In Silico Biology 7 (2007) 187–193
IOS Press

ISN1 Nucleotidases and HAD Superfamily Protein Fold: *In silico* Sequence and Structure Analysis

Bharath Srinivasan and Hemalatha Balaram*

Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore 560 064, India

Edited by E. Wingender; received 11 November 2006; revised 16 February 2007; accepted 24 February 2007; published 28 February 2007

ABSTRACT: cN-II class of 5' purine nucleotidases exhibit specificity for IMP/GMP and belong to the HAD (haloacid dehalogenase) superfamily of hydrolases. The recently identified ISN1 class of IMP specific 5'-nucleotidases occurring in yeast, fungi and certain Plasmodia lack sequence homology with the cN-II class of enzymes. We show from analysis of motif and fold conservation that ISN1s also belong to the HAD superfamily. This identification adds a new novel member to this superfamily.

KEYWORDS: Purine nucleotidase, ISN1 nucleotidase, HAD superfamily, distant homology modeling, phylogenetic analysis

Abbreviations: ISN1, IMP specific 5'-nucleotidase 1, HAD, Haloacid dehalogenase, BLAST, Basic Local Alignment Search Tool, PSI-BLAST, Position Specific Iterative BLAST, PSSM, Position Specific Scoring Matrices

INTRODUCTION

Structural folds and catalytic motifs are highly conserved among members of enzyme families catalyzing similar reactions on related substrate molecules. One such class of enzymes is the family of soluble 5'-nucleotidases (E.C. 3.1.3.5) that catalyzes the hydrolysis of phosphate group esterified at the 5' carbon of the ribose or deoxyribose moiety of nucleoside monophosphates. 5'-nucleotidases are widely distributed among the various kingdoms of life and play a key role in modulating levels of cellular nucleotide pools. Soluble 5'-nucleotidases are located in the cytosolic and mitochondrial compartments of the cell. Cytosolic 5'-nucleotidases (cN) are grouped into 3 classes based on their substrate specificity. cN-I and cN-II class of 5'-nucleotidases with purine base specificity hydrolyze AMP and IMP/GMP [Bretonnet *et al.*, 2005] respectively, while cN-IIIs are pyrimidine nucleotidases. The mitochondrial and cytosolic 5'(3')-deoxyribonucleotidases are referred to as mdN and cdN, respectively [Hunsucker *et al.*, 2005].

All soluble forms of 5'-nucleotidases are known to belong to the haloacid dehalogenase (HAD) superfamily of hydrolases and require the cofactor magnesium for catalysis [Hunsucker et al, 2005].

*Corresponding author. Tel.: +91 80 22082812; Fax: +91 80 22082766; E-mail: hb@jncasr.ac.in.

Electronic publication can be found in *In Silico Biol.* **7**, 0019 <http://www.bioinfo.de/isb/2007/07/0019/>, 28 February 2007.

Evidence for HAD fold in soluble nucleotidases comes from the available structures of cN-II from *Legionella pneumophila* (2BDE), cN-III from *Mus musculus* (2BDU) and mdN from *Homo sapiens* (PDB id, 2BDE, 2BDU and 1Z4I, respectively). HAD superfamily constitutes a variety of enzymes that catalyze the cleavage of substrates with C-Cl, P-C, and P-OC bonds via nucleophilic substitution pathways [Allen and Dunaway-Mariano, 2004]. The HAD superfamily consists of more than 3000 members in whom the overall fold is highly conserved. All the members contain an α/β Rossman core and a cap domain. The active site located in the core domain is formed by four loops, which position substrate and cofactor binding residues as well as the catalytic groups that mediate the chemistry [Allen and Dunaway-Mariano, 2004]. Motifs implicated in catalytic function and Mg^{2+} cofactor binding that are conserved in HADs are also conserved in 5'-nucleotidases.

Saccharomyces cerevisiae lacks a homologue of conventional cN-II and the IMP specific 5'-nucleotidase (ISN) in this organism is grouped under the protein family ISN1 [Itoh *et al.*, 2003]. Homologs of yeast ISN1 are present in *Schizosaccharomyces pombe*, many fungi and some species of Plasmodia. Though ISN1s and cN-IIs are 5' purine nucleotidases with common substrate specificity for IMP, they do not share significant sequence similarity. We show in this paper that ISN1 sequences, like the cN-II class of nucleotidases, are also members of the HAD superfamily. The extreme sequence divergence of ISN1s had prevented the identification of HAD specific motifs in them and therefore, exclusion from HAD membership.

MATERIALS AND METHODS

YOR155c sequence (*S. cerevisiae* ISN1) was obtained from GeneDB (http://www.genedb.org). The sequences of annotated cytosolic nucleotidases were downloaded from the gene database of NCBI. The non-redundant database at NCBI was used to search for homologs of ISN1s using the algorithm BLASTP. Distant homology searches were carried out using PSI-BLAST [Altschul *et al.*, 1997]. T-Coffee (http://www.ebi.ac.uk/t-coffee/) [Notredame *et al.*, 2000] was used for generating multiple sequence alignment profiles. The protein sequences of the ISN1s were submitted to servers nFOLD and mGenTHREADER [Jones, 1999] for fold recognition and initial pairwise alignment of the query sequence with the ten most suitable templates. Phylogenetic and molecular evolutionary analyses were done using MEGA, version 3.1 [Kumar *et al.*, 2004].

RESULTS AND DISCUSSION

Distant homologs of ISN1 sequences

A search of the non-redundant database of NCBI using BLAST, for homologs of *S. cerevisiae* ISN1 yielded 20 retrievals. Except for one hit of ISN1 homolog from *Plasmodium falciparum* all the retrievals were of fungal origin. In the absence of any structural information for the ISN1 family of proteins, the sequences were analyzed by position specific iterative BLAST (PSI-BLAST) [Altschul *et al.*, 1997] to detect remote homologues. On using ISN1 sequences from *S. cerevisiae*, *Candida glabrata*, *Debaryomyces hansenii*, *Candida albicans*, *Kluyveromyces lactis*, and *Gibberella zeae* as query, the search converged in the 2^{nd} iteration without retrieval of any new sequences. However, on using ISN1s from *P. falciparum*, *S. pombe*, *Aspergillus fumigatus*, *Aspergillus nidulans*, *Ashbya gossypii*, and *Yarrowia lipolytica* the PSSM constructed based on the hits from the first iteration, retrieved HAD superfamily

members in its subsequent iterations with E-values better than the threshold of 0.005. Subsequent rounds of iteration, till the search converged, yielded further HAD superfamily members. Examination of the PSI- BLAST results indicated that the HADs retrieved were either phosphomannomutases or HAD members without functional annotation. Phosphomannomutases have canonical HAD fold and are members of HAD subfamily II [Allen and Dunaway-Mariano, 2004]. These results clearly indicated that ISN1 sequences are possible members of the HAD superfamily. However, it should be noted that the PSI-BLAST iterations did not pick up IMP specific 5'-nucleotidases of the cN-II family.

Multiple sequence alignment (Fig. 1) of a representative set of 5 ISN1 members with HAD members highlighted conserved amino acids across these sequences that correspond to the HAD motifs I–IV. The most prominent motif I that contains the DXDX[V/T] sequence is conserved across HADs and ISN1s. In HADs the 1^{st} aspartate of this motif is implicated in catalysis and serves as a nucleophile forming a phosphorylated enzyme intermediate. Further, the available crystal structures of HAD superfamily of hydrolases show that motif II positions a conserved serine/threonine that binds the substrate phosphoryl group while motif III positions a conserved arginine/lysine that shields the charge in the aspartate nucleophile and the phosphoryl group. Motif IV positions 2-3 aspartate residues that bind the Mg^{2+} cofactor [Zhang *et al.*, 2004; Lu *et al.*, 2005]. The conservation of the catalytically essential motifs in ISN1 sequences further supports the results from PSI-BLAST, again indicating that ISN1 sequences belong to HAD superfamily of hydrolases. As cN-IIs are known members of HAD superfamily, all the motifs of HAD superfamily are conserved in this family of nucleotidases also(alignment not shown). However, various alignment tools such as Clustalw, MUSCLE, and Coffee-T failed to align all these motifs across ISN1s and cN-IIs indicating extensive sequence divergence. Manual curation involving deletion of non-homologous N and C terminal regions in ISN1 and cN-II sequences respectively, was needed to obtain an alignment of the DXDX[V/T] motifs across these sequences.

Protein fold of ISN1 members

The presence, location and structure of the cap domain permits classification of HADs into 3 distinct subfamilies viz., type I, II and III. The cap domain in Type I and II subfamilies comprises of an all α bundle and an α/β sandwich, respectively, while proteins in the subfamily III lack the cap altogether. The cap is inserted between loop 1 and 2 of core domain in type I and between loop 2 and 3 in type II HAD subfamilies. HAD IIa differs from HAD IIb in the topological arrangement of the alternating α/β cap domain. To obtain the structural fold of ISN1s, 14 of the ISN1 sequences were submitted to mGenthreader and nFold (http://www.biocentre.rdg.ac.uk/bioinformatics/nFOLD/) servers. Barring the first 130–160 residues, ISN1s from *S. pombe*, *A. fumigatus*, *K. lactis*, *D. hansenii*, *Y. lipolytica*, *A. nidulans*, and *P. falciparum* modeled with high reliability, upon ten scaffolds which were all members of HAD superfamily of hydrolases. The model of ISN1 sequences generated using 2AMY, a phosphomannomutase from *Homo sapiens*, as scaffold was assigned the highest score. Although the identity between the sequence of 2AMY and the modeled ISN1s (14 in number) was only about 13–20%, the alignment length of the scaffold and the query sequences was 229–243 amino acids with significant E-values. The features of predicted secondary structures were highly conserved between the template and the query sequences. The other scaffold proteins were either known phosphatases or sugar hydrolases or hypothetical proteins with HAD fold from structural genomics initiatives. The remaining ISN1 sequences (from *S. cerevisiae*, *C. albicans*, *K. lactis*, *D. hansenii*, *Magnaporthe grisea*, *A. gossypii*, *C. glabrata*, *G. zeae* and *Neurospora crassa*) either threaded onto HAD scaffold with medium to poor scores or failed to thread on any structural scaffold.

Fig. 1. Multiple sequence alignment of the four motifs of HADs with ISN1s. The histogram represents the identity, on a scale of 1–10 (1 is least identity and 10, the maximum identity is represented by *, similarity is represented by $+$) of the residue in a column. The alignment obtained from T-Coffee [Notredame *et al.*, 2000] was submitted to MUSCLE [Edgar, 2004] to generate the figure. 1NF2, 1YMQ, 1RKQ, 1WR8, 1RLM and 1U02 are PDB id's of HAD sequences obtained from mGenThreader. The accession numbers for ISN1 sequences are CAB10798 from *Schizosaccharomyces pombe*, XP_391805 from *Gibberella zeae*, XP_746360 from *Aspergillus fumigatus*, XP_451105 from *Kluyveromyces lactis*, and NP_014798 from *Saccharomyces cerevisiae*.

The model of ISN1s generated using 2AMY as scaffold possesses a Rossmanoid core and a cap slightly altered from that of the HAD subfamily IIb. The Rossman fold of the model overlaps with that in cN-II while the α/β sandwich cap is distinct from all 3 classes of cNs (Fig. 2). Loop 1 in the core domain, following the first β-sheet, has the canonical DXDX[V/T] motif with loop 2 containing the conserved threonine. The cap domain was found to occur between loop 2 and loop 3 as is known for HAD subfamily IIb members. The lysine in loop 3 of core domain was also conserved along with the conserved aspartates in loop 4. This conservation of catalytically important residues in the core domain indicates that the catalytic chemistry is also highly conserved in ISN1s. Despite the high similarity between the yeast and other ISN1s , the yeast enzyme did not model on to the HAD superfamily fold. This probably indicates a higher degree of sequence divergence of yeast ISN1 from HADs preventing threading onto this fold. However, the presence of the HAD signature motifs including the completely conserved DXDX[V/T] motif in the yeast nucleotidase suggests that this also probably belongs to the HAD superfamily.

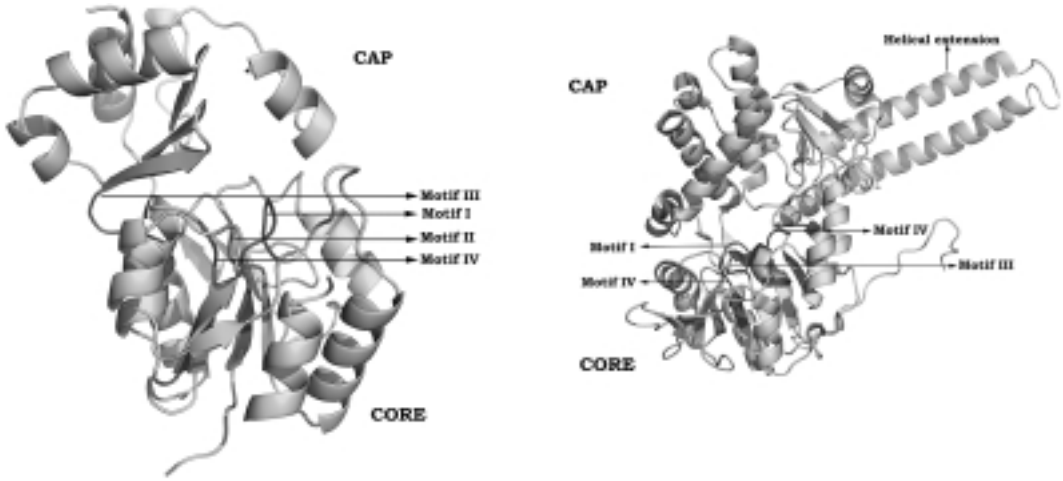

Fig. 2. Structure of cN-II (PDB ID: 2BDE) and the predicted fold of ISN1. Loops 1–4 (Motifs I–IV) are indicated in the predicted ISN1 fold and?? cN-II structure. ISN1 fold was obtained using servers nFOLD and mGenTHREADER [Jones, 1999] and the fold figure was generated using PyMOL v0.99. The arrow in the cN-II structure indicates the carboxy terminal helical segment that is absent in ISN1 structure. The cap and core in both the structures are indicated against the respective domains.

Phylogenetic relationship

cN-II and ISN1s are both IMP specific 5'-nucleotidases and our studies show that ISN1s belong to the HAD superfamily. To understand the relationship we have reconstructed phylogenetic trees using the sequences of cN-IIs, ISN1s, phosphomannomutases (PMMs) and sucrose 6-phosphate phosphatases (S6PPs). Both PMMs and S6PPs are members of HAD subfamily II and hence, were included. The entire sequences of PMMs and S6PPs were used while in the case of cN-IIs and ISN1s the amino and carboxy terminal stretches respectively, unique to these classes were eliminated to generate a meaningful multiple sequence alignment where motifs I-IV were aligned. This alignment was used to generate trees using neighbor joining or maximum parsimony program of MEGA package [Kumar *et al.*, 2004]. The two methods gave the same tree topology. The unrooted tree is shown in Fig. 3. The rooted tree displaying the robustness of each node determined using the bootstrap analysis of MEGA is shown in the inset. The phylogenetic relationship of phosphomannomutase, sucrose-6-phosphate phosphatase, ISN1s and cN-IIs indicate that the ISN1s form a distinct arm from the cN-IIs with substantial divergence among its members while cN-IIs cluster together. The tree also shows that ISN1 and cN-II class of 5'-nucleotidases appear to have evolved from different parent HADs.

The conservation of HAD superfamily sequence motifs and fold in the ISN1 class of nucleotidases and the presence of evolutionary relationships strongly supports a HAD superfamily structure. We propose to rename the ISN1 family as cN-IIb. cN-II class of purine nucleotidases possess an insertion of approximately 70 residues at their carboxy terminii (PDB id:2BDE) that folds into a helical extension shown in Fig. 2. ISN1 nucleotidases lack this segment of sequence but have a unique stretch at the N-terminal end. The model of ISN1 generated in this study covers only residues 161 to 449 leaving the first 160 residues without a structural fold. Attempts to model this domain yielded folds that were fully helical in nature, the significance of which requires experimental verification.

Search of genome databases indicated that ISN1s are present only in fungi and 4 species of Plasmodia viz., *P. knowlesi*, *P. falciparum*, *P. vivax* and *P. gallinaceum*, species that cause simian, human and chicken malaria. Though the genomes of *P. berghei*, *P. chabaudi* and *P. yoelii*, all causative agents of rodent

Fig. 3. Phylogenetic relationship between cN-IIs, ISN1s, phosphomannomutases (PMMs) and sucrose-6-phosphate phosphatases (S6PPs). The multiple sequence alignment used to construct the tree was generated using T-Coffee [Notredame *et al.*, 2000]. The phylogenetic tree was constructed using Neighbor-joining method of MEGA 3.1 [Kumar *et al.*, 2004]. The amino terminus of ∼160 residues in ISN1s and the carboxy terminus of ∼100 residues in cN-II were excluded for the tree construction as these were unique to the specific families. Each node was tested using the bootstrap approach by taking 5000 replications and a random seeding of 64238 to ascertain the reliability of nodes. The numbers indicated are in percentages against each node. The branch lengths are drawn to scale indicated. The accession numbers for the sequences are CAB10798 for ISN1_Sp from *Schizosaccharomyces pombe*, AAS50478 for ISN1_Ag from *Ashbya gossypii*, XP_503839 for ISN1_Yl from *Yarrowia lipolytica*, XP_746360 for ISN1_Af from *Aspergillus fumigatus*, XP_451105 for ISN1_Kl from *Kluyveromyces lactis*, XP_461870 for ISN1_Dh from *Debaryomyces hansenii*, XP_660826 for ISN1_En from *Emericella nidulans*, NP_014798 for ISN1_Sc from *Saccharomyces cerevisiae*, NP_036361.1 for cN-II_Hs from *Homo sapiens*, NP_084086.2 for cN-II_Mm from *Mus musculus*, AAC48784.1 for cN-II_Bt from *Bos taurus*, XP_8635921.1 for cN-II_Cf from *Canis familiaris*, NM_001015773.1 for cN-II_Xt from *Xenopus tropicalis*, NP_001026405.1 for cN-II_Gg from *Gallus gallus*, AAW32903.1 for s6pp_Nt from *Nicotiana tabacum*, CAC43285.1 for s6pp_Nsp. from *Nostoc* sp. PCC 7120, BAA13460.1 for PMM_Hs from *Homo sapiens*, XP_001078730.1 for PMM_Rsp. from *Rattus* species (*norvegicus*).

malaria, have been sequenced, ISN1 homologues could not be retrieved. The presence of ISN1 class of purine nucleotidases in fungi and Plasmodia probably arise from the specific and different biochemical requirements of these organisms. Progress in the biochemical and structural characterization would shed light on the active site, catalytic and regulatory mechanisms operating in this novel class of IMP specific 5'-nucleotidases.

ACKNOWLEDGEMENT

HB acknowledges CSIR and CSIR NMITLI for funding. BS is a recipient of CSIR-JRF fellowship.

REFERENCES

- Allen, K. N. and Dunaway-Mariano, D. (2004). Phosphoryl group transfer: evolution of a catalytic scaffold. Trends Biochem. Sci. **29**, 495-503.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389-3402.
- Bretonnet, A. S., Jordheim, L. P., Dumontet, C. and Lancelin, J. M. (2005). Regulation and activity of cytosolic 5'-nucleotidase II. A bifunctional allosteric enzyme of the Haloacid Dehalogenase superfamily involved in cellular metabolism. FEBS Lett. **579**, 3363-3368.
- Edgar, R. C.(2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**, 1792-1797.
- Hunsucker, S. A., Mitchell, B. S. and Spychala, J. (2005). The 5'-nucleotidases as regulators of nucleotide and drug metabolism. Pharmacol. Ther. **107**, 1-30.
- Itoh, R., Saint-Marc, C., Chaignepain, S., Katahira, R., Schmitter, J.-M. and Daignan-Fornier, B. (2003). The yeast *ISN1* (*YOR155c*) gene encodes a new type of IMP-specific 5'-nucleotidase. BMC Biochem. **4**, 4.
- Jones, D. T. (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J. Mol. Biol. **287**, 797-815.
- Kumar, S., Tamura, K. and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. **5**,150-163.
- Lu, Z., Dunaway-Mariano, D. and Allen, K. N. (2005). HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. Biochemistry **44**, 8684-8696.
- Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302**, 205-217.
- Zhang, G., Morais, M. C., Dai, J., Zhang, W., Dunaway-Mariano, D. and Allen, K. N. (2004). Investigation of metal ion binding in phosphonoacetaldehyde hydrolase identifies sequence markers for metal-activated enzymes of the HAD enzyme superfamily. Biochemistry **43**, 4990-4997.