimization and steric hinderance. The best model predicted was used for further analysis by PROCHECK (13). Ramachandran plot analysis showed that main-chain conformations for 83.4 % of amino acid residues are within the most favored or allowed region, 14.5 % in the allowed and 1.7% in the generously allowed region and only 0.4% in the disallowed region. In general, a score close to 100% implies good stereochemical quality of the model **[15].**

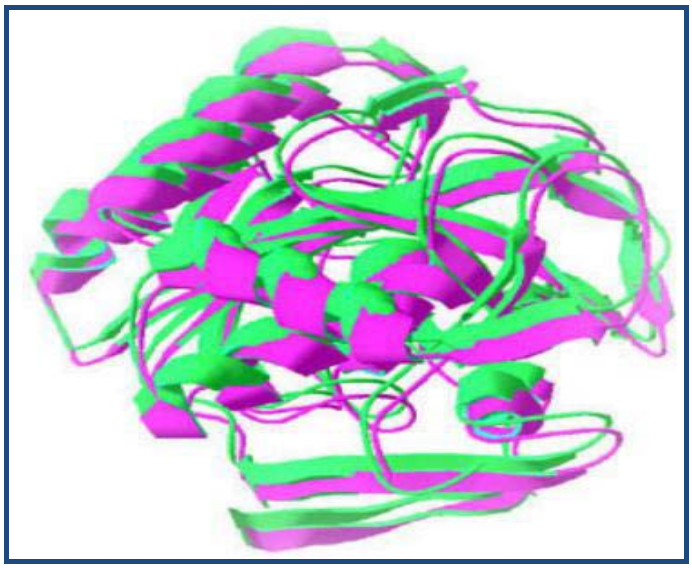

Figure 6: Superposition of template (green) with the predicted model of alkaline protease (pink). The picture has been taken from Swiss PDB viewer

The overall general similarities and subtle difference among the 3D structure of template 3F7O and predicted model can be seen from the backbone superposition. As evident from superposition **(Figure 6),** general folding topology of the structure is similar; however, some structural differences appear between the predicted model and template. These differences are mostly due to insertion and deletions in different loop regions. The RMSD (Root Mean Square Deviation) between predicted model and template is 0.24 Å. The low RMSD between the target and template echoes the presence of strong homology (The lower the value, the more similar the structures are). The z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations **[15].** In order to facilitate interpretation of the z-score of the specified protein, its particular value is displayed in a plot that contains the z-scores of all experimentally determined protein chains in current PDB. Groups of structures from different sources (X-ray, NMR) are distinguished by different colors (NMR with dark blue and X-ray with light blue). This plot can be used to check whether the z-score of the protein in question is within the range of scores usually found for proteins of similar size belonging to one of these groups. It can be seen in **(Figure 4)** that Z-scores value of the obtained model (-7.53) is located within the space of

proteins determined by X ray. This value is extremely close to the value of the template (-7.88) which recommends that the obtained model is reliable and very close to experimentally determined structure.

Conclusion:

In summary, based on the template structure it is clearly observed that the theoretical structure generated is structurally similar to the template structure which is highly sufficient for the development of specific ligand for alkaline protease. Our model of NIR is only a predictive, and needs to be confirmed experimentally.

References:

[1] Chu WH. *J Ind Microbiol Biotechnol*. 2007 **34**: 241 [PMID: 17171551]
[2] Gupta R *et al. Appl Microbiol Biotechnol*. 2002 **59**: 15 [PMID: 12073127]
[3] Ramakrishna DPN *et al. Intl J Biotechnol Biochem*. 2010 **6**: 493
[4] Arulmani M *et al. World J Microbiol Biotechnol*. 2007 **23**: 475
[5] Saeki K *et al. J Biosci Bioeng*. 2007 **103**: 501 [PMID: 17630120]
[6] Cardenas J *et al. J Mol Catal*. 2001 **14**: 111
[7] Mala B *et al. Mol. Biol. Rev*. 1998 **62**: 3597 [PMC98927]
[8] Argos P. *J Mol Biol*. 1987 **193**: 385 [PMID: 3037088]
[9] Barrett AJ. *Methods Enzymol*. 1994 **244**: 1 [PMID: 7845199]
[10] Rawlings ND & Barrett AJ, *Methods Enzymol*. 1995 **248:** 183 [PMID: 7674922]
[11] Guruprasad K *et al. Protein Eng*. 1990 **4**: 155 [PMID: 2075190]
[12] Hubbard TJ & Blundell TL, *Protein Eng*. 1987 **1**: 159 [PMID: 3507702]
[13] Baker D & Sali A, *Science*. 2001 **294**: 93 [PMID: 11588250]
[14] Marsden RL & Orengo CA, *Methods Mol Biol*. 2008 **426**: 3 [PMID: 18542854]
[15] Reddy ChS *et al. Comput Biol Chem*. 2006 **30**: 120 [PMID: 16540373]
[16] Westbrook J *et al. Nucleic Acids Res*. 2002 **1**:245 [PMID: 11752306]
[17] Altschul SF *et al. Nucleic Acids Res*. 1997 **1**: 3389 [PMID: 9254694]
[18] Laskowski RA *et al. J Biomol NMR*. 1996 **8**: 477 [PMID: 9008363]
[19] Morris AL *et al. Proteins*. 1992 **12**: 345 [PMID: 1579569]
[20] Sippl MJ. *Curr Opin Struct* Biol. 1995 **5**: 229 [PMID: 7648326]
[21] Zhang Y & Skolnick J, *Proteins*. 2004 **57**: 702 [PMID: 15476259]

Edited by P Kangueane
Citation: Syed *et al.* Bioinformation 8(4): 175-180 (2012)

BIOINFORMATION

open access

Supplementry material:

Table 1: Primary structural analysis of Alkaline Protease

AA	%tage	AA	AA	AA	%tage	AA	%tage
Ala	12.4	Arg	3.0	Asp	5.5	Lys	6.5
Phe	2.7	Asn	5.2	Cys	0.0	Met	0.7
Pro	3.0	Gly	9.9	Gln	3.5	Trp	1.0
Ser	9.2	His	2.0	Glu	5.2	Tyr	3.7
Thr	6.5	Ile	6.7	Leu	6.2	Val	7.2

Total number of negatively charged residues (Asp + Glu) = 43; Total number of positively charged residues (Arg + Lys) = 38; Grand average of hydropathicity (GRAVY) = -0.236; Number of amino acids = 403; Molecular weight = 42571.4; Theoretical pI: 5.95

Table 2: Identified domain/functional regions in Alkaline Protease

Amino acid position	Domain residues	Domain Name	Active site
158 – 169	AYVVDSGVnvdH	Serine proteases, subtilase family	Aspartic acid
193 – 203	HGThVSGtIAG	Serine proteases, subtilase family	Histidine
347 – 357	GTSmAtPhIVG	Serine proteases, subtilase family	Serine

Table 3: Sequence alignment between the template and the target

Q71RZ0	124	TTQKSAPWGLGSISHKGQQSTDYIYDTSAGEGTYAYVVDSGVNVDHEEFEGRASKAYNAAGGQHVDSIGHGTHVSGTIAG	203
3F7O_A	3	TQQPGAPWGLGRISHRSKGSTTYEYDTSGGSGTCAYVIDTGVEASHPEFEGRASQIKSFISGQNTDGNGHGTHCAGTIGS	82
Q71RZ0	204	KTYGIAKKASILSVKVFQGE-SSSTSVILDGFNWAANDIVSKKRTSKAAINMSLGGGYSKAFNDAVENAFEQGVLSVVAA	282
3F7O A	83	KTYGVAKKTKIYGVKVLDNSgSGSYSGIISGMDFAVQDSKSRSCPKGVVANMSLGGGKAQSVNDGAAAMIRAGVFLAVAA	162
Q71RZ0	283	GNENSDAGQTSPASAPDAITVAAIQKSNNRASFSNFGKVVDVFAPGQDILSAWIGSSsaTNTISGTSMATPHIVGLSLYL	362
3F7O_A	163	GNDNANAANYSPASEPTVCTVGATTSSDARSSFSNYGNLVDIFAPGSNILSTWIGGT--TNTISGTSMATPHIVGLGAYL	240
Q71RZ0	363	AALENLDGPAAVTKRIKELATKDVVKDV-KGSPNLLAYNGN	402
3F7O A	241	AGLEGFPGAQALCKRIQTLSTKNVLTGIpSGTVNYLAFNGN	281

E-value: 1.33e-84, bit-score: 259, aligned-length: 279, Identity to query: 53%