

***In-silico* analysis of caspase-3 and -7 proteases from blood-parasitic *Schistosoma* species (Trematoda) and their human host**

Shakti Kumar^{1,2}, Devendra Kumar Biswal¹ & Veena Tandon^{1,2*}

¹Bioinformatics Centre, North-Eastern Hill University, Shillong 793022, Meghalaya, India; ²Department of Zoology, North-Eastern Hill University, Shillong 793022, Meghalaya, India; Veena Tandon - Email: tandonveena@gmail.com; *Corresponding author

Received August 14, 2012; Revised October 21, 2012; Accepted April 17, 2013; Published May 25, 2013

Abstract:

Proteolytic enzymes of the caspase family, which reside as latent precursors in most nucleated metazoan cells, are core effectors of apoptosis. Of them, the executioner caspases- 3 and -7 exist within the cytosol as inactive dimers and are activated by a process called dimerization. Caspase inhibition is looked upon as a promising approach for treating multiple diseases. Though caspases have been extensively studied in the human system, their role in eukaryotic pathogens and parasites of human hosts has not drawn enough attention. In protein sequence analysis, caspases of blood flukes (*Schistosoma* spp) were revealed to have a low sequence identity with their counterparts in human and other mammalian hosts, which encouraged us to analyse interacting domains that participate in dimerization of caspases in the parasite and to reveal differences, if any, between the host-parasite systems. Significant differences in the molecular surface arrangement of the dimer interfaces reveal that in schistosomal caspases only eight out of forty dimer conformations are similar to human caspase structures. Thus, the parasite-specific dimer conformations (that are different from caspases of the host) may emerge as potential drug targets of therapeutic value against schistosomal infections. Three important factors namely, the size of amino acids, secondary structures and geometrical arrangement of interacting domains influence the pattern of caspase dimer formation, which, in turn, is manifested in varied structural conformations of caspases in the parasite and its human hosts.

Keywords: Dimerization, Caspase-3, Caspase-7, *Schistosoma*.

Background:

Schistosomiasis is a parasitic disease that is caused by blood flukes (trematodes) and leads to chronic ill health. According to the World Health Organization (WHO), it affects at least 240 million people worldwide and more than 700 million people are at risk in endemic areas. The infection is prevalent in tropical and sub-tropical areas among poor communities having no access to potable water and adequate sanitation. Transmission of infection occurs through snails (infesting water bodies), which release the infective free swimming larval stage – the cercaria; the latter enters the mammalian host directly penetrating via a cutaneous route. The causative organisms of

this disease are species of the fluke *Schistosoma*- *S. mansoni*, *S. japonicum* and *S. haematobium*, of which *S. japonicum* is the most prevalent one [1]. Presently, 'Multi-Drug-Administration' (MDA), praziquantel in particular, is used to control schistosomiasis. However, low therapeutic efficacies and side-effects limit the use of these existing drugs. Besides, no drug is effective against both the juvenile and adult stages of the fluke [2]. Immunological studies of this infection do not reveal any clear immune effector mechanism, based on which an effective vaccine could be designed. Hence, the control of schistosomiasis warrants searching for new drug targets that

can be effectively used to combat all developmental stages of the parasite's life cycle in human or mammalian hosts [3].

Induction of apoptosis is one of the most promising strategies for drug development against many diseases. Therefore, it has been considered attractive to target apoptosis for designing novel therapeutics against cancer and other autoimmune diseases [4, 5]. Moreover, activation of apoptosis can also serve as a novel possibility of drug discovery addressing parasite infections as well. Activation of apoptosis highly depends on dimerization of caspases, a family of cysteine proteases, which are of three types: apoptosis initiators, apoptosis executioners and inflammatory mediators. The apoptosis executioner caspases (Caspase-3, -6, -7 and -14) remain in dimeric form even in the inactive state, whereas the apoptosis initiator caspases (Caspase-2, -8, -9 and -10) and inflammatory mediators (Caspase-1, -4, -5, -9, -11, -12, and -13) form dimers in specific conditions or remain as monomers in the cytosol [6]. The activation mechanism of caspases highly depends on the proteolytic cleavage and adoption of loop conformation in the dimer state; each dimer has four surface loops (L1, L2, L3, and L4) that create a groove for directly binding to their substrates. During the process of dimerization an additional loop, called L2', of one monomer comes into proximity with L2 loop of the adjacent monomer [7].

The structural arrangement of a dimer single unit resembles a sandwich model with six-stranded β -sheets surrounded by five α -helices (two on one side and three on the opposite). In a dimer complex, β -sheets of dimer interfaces of both monomers are arranged by an approximately 180° rotation in such a way that the dimer comprises twelve β -sheets contained in a core surrounded by ten α -helices on nearly opposite strands of the protein [8]. In protein-protein interactions, domains interact physically with one another to adopt a proper conformation, and thus are considered functional and structural units of a protein, which perform a specific function and are evolutionarily conserved at the sequence as well as the structure level [9]. Therefore, domains sharing a functional similarity should normally be conserved. However, if there is a considerable sequence diversity of a particular domain, then its geometrical arrangement can also be different, which, in turn, may exhibit alterations in function and stability of the protein structure [10]. Phylogenetic studies of caspases from worms to humans reveal that caspases of worms, considered to be of a primitive order, are highly diverse and different from those of their mammalian hosts [11]. This diversity may affect their dimer conformation at binding interfaces to adopt different dimer conformation and their stability. Thus, Caspase-3 and -7 of blood flukes may emerge as potent drug targets [12]. In the present study, we analysed the interacting domains of Caspase-3 and -7 of *Schistosoma* spp at the sequence and structural levels and performed the protein-protein docking for studying their geometrical arrangement in the dimer state.

Methodology:

Generation of homology model of Caspase-3 and -7 sequences of *Schistosoma* species

Two Caspase-3 sequences (EMBL:CAX76583.1 and ref:XP_002574296.1) from *S. mansoni* and *S. japonicum*, respectively and two Caspase-7 sequences (ref:XP_002574645.1 and ref: XP_002580143.1) from *S. mansoni* were retrieved from

GenBank [13], National Center for Biotechnology Information (NCBI). A total of only 62 and 22 3-D structures for Caspase-3 and Caspase-7, respectively were found from humans in Protein Data Bank (<http://www.pdb.org>). Basic Local Alignment Search Tool (BLAST)p [14] was used to retrieve template hits for homology modelling and only the human caspase crystal structures could be mined for the initial 3-D structure modelling. Sequence similarities and identities of templates with corresponding schistosomal caspases are outlined in **Table 1 (see supplementary material)**. Pairwise alignment of corresponding templates and targets (supplementary material) showed some gaps, which adopted loop structures during homology modelling. Initial 3-D models were generated using automated homology modelling software Modeler9v8 [15] that generate a 3-D model by satisfaction of spatial restraints. The models were further refined by loop modelling, side chain modelling and energy minimization with the aid of Modeler9v8, SCWRL-4.0 [16] and Swiss-PDBViewer4.0 Gromacs96 program [17], respectively for getting more optimized and energetically favoured structures. Structure Analysis and Verification Server (SAVES) [<http://nihserver.mbi.ucla.edu/SAVES/>] and pKNOT v.2 web server [<http://pknot.life.nctu.edu.tw/>] were used for evaluation of the stereochemical qualities and protein knot in the modeled structures [18].

Protein-protein docking simulation

The protein-protein docking simulation was performed twice by RosettaDock Server to get the digested and undigested L2 loop dimer structures [19]. The server performs a local docking search and also assumes that protein backbone conformations typically do not change much upon association. Hence, it was necessary that both protein partners be kept in the same orientation as in their templates. The local perturbation included $\sim \pm 3$ Å distance between the two proteins, ~ 8 Å sliding of proteins relative to each other along their surfaces, $\sim 8^\circ$ tilt and a complete 360° spin around the central axis between the two proteins. The server initiated 1000 independent simulations from this range of random positions.

Calculation of the involved amino acids in dimer interface

The Protein Interaction Calculator (PIC) server was used for tracing the interacting amino acids in the dimer complex [20] and the hydrophobic, hydrophilic and ionic interactions of participating atoms were calculated in the main as well as side chains either within the intra-chain or inter-chain of protein structures utilizing solvent accessible surfaces based on distance methods with default parameters.

Structural alignment and angles, distances and vector orientation calculations between regular secondary structures of the dimer interface

For the structure-based sequence alignment of modeled caspases of *Schistosoma* spp with those of human, MUSTANG programs [21] were used for the purpose of calculating the root mean square deviation (RMSD) amongst modeled structures. Orientation vectors of modeled α -helices and β -sheets as well as crystal structures were calculated by an open source perl script developed by Thomas Holder (www.pymolwiki.org). Additional perl scripts were developed to calculate the angle and distance between the two vectors.

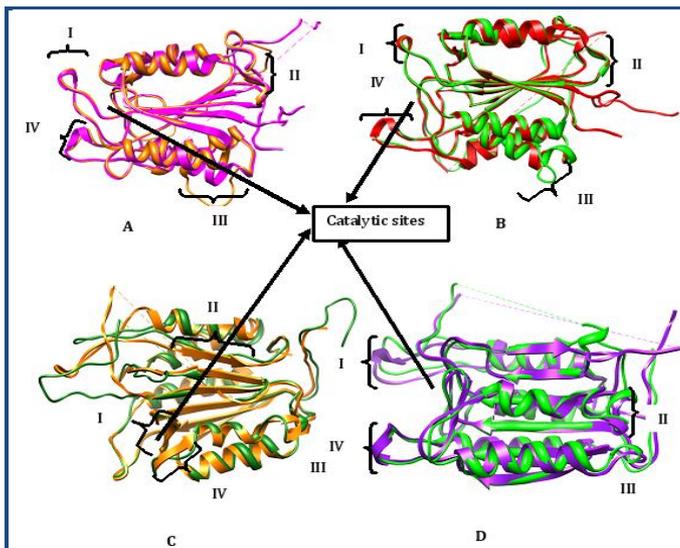


Figure 1: Structural alignment of 3-D modelled caspases -Caspase-3 and -7 of *Schistosoma* species with their respective templates. A and B is 3-D modeled structures of CAX76583.1 (orange) and XP_002574296.1 (green), respectively and their alignment with respective templates: chain A of 1NMQ (pink) and 2J31 (red). C and D are 3-D modeled structures of XP_002574645.1 (forest green) and XP_002580143.1 (green), respectively and their alignment with respective templates: chain A of 1K86 (orange) and 2J31 (purple). There are four regions where structural differences were found in schistosomal modeled caspase structures; I and IV regions of all modeled structures belong to catalytic sites; dotted lines in all alignment structures indicate those regions, which have been cleaved in templates, whereas in modeled structures the corresponding regions have been manually deleted.

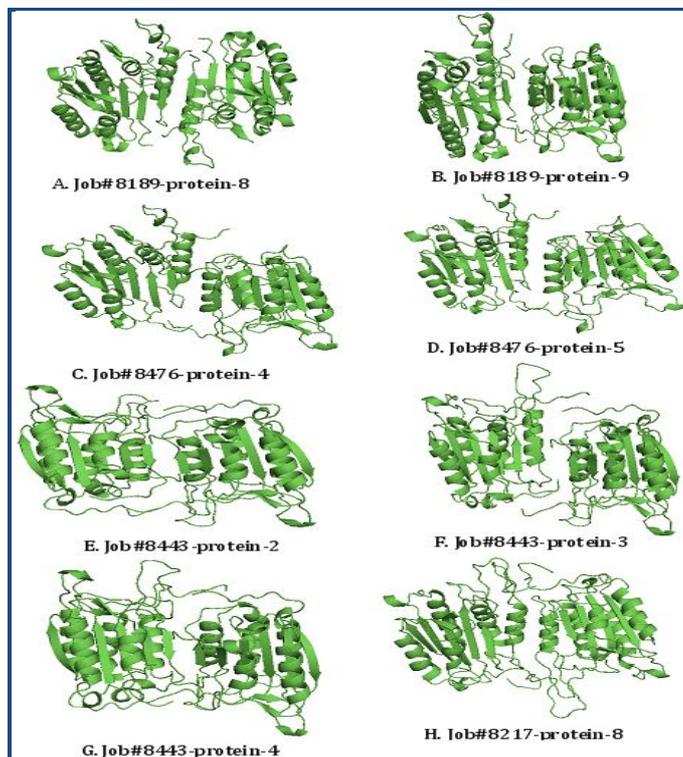


Figure 2: Dimer conformations (eight in all) of *Schistosoma* Caspase-3, and -7 that are similar to human caspases. A-B:

CAX76583.1 (*S. japonicum*); C-D: XP_002574296.1 (*S. mansoni*); E-G: XP_002580143.1 (*S. mansoni*); H: XP_002574645.1 (*S. mansoni*).

Results & Discussion:

Parasite sequence identity and similarity

In similarity search analysis, Caspase-3 sequences of *S. japonicum* [EMBL: CAX76583.1] and *S. mansoni* [ref: XP_002574296.1] shared 70% similarity between themselves. However, the two available Caspase-7 sequences of *S. mansoni* (ref:XP_002574645.1 and ref:XP_002580143.1) surprisingly shared only 19% identity to each other despite belonging to the same species **Table 2** (see supplementary material). In human caspases there is higher percentage of aromatic, hydrophilic and positively and negatively charged amino acids as compared to those in *Schistosoma* caspases. However, with regard to hydrophobic amino acids, the reverse is true **Table 3** (see supplementary material).

Modelling of structure and validation

The residues in Caspase-3 and -7 of *Schistosoma* species were found in allowed regions of the Ramachandran diagram and an overall G-factor of 0.00, -0.01, 0.03 and -0.01 was found in the accessions - EMBL:CAX76583.1, ref:XP_002574296.1, ref:XP_002574645.1 and ref:XP_002580143.1, respectively (**Table 4** (see supplementary material)). Force field energies were calculated and are -1739.008, -7286.563, -2089.667 and 788.141 KJ/mol for the accessions EMBL:CAX76583.1, ref:XP_002574645.1, ref:XP_002574296.1 and ref:XP_00258143.1, respectively. All modelled structures are energetically favoured and an overall G-factor quality approached zero, thereby substantiating the fact that the modeled structures in the present study are suitable for protein-protein docking studies. Acceptable G-factor values in PROCHECK are between 0 and 0.5 for reliable structures, the best model being the one with zero G-factor [22].

Structure similarity and structure alignment

There are some regions in the modeled structures that have adopted different secondary structures from their respective templates (**Figure 1**). The average sequence length of different secondary structural motifs is larger in EMBL:CAX76583.1 compared to ref:XP_002574296.1 of Caspase-3, and in ref:XP_002574645.1 compared to ref:XP_002580143 of Caspase-7, respectively **Table 5** (see supplementary material). When RMSD of modeled structures of Caspase-3 and -7 in *Schistosoma* species were analysed with those of human templates, the RMSD was found to be greater than 2.0 Å **Table 6** (see supplementary material). This observation revealed that human caspases are structurally distant from schistosomal caspases. Similarly, EMBL: CAX76583.1 and ref:XP_002574645.1 are evolutionarily distant from ref: XP_002574296.1 and ref:XP_002580143.1, respectively (**Table 6**). Greater C α RMSDs (greater than 1.05 \pm 0.002 and 1.55 \pm 0.001 Å for orthologs and paralogs, respectively) in homologous structures may affect the binding affinity and stability of ligands [23]. Geometrical changes (as measured by the local RMSD) of the active sites also influence the ligand binding affinity among orthologs and therefore, exhibit a positive correlation, with secondary motif lengths in structural alignments [24].

Protein-Protein Interaction simulation

The loop-digested Caspase-3 and -7 were docked for obtaining the homodimers. RosettaDock server provided the lowest minimum energies containing ten dimer conformations for each modeled structure (i.e., EMBL:CAX76583.1, ref:XP_002574296.1, ref:XP_002574645.1 and XP_002580143.1) (see supplementary files). Out of 40 dimer conformations, 8 (viz., job#8189-protein-8 and job#8189-protein-9 dimer of EMBL:CAX76583.1; job#8476-protein-4 and job#8476-protein-5 of ref:XP_002574296.1; job#8443-protein-2, job#8443-protein-3, job#8443-protein-4 and job#8443-protein-8 of ref:XP_002580143.1 and job#8217-protein-8 of ref:XP_002574645.1) were shown to have similar conformations in proximity with human crystal structures and β -sheets of interacting surfaces of both the chains (**Figure 2**). Other conformations (namely, job#8189-protein-2, job#8189-protein-3, job#8189-protein-5, job#8189-protein-6 and job#8189-protein-7 of EMBL:CAX76583.1; job#8476-protein-0, job#8476-protein-3, job#8476-protein-4, job#8476-protein-5 and job#8476-protein-6 of ref:XP_002574296; job#8443-protein-0, job#8443-protein-1, job#8443-protein-6 and job#8443-protein-7 of ref:XP_002580143.1) of the modeled structures had different dimer structures, wherein helix-helix or helix-sheet domain interactions were dominant at interacting surfaces and these domains were closer to each other unlike interactions involving only β -sheets in humans. With regard to Caspase-7, modeled structure of ref:XP_002574645.1 (*S. mansoni*) along with all dimer conformations did not bear any similarity with those of other modeled structures of *Schistosoma* species or humans, with the exception of one job#8217-protein-8, that exhibited somewhat similar conformations to human Caspase-7. Of all the conformations of this modeled structure, five (viz., job#8217-protein-1, job#8217-protein-3, job#8217-protein-4, job#8217-protein-6 and job#8217-protein-7) showed a striking similarity amongst themselves.

Differences in secondary structure elements in structural alignment of 3-D modeled caspase structures of *Schistosoma* species with their templates were found mostly at the protease activity-performing sites (**Figure 1**). In *Schistosoma* species the catalytic sites of caspases are mostly made up of coiled secondary structures, whereas the human caspases comprise both β -sheets and coils. The solvent accessible surface areas of regular secondary structures (i.e., α -helices and β -sheets) are more in occurrence as compared to coiled secondary structures. Similarly, backbones of regular secondary structures follow a specific arrangement unlike in coiled structures. Hence, the presence of an irregular secondary structure at a catalytic site may have some effect on their binding affinity and cleavage activity [25].

In case of ref:XP_002574645.1 modeled structure, interestingly glycine is at the 107th position amidst 2nd and 3rd motifs and the 231 to 257 region reflects regular occurrence of glycine and proline at the 239th and 246th position, respectively. Previous analyses support destabilization of the β -sheet by glycine and proline, which is why the 4th motif is a coil (i.e., which actively participates in dimer formation), whereas in the template the corresponding region is a β -sheet [26].

Although 3-D structures of caspase sequences of *Schistosoma* species were modeled on the basis of human caspase crystal

structures (see 'Materials and methods'), the dimer forms that were predicted by RosettaDock server are not the same as those of human caspases, thus tempting us to analyze the interacting domains on the basis of three factors: 1) distance and angle between interacting helices and β -sheets of *Schistosoma* caspases, 2) arrangement of secondary structure elements on the dimer interface, and 3) amino acid distribution at the dimer interface. The angle between the β -sheet and α -helix, which participate in dimer formation, is lesser than that in their templates. The arrangement of secondary structure elements at the dimer interface of EMBL:CAX76583.1, ref:XP_002574296.1 and ref:XP_002580143.1 is same as in their templates; however, in ref:XP_002574645.1 in place of β -sheet in the dimer interface there is a long loop. As predicted the solvent-accessible surface area of β -sheets is more than that of loop/coil structures [27]. Hence, there are more chances for a cavity being formed at the dimer interface in XP_00257465.1 but not in other structures. An analysis of amino acid distribution at the dimer interface of parasite caspases exhibited more occurrence of small amino acids (i.e., glycine, alanine, valine, serine, threonine, proline and cysteine) in schistosomal caspases compared to 1NMQ, 2J31 and 1K86 (caspases of human). The distance between β -sheets and α -helices is almost the same in all the structures except ref | XP_002574645.1 but the amino acid distribution was not identical in human and *Schistosoma* species (**Figure 3**). Due to their high occurrence in schistosomal caspases, amino acids having small side chains are not able to fill the distance between the β -sheets and α -helices; therefore, a large cavity is formed at the dimer interface contrary to human caspases with high occurrence of amino acids with long side chains, which are able to create bridges between α -helices and β -sheets. Therefore, during protein-protein docking simulation the interacting α -helix of one partner acts as an interacting ligand in the cavity created at the dimer surfaces of its counterpart. Thus, either helix-helix or helix-sheet interactions are more prominent than β -sheet- β -sheet interactions during dimer conformations of Caspase-3 and -7 of *Schistosoma* species, which are found to be different from human caspases.

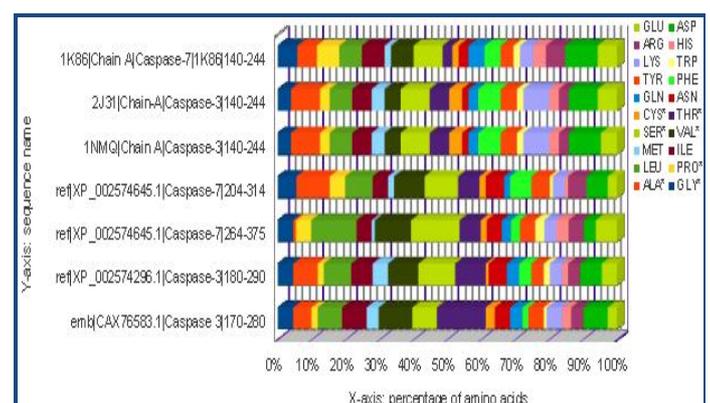


Figure 3: Amino acid distribution at the dimer interface of the modeled structure of schistosomal caspases and their templates. X-axis and y-axis represent twenty amino acids and their occurrence number in schistosomal and human caspases respectively.

Conclusion:

On the basis of sequence identity and similarity Caspase-3 and -7 of *Schistosoma* species can be considered homologs of human caspases. The schistosomal caspases are unique in comparison

to those of their host in the following salient points: 1) At the secondary structure level, schistosomal caspase active sites have different secondary structural elements (combination of β -sheets and coils) and these structural differences may affect their protease activity. In XP_002574645.1 (Caspase-7; *S. mansoni*) structure, the 244-248 region is a loop owing to regular occurrence of proline and glycine amino acids; 2) A large cavity with a rough surface formed due to the occurrence of a large number of small amino acids in schistosomal caspases at the dimer interface may be responsible for adopting different dimer conformation structures during the protein-protein docking simulation. Large cavities are in a way responsible for increasing the interfacial atomic distance, thereby reducing the dimer stability and, affecting apoptosis initialization; 3) At the tertiary level, the caspase dimer formations in *Schistosoma* species are governed by the nature of amino acids involved, secondary structure elements of the interacting domains and the angle and distance between the latter; 4) The dimer conformations of helminth caspases, being different from those of their host (human) would also have cavities with different sizes and shapes at the dimer interface, which, in turn, would alter their binding properties as well that seems promising for design of molecules more specific to helminth caspase active sites. Thus, the parasite executioner caspases may be exploited as potent drug targets of chemotherapeutic use.

Authors' Contributions:

SK performed the designing of the study, computational analysis and literature screening. VT and DKB initiated the idea and contributed to the preparation and writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgment:

This work was supported by a Department of Information Technology, Government of India - funded project 'North East Parasite Information and Analysis Centre (NEPIAC): in silico approach' to VT & DKB.

References:

- [1] Jia TW *et al. Bull World Health Organ.* 2007 **85**: 465 [PMID: 17639243]
- [2] Hotez PJ *et al. J Clin Invest.* 2008 **118**:1311 [PMID: 18382743]
- [3] Caffrey CR *et al. PLoS ONE.* 2009 **4**: 4413 [PMCID: PMC2635471]
- [4] Nicholson DW, *Nature.* 2000 **407**: 810 [PMID: 11048733]
- [5] MacKenzie SH *et al. Curr Opin Drug Discov Devel.* 2010 **13**: 568 [PMID: 20812148]
- [6] FAN TJ *et al. Acta Biochim Biophys Sin (Shanghai).* 2005 **37**: 719 [PMID: 16270150]
- [7] Witkowski WA & Hardy JA, *Protein Sci.* 2009 **18**: 1459 [PMID: 19530232]
- [8] Riedl SJ *et al. Proc Natl Acad Sci U S A.* 2001 **98**: 14790 [PMID: 11752425]
- [9] Fuentes-Prior P & Salvesen GS, *Biochem J.* 2004 **384**: 201 [PMID: 15450003]
- [10] Mika S & Rost B, *PLoS Comput Biol.* 2006 **2**: 0698 [PMID: 16854211]
- [11] Lamkanfi M *et al. Cell Death Differ.* 2002 **9**: 358 [PMID: 11965488]
- [12] Mohapatra AD *et al. PLoS Negl Trop Dis.* 2011 **5**: 1306 [PMID: 21931872]
- [13] Altschul SF *et al. J Mol Biol.* 1990 **215**: 403 [PMID: 2231712]
- [14] Benson DA *et al. Nucleic Acids Res.* 2011, **39**: D32 [PMID: 21071399]
- [15] Fiser A & Sali A, *Meth Enzymol.* 2003 **374**: 461 [PMID: 14696385]
- [16] Krivov GG *et al. Proteins.* 2009 **77**: 778 [PMID: 19603484]
- [17] Guex N & Peitsch MC, *Electrophoresis.* 1997 **18**: 2714 [PMID: 9504803]
- [18] Lai YL *et al. Nucleic Acids Res.* 2007 **35**: 420 [PMID: 17526524]
- [19] Lyskov S & Gray JJ, *Nucleic Acids Res.* 2008 **36**: 233 [PMID: 18442991]
- [20] Tina KG *et al. Nucleic Acids Res.* 2007 **35**: 473 [PMID: 17584791]
- [21] Konagurthu AS *et al. Proteins.* 2006 **64**: 559 [PMID: 16736488]
- [22] Hazai E & Bikadi Z, *J Struct Biol.* 2008 **162**: 63 [PMID: 18249138]
- [23] Peterson ME *et al. Protein Sci.* 2009 **18**: 1306 [PMID: 19472362]
- [24] Brylinski M *et al. J Struct Biol.* 2011 **173**: 558 [PMID: 20850544]
- [25] Saraboji K *et al. Comput Biol Chem.* 2005 **29**: 25 [PMID: 15680583]
- [26] Li SC *et al. Natl Acad Sci U S A.* 1996 **93**: 6676 [PMID: 8692877]
- [27] Singh SK *et al. Proteins.* 2003 **51**: 167 [PMID: 12660986]

Edited by P Kanguane

Citation: Kumar *et al.* Bioinformation 9(9): 456-463 (2013)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Best hits obtained by protein-protein BLAST result with PDB database

S.N.	Target sequence	Template (best hit) name	Seq identity (%)	Seq similarity (%)	Target coverage	E-value
1	gi 256076694 ref XP_002574645.1 Caspase-7 (C14 family) [<i>Schistosoma mansoni</i>]	pdb 1K86 Chain A	84/265 (32%)	134/265 (51%)	67%	4e-36
2	gi 256087989 ref XP_002580143.1 Caspase-7 (C14 family) [<i>Schistosoma mansoni</i>]	pdb 2J31 Chain-A	102/250 (42%)	155/250 (61%)	75%	1e-65
3	gi 256075986 ref XP_002574296.1 Caspase-3 (C14 family) [<i>Schistosoma mansoni</i>]	pdb 2J31 Chain-A	105/237 (44%)	145/237 (61%)	79%	4e-66
4	gi 226469514 emb CAX76583.1 Caspase-3, apoptosis-related cysteine peptidase [<i>Schistosoma japonicum</i>]	pdb 1NMQ Chain A	91/233 (39%)	137/233 (59%)	80%	2e-53

Table 2: Caspase-3 and -7 sequences of *Schistosoma* species and their templates: identity percentage matrix

S.N	Protein name	1	2	3	4	5	6	7
1	gi 226469506 emb CAX76583.1 Caspase-3 Sch_jap		70.0	23.0	27.0	35.0	35.0	32.0
2	gi 256075986 ref XP_002574296.1 Caspase-3 Sch_man			23.0	31.0	40.0	44.0	35.0
3	gi 256076694 ref XP_002574645.1 Caspase-7 Sch_man				19.0	28.0	28.0	30.0
4	gi 256087989 ref XP_002580143.1 Caspase-7 Sch_man					40.0	42.0	39.0
5	1NMQ_A PDBID CHAIN SEQUENCE Caspase-3						99.0	54.0
6	2J31_A PDBID CHAIN SEQUENCE Caspase-3							54.0
7	1K86_A PDBID CHAIN SEQUENCE Caspase-7							

Table 3: Protein sequences and their amino acid composition: Caspase-3 and -7 of human and *Schistosoma* species

S.N.	Protein name	Length	% of hydrophobic amino acids	% of aromatic amino acids	% of hydrophilic amino acids	% of positively charged amino acids	% of negatively charged amino acids
1	1NMQ:A chainID Caspase-3 <i>Homo sapiens</i>	249	35.341	10.843	24.096	16.064	13.655
2	2J31:A chainID Caspase-3 <i>Homo sapiens</i>	254	35.600	10.800	24.000	16.000	13.600
3	1K86:A chainID Caspase-7 <i>Homo sapiens</i>	253	37.549	10.277	22.925	15.415	13.834
5	ref XP_002574645.1 Caspase-7 [<i>Schistosoma mansoni</i>]	379	42.216	6.596	22.691	15.831	12.665
6	ref XP_002580143.1 Caspase-7 [<i>Schistosoma mansoni</i>]	329	41.033	8.815	28.267	11.854	10.030
7	emb CAX76583.1 Caspase-3, [<i>Schistosoma japonicum</i>]	288	38.194	9.722	27.778	12.847	11.458

Table 4: Stereochemical quality of modeled Caspase-3 and -7 3-D structures by Procheck Server

SN	Protein name	Most favoured region	Additional favoured region	Generously favoured region	Disallowed region	Overall quality G-factor
1.	CAX76583.1 Caspase-3	92.5%	6.6%	0.9%	0.0%	0.00
2.	XP_002574296.1 Caspase-3	90.8%	8.3%	0.9%	0.0%	-0.01
3.	XP_002574645.1 Caspase-7	93.0%	7.0%	0.0%	0.0%	0.03
4.	XP_002580143.1 Caspase-7	96.4%	3.6%	0.0%	0.0%	-0.01

Table 5: Regions of schistosomal Caspase-3 and -7, whose Secondary Structure Elements (SSEs) differ from their templates

S.N.	Target and templates	Region-1	SSE	Region-2	SSE	Region-3	SSE	Region-4	SSE
Caspase-3									
1.	EMBL:CAX76583.1 PDB: 1NMQ Chain A	12-16	Loop	63-66	Helix	179-187	Loop	213-217	Loop
		56-60	Helix	107-111	Loop	221-226	Helix	253-256	Helix
2.	Ref:XP_002574296.1 PDB: 2J31 Chain A	12-16	Loop	64-68	Helix	104-107	Loop	185-189	Helix
		57-60	Helix	107-111	Loop	149-151	Helix	225-229	Helix-Coil
Caspase-7									
3.	Ref:XP_002574645.1 PDB: 1K86 Chain A	92-99	Loop	106-111	Loop	120-130	Loop-Helix-Loop	244-248	Loop
		147-152	Sheet	154-158	Sheet	167-176	Helix	290-294	Sheet
4.	Ref:XP_002580143.1 PDB: 1K86 Chain A	74-77	Helix	93-110	Loop	112-117	Loop	118-122	Helix
		130-134	Loop	149-158	Sheet-Loop-Sheet	168-174	Helix	174-177	Helix-Loop

Table 6: Structural alignment of schistosomal Caspase-3 and -7 with their templates and RMSD matrix.

S.N.	Modeled structures of Caspase-3 and -7	1	2	3	4	5	6	7
1	EMBL: CAX76583.1 Caspase-3		0.6	2.2	0.9	2.4	2.3	2.4
2	Ref:XP_002574296.1 Caspase-3			2.1	0.9	2.1	2.5	2.2
3	Ref:XP_002574645.1 Caspase-7				2.4	2.4	2.3	1.9
4	Ref:XP_002580143.1 Caspase-7					2.3	2.4	2.1
5	PDB:1NMQ chain-A Caspase-3						0.2	1.3
6	PDB:2J31 chain-A Caspase-3							1.3
7	PDB:1K86 chain-A Caspase-7							

Sequence alignments:

```
>pdb|2J31|A Structure related to 2J31_A Chain A, The Role Of Loop Bundle
Hydrogen Bonds In The Maturation
And Activity Of (Pro) caspase-3
Length=250
sbjct = ref|XP_002574296|caspase-3
Query = pdb|2J31|A
Score = 207 bits (526), Expect = 4e-66, Method: Compositional matrix
adjust.
Identities = 105/237 (44%), Positives = 145/237 (61%), Gaps = 8/237
(3%)

Query 57 GECILINQRDFHPSTNQSRDGTDVDADRVERVFKSLNYKVTRILNITKSVLHQTLLEAS
116
Sbjct 17 G CI+IN ++FH ST + R GTDVDA + F++L Y+V ++T+ + + + + S
76 GLCIIINNKNFHKSTGMTSRSGTDVDAANLRETFRNLYEVRNKNDLTREEIVELMRDVS

Query 117 QADHSSYDSFIFVMLSHGDNNIIYANDGEVLTSYIMAFFRGDRCPSLIAKPKLFFFQACR
176
Sbjct 77 KEDHSKRSSFVCVLLSHGEEGIIFGTNGPVDLKKITNFFRGDRCSRSLTGKPKLFI IQACR
136

Query 177 GAAFDKGVSTMVTDAG--EDLIVHKLPIEADILVAYSTVPGFFAWRNSSSGSWFIQELCN
234
Sbjct 137 G A D G+ TD+G +D+ HK+P+EAD L AYST PG+++WRNS GSWFIQ LC
193 GTALDCGIE---TDSGVDDMACHKIPVEADFLYAYSTAPGYYSWRNSKDGSWFIQSLCA

Query 235 ALESDMKNANHSDIMSLT VVARVVAYQYRSNTGQIETDNMKQMTSTVSTLTRLFYI
291
Sbjct 194 L+ + A+ + M +LT V R VA ++ S + KQ+ VS LT+ Y
247 MLK---QYADKLEFMHILTRVNRKVATEFESFSFDFATFHAKKQIPCI VSM LTKELYF
```

```
>pdb|2J31|A Structure related to 2J31_A Chain A, The Role Of Loop Bundle
Hydrogen Bonds In The Maturation
And Activity Of(Pro)caspase-3
Length=250
Query = ref|XP_002580143.1|caspase-7
Sbjct = pdb|2J31|A
Score = 207 bits (526), Expect = 1e-65, Method: Compositional matrix
adjust.
Identities = 104/247 (42%), Positives = 150/247 (61%), Gaps = 8/247
(3%)

Query 71 YRTDGAERGLCVVFSVDNFDPI LCLPPRNGSLVDVQNIKRAFANLDFRVIVHSNPTSESL
130
Sbjct 9 Y+ D E GLC++ + NF + R+G+ VD N++ F NL + V ++ T E +
68 YKMDYPEMGLCII IINNKNFHKSTGMTSRSGT DVDAANLRETFRN LKYEVRNKNDLTREEI

Query 131 FSLIRTISTQNLRDHDCFACVILSHGDEGGLIYATDGSIPVDRIIAPFRGDQCLDLRGKP
190
Sbjct 69 L+R +S ++ F CV+LSHG+EG +I+ T+G + + +I FRGD+C L GKP
127 VELMRDVS KEDH SKRSSFVCVLLSHGEEG-IIFGTNGPVLDLKKITNFFRGDRCSRSLTGKP

Query 191 KLFFIQACRGMALDDGVFVCDGPLSNSKSADAATTIRRI PVEADLFIAYAVQPGYYAFRN
250
Sbjct 128 KLF IQACRG ALD G+ G D +IPVEAD AY+ PGYY++RN
180 KLFFIQACRGTALDCGIETDSG-----VDDDMACHKIPVEADFLYAYSTAPGYYSWRN

Query 251 SVNGSWFIRALSDVLLRYGNTLDLLSIMTRVNYDVAFEYESTAANPSFCGKKQMP SFVST
310
Sbjct 181 S +GSWFI++L +L +Y + L+ + I+TRVN VA E+ES + + +F KKQ+P VS
240 SKDGSWFIQSLCAMLKQYADKLEFMHILTRVNRKVATEFESFSFDATFHAKKQIPCIVSM

Query 311 LTKHVIF 317
Sbjct 241 LTK + F
LTKELYF 247
```

```
>pdb|1NMQ|A Structure related to 1NMQ_A PubChem BioAssay Info linked to
1NMQ_A Chain A, Extended Tethering: In Situ Assembly Of Inhibitors
Length=249
Query = emb|CAX76583.1|caspase 3
Sbjct = pdb|1NMQ|A
Score = 174 bits (441), Expect = 2e-53, Method: Compositional matrix
adjust.
Identities = 91/233 (39%), Positives = 137/233 (59%), Gaps = 8/233 (3%)

Query 46 GECLLINQRDFDPSVQSPRDGTDIDADSI ERVFRSLNYKVTRILNITKKALDQTLVEAS
105
Sbjct 17 G C++IN ++F S G + R GTD+DA ++ FR+L Y+V ++T++ + + + + S
76 GLCII IINNKNFHKSTGMTSRSGT DVDAANLRETFRN LKYEVRNKNDLTREEI VELMRDVS

Query 106 QKDHL SYDSFV FIMLTHGGNNIVYTSDDKVQTSYIMSF FRGDRCP SLIRKPKLFFFQVSR
165
Sbjct 77 ++DH SFV ++L+HG I++ ++ V I +FFRGDRC SL KPKLF Q R
136 KEDH SKRSSFVCVLLSHGEEGIIIFGTNGPVLDLKKITNFFRGDRCSRSLTGKPKLFFIIQACR

Query 166 GAS YDKGVSTMVADAG--DDVIVHKLPVDADILISYSTVPGYITCCNTTKG SWYIQELCN
223
Sbjct 137 G D G+ T D+G DD+ HK+PVDAD L +YST PGY + N+ GSW+IQ LC
193 GTELDCGIET---DSGVDDDMACHKIPVDADFLYAYSTAPGYYSWRNSKDGSWFIQSLCA

Query 224 ALETDIKNGNHTDIMTLLTVVARVVAYQYRSSTRQVETDNMKQMTSTFSTLTR 276
Sbjct 194 L+ + + + M +LT V R VA ++ S + KQ+ S LT+
MLK---QYADKLEFMHILTRVNRKVATEFESFSFDATFHAKKQIPCIVSM LTK 243
```