Journal of Structural Biology 190 (2015) 38–46

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

The insertion domain 1 of class IIA dimeric glycyl-tRNA synthetase is a rubredoxin-like zinc ribbon

Gurmeet Kaur, Srikrishna Subramanian *

CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh, India

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form 11 February 2015
Accepted 12 February 2015
Available online 23 February 2015

Keywords:
Glycyl-tRNA synthetase
Ap4A
Zinc finger
Insertion domain 1
Aminoacyl tRNA synthetase
Hereditary motor neuropathy
Charcot–Marie–Tooth disease

ABSTRACT

The insertion domain 1 (ID1) of class IIA dimeric glycyl-tRNA synthetase (α_2GRS) is an appended domain in the core catalytic region of the enzyme. ID1 has been shown to play a role in tRNA aminoacylation, mediating interaction with the acceptor arm of tRNA and diadenosine tetraphosphate (Ap4A) synthesis. Mutations in α_2GRS, including those in the ID1 region, have been implicated in distal hereditary motor neuropathy-V (dHMN-V) and Charcot–Marie–Tooth (CMT) disease. Through sequence and structure based evolutionary analysis, we show that ID1 of α_2GRS is a rubredoxin-like zinc ribbon domain. The zinc-chelating cysteines of ID1 are well conserved in all archaeal versions of the enzyme and also in several eukaryotes, which most likely have acquired them via horizontal gene transfer from bacteria; but in all other eukaryotes, the zinc-chelating residues are not preserved. ID1 from bacteria display a selective preservation of zinc-binding residues, ranging from complete conservation to complete loss. The ID1 from different organisms harbor variable-sized non-conserved insertions between the two zinc-binding half-sites of the zinc ribbon. Three of the previously identified CMT-associated mutations in α_2GRS, viz., human D146N, mouse C157R and human S211F, are located in the zinc ribbon region of ID1. Interestingly, human Asp146 which is implicated in the synthesis of Ap4A, a molecule known to act during neuronal transmission, has also been reported to be mutated in dHMN-V, suggesting a possible link between hereditary motor neuropathy and Ap4A synthesis.

1. Introduction

Aminoacyl-tRNA synthetases (aaRSs) are enzymes that help in the translation of genetic information (Söll and Schimmel, 1974). Glycyl-tRNA synthetase (GRS) is an interesting enzyme in that two phylogenetically and structurally different forms are known (O'Donoghue and Luthey-Schulten, 2003; Shiba, 2005; Woese et al., 2000). A tetrameric form ($(\alpha\beta)_2$) of GRS is exclusively found in bacteria and a dimeric (α_2) form is present in eukaryotes, archaea and some bacteria (Fig. 1A) (O'Donoghue and Luthey-Schulten, 2003; Woese et al., 2000). The $(\alpha\beta)_2$GRS is related to the class IIC aaRSs (Phe and $(\alpha\beta)_2$GRS) and the α_2GRS belongs to class IIA (Ser, Pro, Thr, His and α_2GRS) (O'Donoghue and Luthey-Schulten, 2003). In the $(\alpha\beta)_2$GRS, the α-subunit harbors the catalytic domain and the β-subunit is the anticodon-binding domain, whereas α_2GRS displays a modular arrangement with both the catalytic and anticodon-binding domains present in a single polypeptide chain (Cader et al., 2007; Logan et al., 1995; Mazauric et al., 1998; O'Donoghue and Luthey-Schulten, 2003) (Fig. 1A).

The catalytic domain of α_2GRS has a typical class II aaRS fold (SCOP identifier 55680) made up of a seven-stranded antiparallel β-sheet (Logan et al., 1995) (Fig. 1B). Three consensus sequence motifs define class II aaRSs; wherein motif 1 (Gϕxxϕxx<u>P</u>ϕϕ) mediates dimerization while motif 2 (F<u>R</u>xE-loop-(H/R)xxxFxxx(D/E)) and motif 3 (GϕGϕGϕ(D/E)<u>R</u>ϕϕϕϕ) (where x is any aminoacid, ϕ is a hydrophobic aminoacid, underlined residues are absolutely conserved) contribute crucial active site residues (Cusack, 1995; Eriani et al., 1990) (Fig. 1A). Motif 2 and motif 3 are preserved in α_2GRS, but motif 1 is atypical, in that the invariant proline residue can be substituted by other aminoacids (Logan et al., 1995; Mazauric et al., 1998). The structure of the anticodon binding domain consists of a mixed β-sheet flanked by α-helices (SCOP fold identifier 52953) (Cader et al., 2007; Logan et al., 1995). In some α_2GRS, N- and C-terminal extensions and additional appended domains are also observed (Logan et al., 1995; Mazauric et al., 1998; Shiba, 2005).

Insertion domain 1 (ID1) is an ~85 residue domain in the structurally characterized α_2GRS (residues 144–228 in human, PDB

* Corresponding author at: CSIR-Institute of Microbial Technology (IMTECH), Sector 39-A, Chandigarh 160036, India. Fax: +91 1722695215.
E-mail address: krishna@imtech.res.in (S. Subramanian).

http://dx.doi.org/10.1016/j.jsb.2015.02.004
1047-8477/

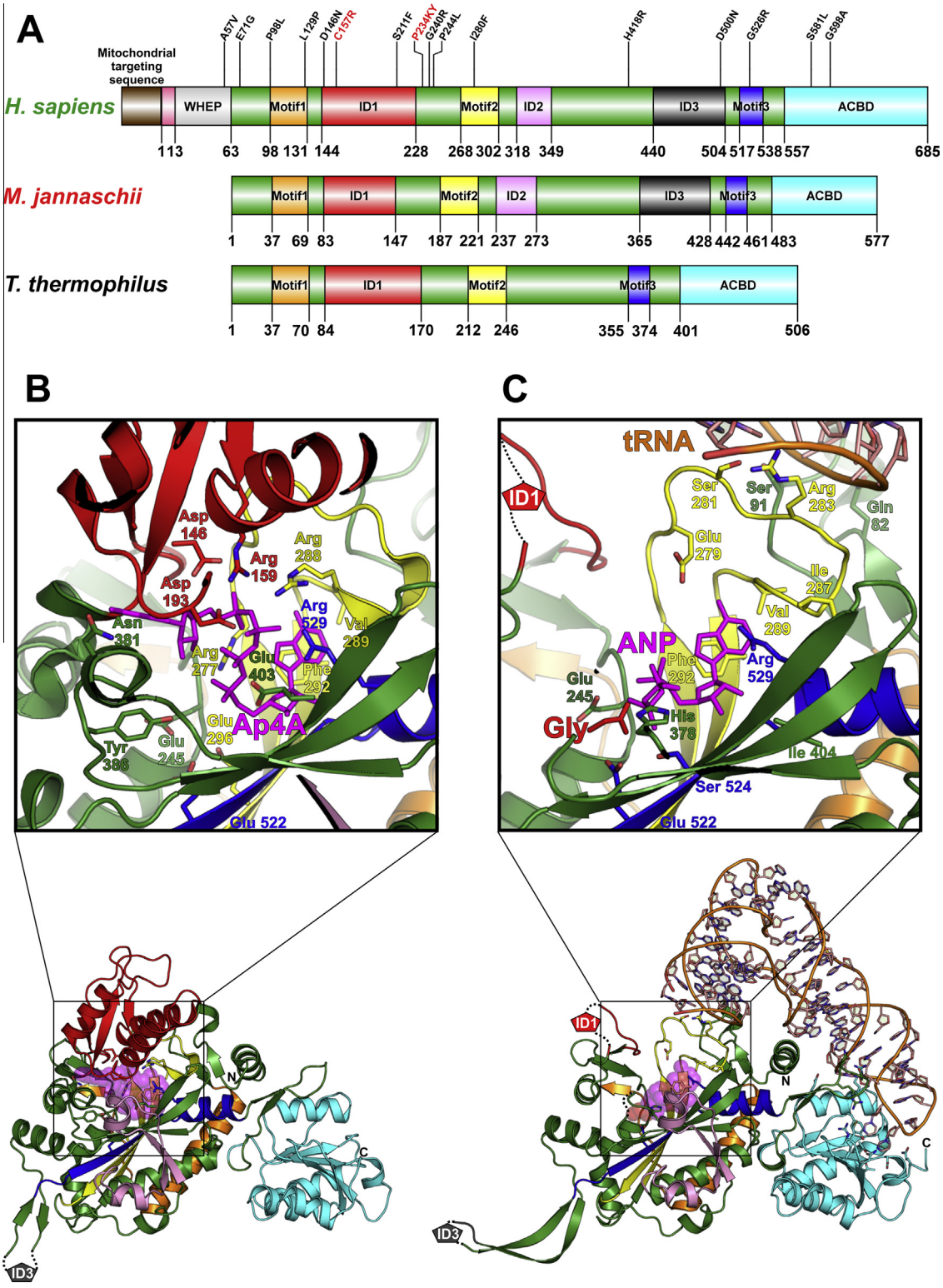

Fig.1. Domain organization of α_2GRS. (A) The domain architectures of representative members of the class IIA GRS are shown. Catalytic domain is colored green and various inserted domains and motifs are indicated. Sixteen known mutations in GRS which lead to CMT and related neuropathies are marked on the human α_2GRS. The mutations that have been identified in humans are labeled black and those identified in mouse are labeled red. (B) Ribbon diagram of human α_2GRS with Ap4A bound at the active site (PDB identifier 2ZT5). (C) Ribbon diagram of human α_2GRS with tRNA, ANP and glycine bound at the active site (PDB identifier 4KR3). Insets reveal a detailed view of the active site pockets and interactions of various residues with the bound moieties. Ap4A and ANP are colored in magenta and glycine at the active site is colored red and shown as spheres and sticks in the full structure and in the inset, respectively. The side chains of various interacting residues are shown as sticks. ANP, Ap4A, tRNA and Gly at the active site and various key residues are labeled. The coloring scheme of the various regions of GRS structures in (B, C) follows that used in (A). Abbreviations: ID (insertion domain), ACBD (anticodon binding domain), ANP (phosphoaminophosphonic acid-adenylate ester), *H. sapiens* (*Homo sapiens*), *M. jannaschii* (*Methanococcus jannaschii*), *T. thermophilus* (*Thermus thermophilus*).

identifier 2ZT5_A; residues 83–169 in *Thermus thermophilus*, PDB identifier 1ATI_A), inserted in the catalytic domain in between motifs 1 and 2 (Logan et al., 1995; Wu et al., 2013) (Fig. 1B). ID1 is exclusively present in all α_2GRS and is absent from class IIC GRS and other aaRSs (Mazauric et al., 1998). ID1 is also absent in PolγB, an accessory subunit of animal mitochondrial DNA polymerase γ, which shows extensive sequence and structural similarity to α_2GRS (Carrodeguas et al., 2001; Wolf and Koonin, 2001). ID1 is of particular interest as it has been implicated in various functions performed by α_2GRS, viz., aminoacylation of tRNA, interaction with the acceptor arm of tRNA and interactions with the second ATP molecule during diadenosine tetraphosphate (Ap4A) synthesis (Guo et al., 2009; Logan et al., 1995; Mazauric et al., 1998; Qin et al., 2014) (Fig. 1B and C). Ap4A is a diadenosine oligophosphate (ApnA) whose synthesis is catalyzed by aaRSs in addition to their primary role of aminoacylating tRNA (Guo et al., 2010; Park et al., 2005). Ap4A has previously been identified as a signaling molecule and its role as a neurotransmitter/neuromodulator is also well established (Kisselev et al., 1998; Miras-Portugal et al., 1998). Deletion of ID1 from human α_2GRS has been shown to result in almost complete loss of aminoacylation activity of the enzyme (Qin et al., 2014). Further, sixteen mutations in α_2GRS including three in the ID1 region have been implicated in Charcot–Marie–Tooth subtype 2D (CMT-2D) and distal hereditary motor neuropathy-V (dHMN-V) (Achilli et al., 2009; Griffin et al., 2014; Lee et al., 2012; Motley et al., 2010; Yao and Fox, 2013).

The ID1 region in the initially characterized structures of α_2GRS from *T. thermophilus* (PDB identifiers 1ATI, 1B76, 1GGM) was mostly disordered (Arnez et al., 1999; Logan et al., 1995). This disorder in structure is now suggested to be a consequence of the conformational flexibility exhibited by this domain, which facilitates Ap4A synthesis and binding to the acceptor arm of tRNA (Guo et al., 2009; Qin et al., 2014). Current protein sequence and structure classification schemes such as Conserved Domain Database (CDD) (Marchler-Bauer et al., 2011), Pfam (Punta et al., 2012), SCOP (Murzin et al., 1995) and CATH (Sillitoe et al., 2013), co-classify ID1 along with the core catalytic domain of α_2GRS. Since ID1 is not a part of the conserved catalytic core of class II aaRS but none-the-less appears to play important functional roles, we were interested in understanding the evolutionary origin of this domain. Using sequence and structure similarity, we show that the ID1 of α_2GRS is a rubredoxin-like zinc ribbon domain in which the metal-chelating aminoacids are not absolutely conserved in all α_2GRS family members.

2. Methods

2.1. Sequence similarity searches

The NCBI non-redundant (NR) protein sequence database (NR; Nov 17, 2014; 52,724,511 sequences; 18,962,362,785 total letters) was searched using the PSI-BLAST program (Altschul et al., 1997) to retrieve sequences similar to ID1 domain of α_2GRS (PDB identifier 2ZT5_A; residues 144-228). An *E*-value threshold of 0.001 was used to run iterative PSI-BLAST searches until convergence was reached. Homologous protein sequences obtained from this step were used to initiate new PSI-BLAST searches to obtain additional α_2GRS sequences.

Additional sequence similarity searches initiated with the ID1 region of human α_2GRS (PDB identifier 2ZT5_A; residues 144-228) were performed using JackHMMER program of the HMMER3 package (Finn et al., 2011) (against: NR version 2014-03-08; *E*-value threshold of 0.01), FFAS server (Jaroszewski et al., 2005) (against PDB, Pfam and SCOP databases) and the HHpred server (Soding et al., 2005) (against: PDB70_29May14 and PfamA_27.0, using MSA generation method HHblits run for 5 iterations, *E*-value threshold of 0.001).

2.2. Multiple sequence alignment

Sequences similar to human ID1 retrieved in the JackHMMER search were analyzed manually. Owing to large-sized insertions in the non-conserved regions of ID1 automated JackHMMER search was not able to align equivalent regions in ID1s from several homologous α_2GRS. Thus for these proteins, regions equivalent to human ID1 (residues 144–228 in PDB identifier 2ZT5) were manually delineated using the conserved motifs of α_2GRS as a guide. The manually refined dataset of ID1 sequences were used to create multiple sequence alignment (MSA) using the ClustalW program (Larkin et al., 2007) within the BioEdit software package (version 7.2.2) (Hall, 1999) with default parameters. Misaligned regions in the ClustalW-generated MSA were adjusted manually.

2.3. Conserved domain search

Cd-hit tool (Li and Godzik, 2006) was used to perform a sequence-based clustering of the ID1 sequences. Clustering was performed at a cut-off of 80% sequence identity with a word size of 5 and other default parameters. Conserved domain (CD) search of representative ID1 sequences obtained after cd-hit clustering was performed against the CDD with a cut-off *E*-value of 0.01 (Marchler-Bauer et al., 2011) to infer evolutionary connections of ID1 with other CDs.

2.4. Structure based methods

Dali, TM-align and Fr-TM-align tools were used to evaluate structural similarity of ID1 with other proteins (Holm and Sander, 1995; Zhang and Skolnick, 2005; Pandit and Skolnick, 2008). Fifteen structures of full length α_2GRS from human and *T. thermophilus* were available in the Protein Data Bank (PDB) as of 30 May 2014. The structures of human α_2GRS with structurally ordered ID1 (PDB identifier 2ZT5_A, 2PMF_A) were used as queries for automated structure similarity searches. The structures were visually compared and manually superimposed with other zinc ribbon domains by defining the equivalent regions using the pair fitting command of the molecular visualization program PyMOL.

2.5. Delineating domain architectures

DOG (Domain Graph, version 1.0) software (Ren et al., 2009) was used to build the domain architectures of GRSs (Fig. 1). Domain boundaries and functionally important sites were obtained from available literature and similarity of the CDs to Pfam families. Sequences of α_2GRS from *Homo sapiens* (PDB identifier 2ZT5_A), *Methanococcus jannaschii* (UniProt ID: Q57681) and *T. thermophilus* (UniProt ID: P56206) were used for building the domain architectures.

3. Results and discussion

The secondary structure of ID1 (PDB identifier 2ZT5_A), consists of an N-terminal β-hairpin, followed by a three-helical bundle, a small knuckle and a C-terminal β-strand. The N-terminal β-hairpin and the C-terminal β-strand form an antiparallel β-sheet which connects to the catalytic core of the GRS. Although there are no zinc-chelating aminoacids in the structurally characterized ID1s, the overall shape of the zinc knuckles is preserved (Fig. 2A). The first zinc-chelating half-site (residues: 151–155, PDB identifier 2ZT5_A) is present at the turn of the N-terminal β-hairpin and

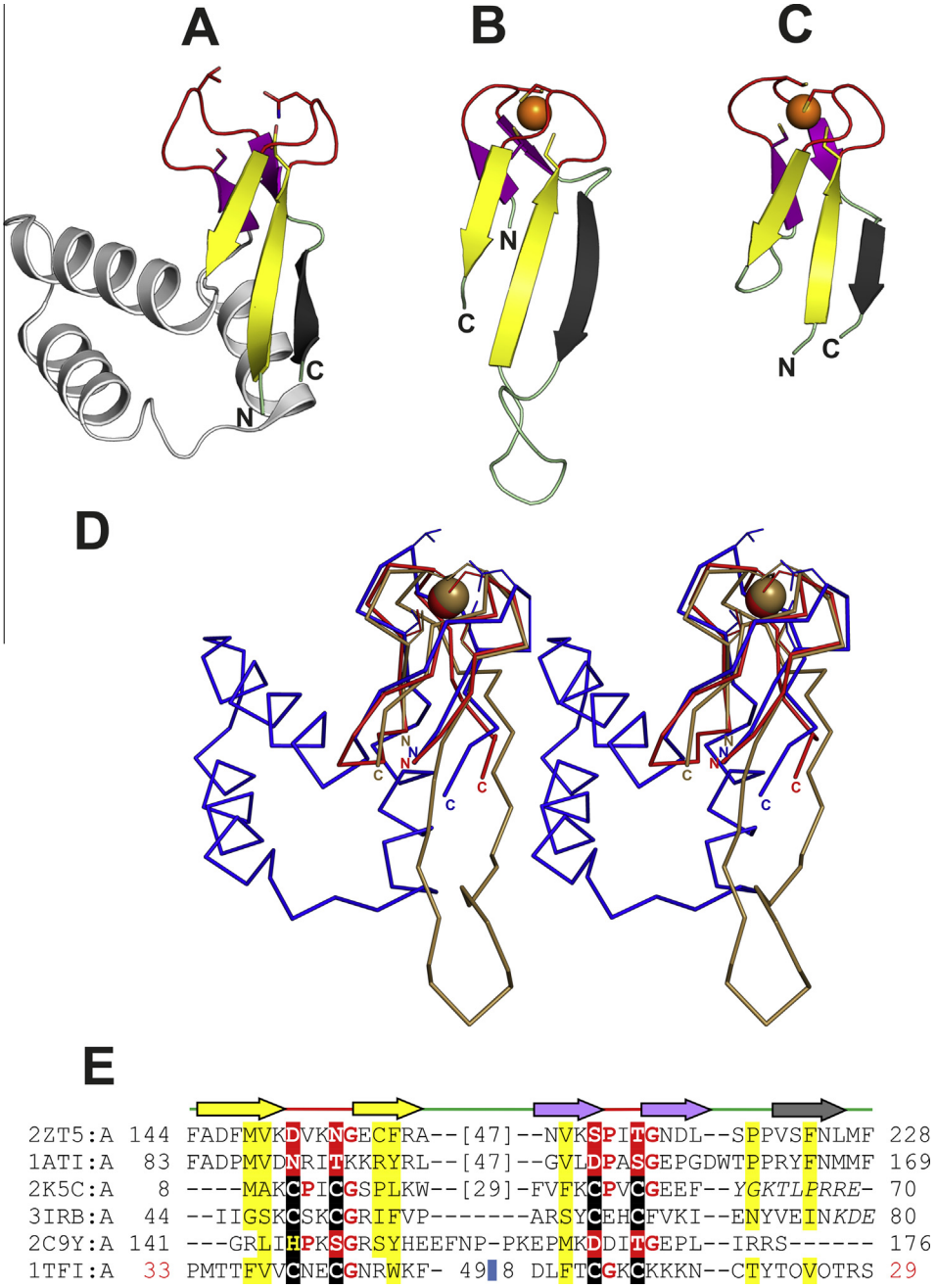

Fig.2. Structure and sequence comparison of ID1 and other zinc ribbon domains. (A) ID1 from human α_2GRS (PDB identifier 2ZT5_A). (B) A classical zinc ribbon domain (PDB identifier 1TFI_A). (C) A circularly permuted rubredoxin-like zinc ribbon domain (PDB identifier 3IRB_A). In these figures, the N-terminal β-hairpin of the classical zinc ribbon is colored purple, the C-terminal β-hairpin is colored yellow, the additional β-strand that forms three-stranded β-sheet with the N-terminal β-hairpin is colored grey and the zinc knuckles are colored red. The secondary structure elements that are not a part of the core of the zinc ribbon domain are colored white. Equivalent secondary structure elements in (A, B, C) are colored alike. Zinc ion is shown as an orange sphere and side chains of zinc-chelating aminoacids and equivalent residues in ID1 are represented in stick form. (D) Stereo diagram of manually superimposed circularly permuted rubedoxin-like zinc ribbon domain (PDB identifier 3IRB_A, colored red), ID1 of GRS (PDB identifier 2ZT5_A, colored blue) and classical zinc ribbon domain of TFIIS (PDB identifier 1TFI_A, colored sand). The zinc ions are shown as spheres and the side chains of zinc-chelating ligands and equivalent residues in ID1 are shown as lines in the respective colors of the proteins. (E) Structure based sequence alignment of the aforementioned proteins. Secondary structure diagram is shown above the alignment. PDB identifiers and start and end aminoacid numbers are indicated. The zinc-binding residues have been highlighted black and other aminoacids at equivalent position are highlighted red. Regions where the structures are not superimposable are shown in italics. Regions of circular permutation (in PDB identifier 1TFI) are separated by a small blue rectangular box and sequence numbers of the regions around the circular permutation are indicated in red. Small amino acids (Gly, Pro) in the vicinity of the zinc-binding ligands are colored in red. Uncharged residues (all aminoacids except Asp, Glu, Lys and Arg) in mostly hydrophobic sites are highlighted in yellow. Long insertions are not shown and the number of omitted residues is specified in brackets.

the second half-site (residues: 211–215, PDB identifier 2ZT5_A) is present as a small knuckle between the helical bundle and C-terminal β-strand. The structure of ID1 resembles a zinc ribbon domain with an inserted helical bundle.

Zinc ribbons are one of several zinc-stabilized protein folds and consist of an all-β tertiary structure (Krishna et al., 2003). Characteristic zinc knuckle motifs (CPxCG) present at the turns of the β-hairpins chelate a structure-stabilizing zinc ion (Fig. 2B and C) (Krishna et al., 2003). Zinc ribbon domains are found in a large number of proteins that are involved in diverse cellular processes and predominantly function to mediate interactions with nucleic acids and proteins (Krishna et al., 2003; Krishna et al., 2010).

Dali structure similarity search initiated with the core zinc ribbon region of ID1 (i.e. with the metal-chelating regions only; residues: 144–159 and 207–228, PDB identifier 2ZT5_A), is able to find matches to many other zinc ribbons domains from proteins, such as the DUF35 family protein (PDB identifier 3IRB_B, *Z*-score = 3.0, RMSD = 1.8 Å, nali = 35), methionyl-tRNA synthetase (PDB identifier 3KFL_A, *Z*-score = 2.9, RMSD = 2.4 Å, nali = 35), endoribonuclease Nob1 (PDB identifier 2LCQ_A, *Z*-score = 2.3, RMSD = 2.3 Å, nali = 33) and adenylate kinase (PDB identifier 3NDP_B, *Z*-score = 2.2, RMSD = 2.6 Å, nali = 31) (Holm and Sander, 1995; Krishna et al., 2003; Krishna et al., 2010; Veith et al., 2012). The helical insertion in between the two halves of the zinc ribbon of ID1 was not included while searching for structural similar proteins as it retrieves a large number of spurious matches to unrelated helical regions from different proteins. Comparable structural similarity results were obtained by using TM-align and Fr-TM-align algorithms (Zhang and Skolnick, 2005; Pandit and Skolnick, 2008) to superimpose ID1 (residues: 144–228 of PDB identifier 2ZT5_A) on the zinc ribbon domain matches obtained from Dali. Most of these structures could be superimposed with a TM-align score of more than 0.5 suggesting a fold level similarity between ID1 and bonafide zinc ribbon domains.

The zinc ribbon of ID1 represents a rubredoxin-like circularly permuted form of the classical zinc ribbon domain present in many DNA binding transcription factors such as TFIIS, TFIIB and eIF2β (Fig. 2B and C) (Krishna et al., 2003). Circularly permuted zinc ribbons, typified by the rubredoxins, are present in a number of enzymes including Class I aaRSs such as Val, Met, Ile and Leu aaRSs (Krishna et al., 2003). In all these aaRSs the zinc ribbon domain, similar to ID1, is inserted in the catalytic core and plays important roles in activation of aminoacids and charging of tRNA (Ibba et al., 2005; Krishna et al., 2003; Zhou et al., 2008). Considering a circular permutation allows for proper superimposition of all the structurally and functionally equivalent secondary structure elements of the zinc ribbon fold, namely a β-hairpin on one side and a three-stranded β-sheet on the other (Krishna et al., 2003). Since automated structure alignment programs do not always consider circular permutations, thus, often satisfactory structural alignments can only be obtained by manually defining equivalent regions for structure superimposition. For example, we could manually superimpose the rubredoxin-like zinc ribbon of ID1 (PDB identifier 2ZT5_A) and that of DUF35 protein (PDB identifier 3IRB_A) with an RMSD = 1.6 Å over 30 backbone C_α atoms (Fig. 2D). This superimposition is comparable to that obtained by automated methods. However, the rubredoxin-like permuted zinc ribbon of ID1 (PDB identifier 2ZT5_A) and classical zinc ribbon domain from TFIIS (PDB identifier 1TFI_A) could not be superimposed by using automated pairwise Dali tool and were misaligned using TM-align. These structures could only be aligned manually upon assuming a circular permutation with an RSMD = 1.5 Å over 23 pairs of backbone C_α atoms.

Sequence similarity based assessment likewise suggests a zinc ribbon origin for ID1. Sequence similarity searches were carried out to detect remote homologs of α_2GRS ID1 (PDB identifier 2ZT5_A) (see Section 2 for details). A structure-guided MSA of representative homologs shows that, despite the fact that ID1 of structurally characterized α_2GRS lacks metal-chelating aminoacids, two pairs of cysteine residues are present in many homologous proteins that map onto the zinc knuckles and would likely help bind a divalent metal ion (Fig. 3). The MSA reveals large variations in the length of ID1 in different organisms, ranging from 39 to 165 aminoacids. The most conserved regions, however, encompass only the core zinc ribbon domain. The N-terminal β-hairpin of the zinc ribbon contains a highly conserved Asp and Arg residue (Asp146, Arg159 in human α_2GRS) that have been implicated in binding the phosphate group adjacent to the second adenosine in Ap4A (labeled * symbol in Fig. 3) (Guo et al., 2009). Besides these, conserved aromatic residues are present at four positions in the ID1 zinc ribbon domain, two on the N-terminal β-hairpin and two on the C-terminal strand (highlighted cyan in Fig. 3).

We examined the pattern of the loss of metal-chelating residues in ID1 by mapping the number of metal-chelating residues present in various species that have been used in the phylogenetic trees of α_2GRS by Woese et al. (2000) and Wu et al. (2013). Our analysis reveals that zinc chelating residues are mostly preserved in all archaea and lost in a majority of eukaryotes. The bacterial clade displays a selective distribution with respect to the preservation of metal-chelating residues across different lineages ranging from complete conservation to complete loss of zinc-chelating residues. Further, we analyzed all the ID1s from α_2GRS proteins (3653 unique sequences) that were obtained in the JackHMMER search. Six sequences annotated as synthetic constructs and from uncultured organisms without any taxonomic classification were not considered in the analysis. The average length of ID1 is 66 residues in archaea (shortest: 62, longest: 73, total number of sequences: 293), 95 residues in eukaryotes (shortest: 62, longest: 165, total number of sequences: 719) and 74 residues in bacteria (shortest: 39, longest: 142, total number of sequences: 2635). In some Nanoarchaea (gi_41615203, gi_490715429) and Crenarchaea (gi_519044377) partial degradation of the zinc-chelation site is observed as ID1 from these organisms have only three zinc-chelating cysteines, although Asp and His residues in the vicinity of the zinc-binding site could potentially help chelate a divalent metal-ion.

Although most of the eukaryotic α_2GRS have lost all zinc-chelating cysteines, some like *Acanthamoeba castellanii* (gi_470397308) and *Rhizopus delemar* (gi_ 384499000) exhibit partial conservation of the zinc-chelating residues and have two and one zinc-chelating cysteines, respectively. It was also observed that many eukaryotes possess more than one copy of α_2GRS, as suggested previously (Shiba, 2005). In *Saccharomyces cerevisiae*, a duplication event has been suggested to have resulted in the formation of two copies of α_2GRS (Wu et al., 2013). This proposition is supported by the BLAST sequence similarity search, in which searches initiated with either GRS1 or GRS2 from *S. cerevisiae* are able to find each other among top matches that were retrieved. However, gene duplication is unlikely to be the sole reason for the presence of multiple copies of α_2GRS in all cases. For example, many members of the Stramenopiles lineage have more than one α_2GRS-like protein with none to all four zinc-binding cysteines in their ID1s. Based on BLAST sequence similarity search and phylogenetic analysis (Supplementary methods and Supplementary Fig. 1), the copy of α_2GRS with all four zinc-chelating cysteines in ID1of Stramenopiles species, such as *Saprolegnia diclina* VS20, *Aphanomyces invadans* and *Phytophthora infestans* T30-4 (gi|574470727, gi|301109960, gi|530742713, respectively), appear to have been acquired though horizontal gene transfer (HGT) from bacteria.

In the bacterial α_2GRS, different clades exhibit varied levels of conservation of metal-chelating residues at the zinc knuckles. For example 75% of Actinobacteria, 50% of Firmicutes, 50% of Tenericutes and most of the Proteobacteria, Spirochaetes and Chloroflexi sequences retrieved by the JackHMMER search, possess four zinc-chelating cysteines, whereas most members of the Bacteroidetes, Chlorobi, Verrucomicrobia, Deinococcus-Thermus and Planctomycetes clades have lost one or more of the zinc-chelating cysteines. Previous reports have suggested the absence of ID1 from some archaeal species such as *Thermoplasma volcanium* and *Thermoplasma acidophilum* and from proteobacteria (Shiba, 2005; Wolf and Koonin, 2001). However, our sequence analysis reveals the presence of ID1 in all archaeal species and in some α-, γ- and δ-proteobacterial lineages, most of which have all four metal-chelating cysteines in the ID1.

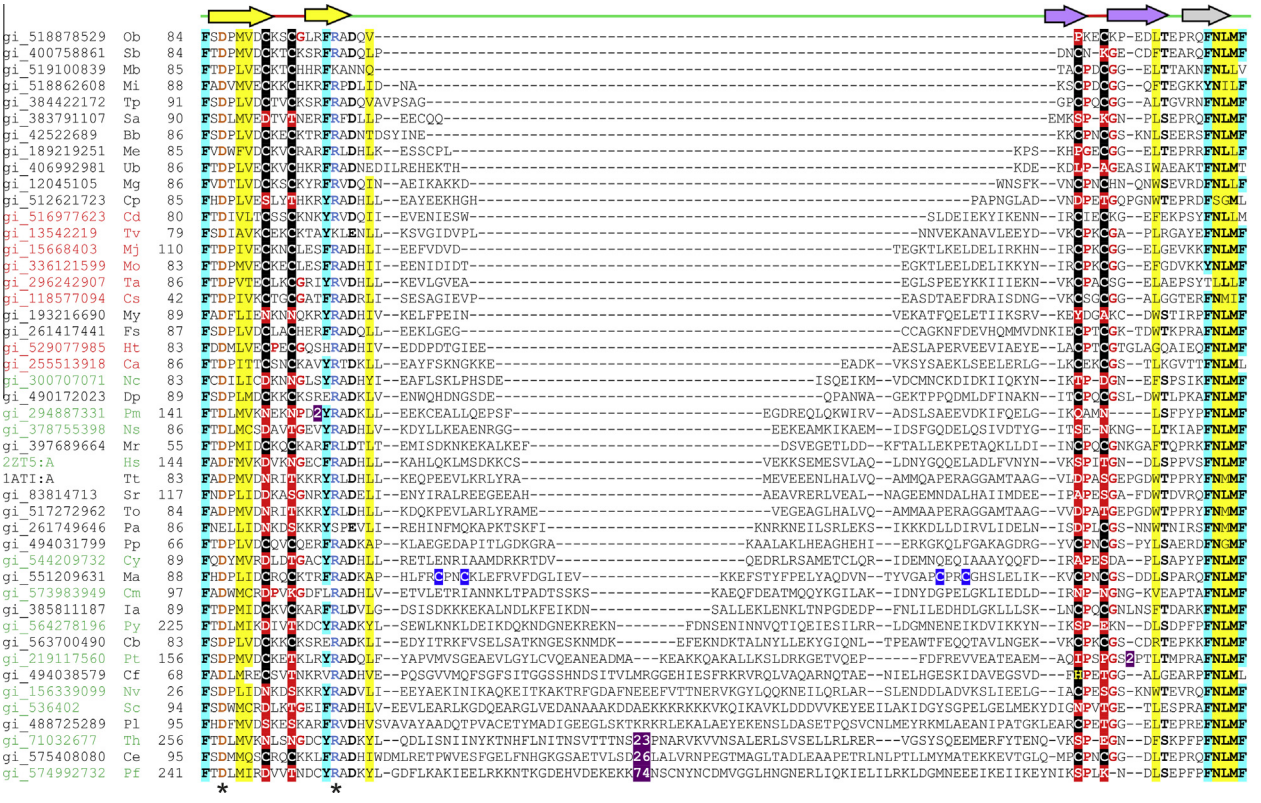

Fig.3. Structure-based multiple sequence alignment of representative sequences of α_2GRS ID1. Gene identification (gi) number/PDBid, organism name abbreviation, the first and the last residue numbers of the regions depicted in the alignment are indicated for each sequence. The secondary structural elements of the zinc ribbon are indicated above the alignment. Long insertions are not shown and the numbers of omitted residues are represented by numbers boxed in violet. Potential metal-binding ligands are boxed in black and non-metal-binding residues at the same position are boxed in red. Secondary zinc binding sites are boxed in blue. Uncharged residues (all aminoacids except Asp, Glu, Lys and Arg) in mostly hydrophobic sites are highlighted yellow and conserved aromatic residues are highlighted cyan. The conserved residues involved in interaction with Ap4A are indicated by an asterisk sign (*). Gi's and organism name of protein sequences from archaea, bacteria and eukaryotes are colored in red, black and green respectively. The organism abbreviations are: Ob – *Oceanospirillales bacterium* SCGC AAA298-N10, Sb – SAR86 cluster bacterium SAR86B, Mb – *Microgenomates bacterium* SCGC AAA040-P11, Mi – *Microgenomates bacterium* SCGC AAA011-J08, Tp – *Treponema pallidum* subsp. *pallidum* str. Chicago, Sa – *Spirochaeta africana* DSM 8902, Bb – *Bdellovibrio bacteriovorus* HD100, Me – *Methylacidiphilum infernorum* V4, Ub – uncultured bacterium, Mg – *Mycoplasma genitalium* G37, Cp – *Corynebacterium pyruviciproducens*, Cd – candidate division YNPFFA, Tv – *Thermoplasma volcanium* GSS1, Mj – *Methanocaldococcus jannaschii* DSM 2661, Mo – *Methanothermococcus okinawensis* IH1, Ta – *Thermosphaera aggregans* DSM 11486, Cs – *Cenarchaeum symbiosum* A, My – *Mycoplasma arthritidis* 158L3-1, Fs – *Fibrobacter succinogenes* subsp. *succinogenes* S85, Ht – *Halorhabdus tiamatea* SARL4B, Ca – *Candidatus Micrarchaeum acidiphilum* ARMAN-2, Nc – *Nosema ceranae* BRL01, Dp – *Desulfobacter postgatei*, Pm – *Perkinsus marinus* ATCC 50983, Ns – *Nematocida* sp. 1 ERTm2, Mr – *Melioribacter roseus* P3M-2, Hs – *Homo sapiens*, Tt – *Thermus thermophilus*, Sr – *Salinibacter ruber* DSM 13855, To – *Thermus oshimai*, Pa – *Blattabacterium* sp. (*Periplaneta americana*) str. BPLAN, Pp – *Plesiocystis pacifica*, Cy – *Cyanidioschyzon merolae* strain 10D, Ma – *Marinimicrobia bacterium* JGI 0000039-D08, Cm – *Cordyceps militaris* CM01, Ia – *Ignavibacterium album* JCM 16511, Py – *Plasmodium yoelii* 17X, Cb – candidate division SR1 bacterium RAAC1_SR1_1, Pt – *Phaeodactylum tricornutum* CCAP 1055/1, Cf – *Chthoniobacter flavus*, Nv – *Nematostella vectensis*, Sc – *Saccharomyces cerevisiae*, Pl – *Planctomyces maris*, Th – *Theileria parva* strain *Muguga*, Ce – *Candidatus Entotheonella* sp. TSY1, Pf – *Plasmodium falciparum* Palo Alto/Uganda.

The region in between the two zinc-binding half-sites is poorly conserved, both in terms of identity of aminoacid residues and its length, which leads to the large variations seen in full-length ID1s across different species, even among close homologs. The ID1 in the structurally characterized α_2GRS (PDB identifier 2ZT5_A) has an all-helical insertion in between the two zinc-binding half sites. However, the corresponding region in other α_2GRS might not have identical tertiary structure. For example, we predict that the region between the two zinc knuckles of ID1 of *Marinimicrobia bacterium* JGI 0000039-D08 (see gi|551209631 in Fig. 3) is likely to contain a second zinc ribbon domain, probably related by duplication to the zinc ribbon of ID1. α_2GRS1 from *S. cerevisiae*, which has a much longer sequence stretch between the two zinc-binding half sites, has also been proposed to have additional β-pleated regions within the ID1 (Wu et al., 2013). The inserted regions within the zinc ribbon would likely help enhance the functions of ID1 and/or perform additional functions in an organism specific manner. For example, the helical region in the ID1 of human and *T. thermophilus* has been suggested to assist in tRNA interaction through its positively charged aminoacids (Cader et al., 2007; Logan et al., 1995). Similarly in *S. cerevisiae*, the additional β-pleated region within the ID1 of α_2GRS1 possesses a lysine rich K-motif (see gi|536402 in Fig. 3) which is suggested to play a role in catalysis and binding to cognate-tRNA (Wu et al., 2013). Asp193 from the α-helical region of human α_2GRS ID1 has been shown to form a hydrogen bond with the adenosine base of the second ATP during Ap4A synthesis (Guo et al., 2009).

Ap4A synthesis by α_2GRS, in contrast to a similar process mediated by other aaRSs, does not require the presence of cognate aminoacid, resulting in a direct condensation of two ATP molecules (Guo et al., 2009). This exclusive feature of α_2GRS has been proposed to have implications in homoeostasis of Ap4A (Guo et al., 2009). ID1 assists this process by contributing to the formation of a unique binding pocket for a second substrate ATP molecule, that condenses with an ATP molecule bound at its typical class II aaRS active site (Guo et al., 2009). Side chains of Asp146 and Arg159 from the zinc ribbon of human ID1 specifically interact with the fourth phosphate moiety of Ap4A (Guo et al., 2009). These two aminoacids are suggested to be strictly conserved in all α_2GRS, which indicates that Ap4A synthesis may also be a conserved function of all α_2GRS (Guo et al., 2009). However, MSA of ID1 (Fig. 3) reveals that in some organisms such as *Blattabacterium*

sp. (*Periplaneta americana*) str. BPLAN and *Microgenomates bacterium* SCGC AAA040-P11, substitutions of these two aminoacids may occur. Although most of the substitutions are conservative in nature with chemically similar aminoacids making the replacements (see gi_519100839, gi_261749646 in Fig. 3), some ID1 sequences with non-conservative aminoacid substitutions were also observed in our analysis (for example, gi_261749646 in Fig. 3, which has Ser in place of Arg). Further, the Ap4A synthesis by human α_2GRS has been shown to be inhibited by the presence of zinc ions, contrary to effects seen on the two most efficient Ap4A producers, LysRS and PheRS (Goerlich et al., 1982; Guo et al., 2009). This is intriguing as ID1s in organisms with four zinc-chelating cysteines would actually bind zinc with a structural consequence. Thus, it remains to be seen how zinc would actually regulate ApnA production by α_2GRS in those organisms.

Of all the identified mutations in the gene coding for α_2GRS, *GARS*, which have been previously identified to cause neuromuscular disorders such as CMT-2D and dHMN-V (Antonellis et al., 2003; Griffin et al., 2014; Lee et al., 2012; Motley et al., 2010; Yao and Fox, 2013), three are located in the conserved zinc ribbon region of ID1 (Achilli et al., 2009; Lee et al., 2012; Qin et al., 2014). Of the three disease-related mutations identified in ID1, Asp200 of human α_2GRS (equivalent to Asp146 in PDB identifier 2ZT5) is mutated to Asn in some diseased patients (Lee et al., 2012). Asp146, as discussed above, is also involved in Ap4A synthesis and is highly conserved in all but one α_2GRS homologs identified in our study, namely that in an uncultured bacterium (gi_406914003) which has a natural D200N mutation. The connection of Asp146 to both disease and Ap4A synthesis strongly supports a probable underlying relation between CMT-mutations in α_2GRS and Ap4A synthesis and homeostasis by α_2GRS. The other two CMT-associated mutations on the ID1 region, human S265F (Lee et al., 2012) and mice C201R (Achilli et al., 2009) (equivalent to Ser211 and Cys157 in PDB identifier 2ZT5, respectively) are not conserved among homologous α_2GRS. Also, ID1 is partially covered by one of the eight hotspot regions that undergo significant structural conformational change due to the *GARS* mutations and have been suggested to mediate neomorphic disease-associated pathological interactions (He et al., 2011).

Further, in order to analyze the mode of Ap4A binding in different zinc-finger proteins, a PDB scan was done to retrieve structures that had a bound Ap4A molecule as a ligand. All these structures were visualized in PyMOL to compare the Ap4A binding. In one of the structures, viz. that of adenylate kinase (AK), an Ap4A molecule was present as an inhibitor at the active site (PDB identifier 2C9Y_A). AK are studied as models for protein conformational changes and domain movement exhibited by its lid domain which also, like the ID1, adopts a rubredoxin-like zinc ribbon fold (Hayward, 2004; Krishna et al., 2003). The zinc ribbon of the lid domain has two conserved arginine residues at its hinge region that mediate electrostatic interactions with the phosphate groups of the substrate ATP molecule, thus orienting it properly in the active site (Berry and Phillips, 1998; Gerstein et al., 1993; Mukhopadhyay et al., 2011) (Fig. 4). Although the mode of interaction with Ap4A is different in the ID1 of α_2GRS and the lid domain of AK, none-the-less they exemplify the flexibility exhibited by the appended zinc ribbon domains which assists in the overall function of the respective proteins. ID1 interacts with the second adenine base of the Ap4A molecule and the phosphate adjacent to it via a single residue from the non-conserved helical insertion and two residues from the base of the N-terminal β-hairpin, respectively (Guo et al., 2009) (Fig. 4). The lid domain of AK interacts with the adenine base and the phosphate group of the Ap4A molecule via a relatively extensive surface in contrast to ID1, involving residues from the second β-strand of the N-terminal β-hairpin, the loop joining the N-terminal β-hairpin to the C-terminal β-hairpin and a single Arg residue from the additional C-terminal β-strand (Fig. 4). The interaction of the zinc ribbon with Ap4A of these two proteins effectively illustrates how the same ligand and domain pair could interact in very different ways.

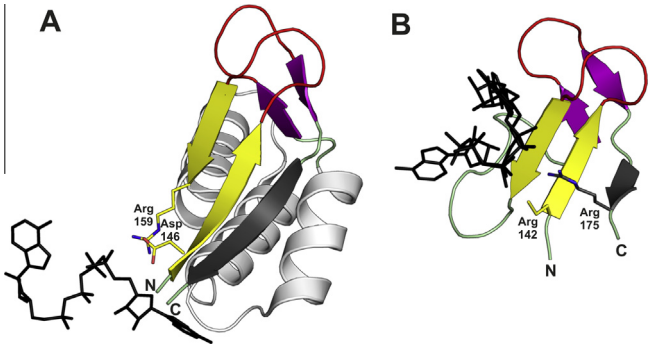

Fig.4. Zinc ribbons interacting with Ap4A. (A) Zinc ribbon from lid domain of AK (PDB identifier 2C9Y_A). (B) Zinc ribbon from ID1 of α_2GRS (PDB identifier 2ZT5_A). Coloring scheme follows that used for Fig. 1. Ap4A ligand is shown in stick representation and colored in black. The side chains of the conserved residues proposed to be involved in interactions with phosphate groups of Ap4A are shown in stick representation and labeled.

The full length ID1 (PDB identifier 2ZT5_A) does not find significant sequence similarity to any other protein domains other than α_2GRS in sequence similarity searches initiated using tools such as FFAS (Jaroszewski et al., 2005), HMMER (Finn et al., 2011) and PSI-BLAST (Altschul et al., 1997), performed against the PDB and/or NR databases. An HHpred search (Soding et al., 2005), however, could relate ID1 to an uncharacterized protein PF0385 from *Pyrococcus furiosus* (PDB identifier 2K5C_A) with an *E*-value = 0.0042 and sequence identity of 19%. The structure of PF0385 consists of a zinc ribbon domain with three helices inserted in between the zinc-binding half-sites of the zinc ribbon domain, similar to the tertiary structure of ID1 of human α_2GRS. We were able to manually superimpose (see Section 2 for details) the structure of PF0385 (PDB identifier 2K5C_A, total 95 residue) over the human ID1 (PDB identifier 2ZT5_A, total 85 residues) with an RMSD of 2.0 Å over 29 structurally-equivalent backbone C_α atoms, most of which are from the core of the zinc ribbon. The structural alignment can be further extended to include residues from the α-helical insertion, with an RMSD of 3.43 Å over 55 backbone C_α atoms. PF0385 has a zinc-ion bound in its structure, and its zinc-chelating half-sites superimpose very well with the zinc knuckles of ID1. The sequential positions of the cysteines that coordinate the zinc ion in PF0385 recapitulate the arrangement of cysteine pairs found in α_2GRS homologs. PF0385 had been previously used as a CASP8 target (CASP identifier T0476) and interestingly, the template for this protein was the ID1 from human α_2GRS (PDB identifier 2Q5H_A) (Ben-David et al., 2009), providing additional support for a zinc ribbon origin of ID1. We also performed a CD search with ID1 sequences clustered at 80% sequence identity, to infer homology of the inserted regions in between the zinc-binding sites of ID1 with any known protein domains in the CDD (Marchler-Bauer et al., 2011). The CD search revealed matches of the full length ID1 sequences only to regions annotated to be present in the GRS catalytic domain. The insertions in between the two zinc-binding half-sites could not be linked to any other CDs but the zinc ribbon region of some ID1 sequences could find matches to several other zinc ribbon domains. This correlates with our sequence analysis based results that the conserved core of ID1 comprises of only a zinc ribbon fold and the insertions in between the two zinc-binding half-sites do not form a part of any other evolutionarily conserved domain.

4. Conclusion

Our analysis reveals a zinc ribbon origin for the ID1 of α_2GRS. This zinc finger domain forms the structural core of ID1 with variable-sized non-conserved insertions in between the two zinc-binding half sites. The inserted regions in the zinc ribbon likely provide accessory functions and do not affect the stability or expression of the enzyme as observed previously for yeast GRS (Wu et al., 2013), although ID1 as a whole is required for the various functions performed by the enzyme (Qin et al., 2014). ID1 from archaea possess cysteines at the zinc knuckles which would be capable of metal-chelation, but the zinc-chelating cysteines are mostly absent in eukaryotes. Bacterial species show an uneven pattern of the conservation of metal-chelating residues across different lineages. Some eukaryotes from the Stramenopiles lineage appear to have acquired additional α_2GRS via an HGT from bacteria. The zinc ribbon domains in different proteins have been shown to interact with DNA, RNA, proteins and other small molecules (Krishna et al., 2003; Krishna et al., 2010). Two conserved residues from the zinc ribbon of ID1 play an important role in binding the phosphate group of the second ATP molecule during Ap4A synthesis (Guo et al., 2009) and interestingly, one of them (human Asp146) is also mutated in dHMN-V (Lee et al., 2012) suggesting a possible link between Ap4A regulation and hereditary motor neuropathies.

Acknowledgments

This work was supported by intramural funds (Project number OLP_0072) from the Council of Scientific and Industrial Research (CSIR) – Institute of Microbial Technology, Chandigarh, India. G.K. is supported by the Shyama Prasad Mukherjee Fellowship of CSIR, India.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2015.02.004.

References

Achilli, F., Bros-Facer, V., Williams, H.P., Banks, G.T., AlQatari, M., Chia, R., Tucci, V., Groves, M., Nickols, C.D., Seburn, K.L., Kendall, R., Cader, M.Z., Talbot, K., van Minnen, J., Burgess, R.W., Brandner, S., Martin, J.E., Koltzenburg, M., Greensmith, L., Nolan, P.M., Fisher, E.M., 2009. An ENU-induced mutation in mouse glycyl-tRNA synthetase (GARS) causes peripheral sensory and motor phenotypes creating a model of Charcot-Marie-Tooth type 2D peripheral neuropathy. Dis. Models Mech. 2, 359–373.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Antonellis, A., Ellsworth, R.E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S.Q., Jordanova, A., Kremensky, I., Christodoulou, K., Middleton, L.T., Sivakumar, K., Ionasescu, V., Funalot, B., Vance, J.M., Goldfarb, L.G., Fischbeck, K.H., Green, E.D., 2003. Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. Am. J. Hum. Genet. 72, 1293–1299.

Arnez, J.G., Dock-Bregeon, A.C., Moras, D., 1999. Glycyl-tRNA synthetase uses a negatively charged pit for specific recognition and activation of glycine. J. Mol. Biol. 286, 1449–1459.

Ben-David, M., Noivirt-Brik, O., Paz, A., Prilusky, J., Sussman, J.L., Levy, Y., 2009. Assessment of CASP8 structure predictions for template free targets. Proteins 77 (Suppl. 9), 50–65.

Berry, M.B., Phillips Jr., G.N., 1998. Crystal structures of *Bacillus stearothermophilus* adenylate kinase with bound Ap5A, Mg^{2+} Ap5A, and Mn^{2+} Ap5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg^{2+} and Mn^{2+}. Proteins 32, 276–288.

Cader, M.Z., Ren, J., James, P.A., Bird, L.E., Talbot, K., Stammers, D.K., 2007. Crystal structure of human wildtype and S581L-mutant glycyl-tRNA synthetase, an enzyme underlying distal spinal muscular atrophy. FEBS Lett. 581, 2959–2964.

Carrodeguas, J.A., Theis, K., Bogenhagen, D.F., Kisker, C., 2001. Crystal structure and deletion analysis show that the accessory subunit of mammalian DNA polymerase gamma, Pol gamma B, functions as a homodimer. Mol. Cell 7, 43–54.

Cusack, S., 1995. Eleven down and nine to go. Nat. Struct. Biol. 2, 824–831.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., Moras, D., 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347, 203–206.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, 18.

Gerstein, M., Schulz, G., Chothia, C., 1993. Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers. J. Mol. Biol. 229, 494–501.

Goerlich, O., Foeckler, R., Holler, E., 1982. Mechanism of synthesis of adenosine(5′)tetraphospho(5′)adenosine (AppppA) by aminoacyl-tRNA synthetases. Eur. J. Biochem./FEBS 126, 135–142.

Griffin, L.B., Sakaguchi, R., McGuigan, D., Gonzalez, M.A., Searby, C., Zuchner, S., Hou, Y.M., Antonellis, A., 2014. Impaired function is a common feature of neuropathy-associated glycyl-tRNA synthetase mutations. Hum. Mutat. 35, 1363–1371.

Guo, M., Yang, X.L., Schimmel, P., 2010. New functions of aminoacyl-tRNA synthetases beyond translation. Nat. Rev. Mol. Cell Biol. 11, 668–674.

Guo, R.T., Chong, Y.E., Guo, M., Yang, X.L., 2009. Crystal structures and biochemical analyses suggest a unique mechanism and role for human glycyl-tRNA synthetase in Ap4A homeostasis. J. Biol. Chem. 284, 28968–28976.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids Symp. Ser. 41, 95–98.

Hayward, S., 2004. Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. J. Mol. Biol. 339, 1001–1021.

He, W., Zhang, H.M., Chong, Y.E., Guo, M., Marshall, A.G., Yang, X.L., 2011. Dispersed disease-causing neomorphic mutations on a single protein promote the same localized conformational opening. Proc. Natl. Acad. Sci. U.S.A. 108, 12307–12312.

Holm, L., Sander, C., 1995. Dali: a network tool for protein structure comparison. Trends Biochem. Sci. 20, 478–480.

Ibba, M., Francklyn, C., Cusack, S., 2005. The Aminoacyl-tRNA Synthetases. Landes Bioscience.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., Godzik, A., 2005. FFAS03: a server for profile–profile sequence alignments. Nucleic Acids Res. 33, W284–W288.

Kisselev, L.L., Justesen, J., Wolfson, A.D., Frolova, L.Y., 1998. Diadenosine oligophosphates (ApnA), a novel class of signalling molecules? FEBS Lett. 427, 157–163.

Krishna, S.S., Majumdar, I., Grishin, N.V., 2003. Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. 31, 532–550.

Krishna, S.S., Aravind, L., Bakolitsa, C., Caruthers, J., Carlton, D., Miller, M.D., Abdubek, P., Astakhova, T., Axelrod, H.L., Chiu, H.J., Clayton, T., Deller, M.C., Duan, L., Feuerhelm, J., Grant, J.C., Han, G.W., Jaroszewski, L., Jin, K.K., Klock, H.E., Knuth, M.W., Kumar, A., Marciano, D., McMullan, D., Morse, A.T., Nigoghossian, E., Okach, L., Reyes, R., Rife, C.L., van den Bedem, H., Weekes, D., Xu, Q., Hodgson, K.O., Wooley, J., Elsliger, M.A., Deacon, A.M., Godzik, A., Lesley, S.A., Wilson, I.A., 2010. The structure of SSO2064, the first representative of Pfam family PF01796, reveals a novel two-domain zinc-ribbon OB-fold architecture with a potential acyl-CoA-binding role. Acta Crystallogr. F 66, 1160–1166.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Lee, H.J., Park, J., Nakhro, K., Park, J.M., Hur, Y.M., Choi, B.O., Chung, K.W., 2012. Two novel mutations of *GARS* in Korean families with distal hereditary motor neuropathy type V. J. Peripher. Nerv. Syst.: JPNS 17, 418–421.

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.

Logan, D.T., Mazauric, M.H., Kern, D., Moras, D., 1995. Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. EMBO J. 14, 4156–4167.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C., Bryant, S.H., 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39, D225–D229.

Mazauric, M.H., Keith, G., Logan, D., Kreutzer, R., Giege, R., Kern, D., 1998. Glycyl-tRNA synthetase from *Thermus thermophilus* – wide structural divergence with other prokaryotic glycyl-tRNA synthetases and functional inter-relation with prokaryotic and eukaryotic glycylation systems. Eur. J. Biochem./FEBS 251, 744–757.

Miras-Portugal, M.T., Gualix, J., Pintor, J., 1998. The neurotransmitter role of diadenosine polyphosphates. FEBS Lett. 430, 78–82.

Motley, W.W., Talbot, K., Fischbeck, K.H., 2010. *GARS* axonopathy: not every neuron's cup of tRNA. Trends Neurosci. 33, 59–66.

Mukhopadhyay, A., Kladova, A.V., Bursakov, S.A., Gavel, O.Y., Calvete, J.J., Shnyrov, V.L., Moura, I., Moura, J.J., Romao, M.J., Trincao, J., 2011. Crystal structure of the zinc-, cobalt-, and iron-containing adenylate kinase from *Desulfovibrio gigas*: a novel metal-containing adenylate kinase from Gram-negative bacteria. J. Biol. Inorg. Chem.: JBIC 16, 51–61.

Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536–540.

O'Donoghue, P., Luthey-Schulten, Z., 2003. On the evolution of structure in aminoacyl-tRNA synthetases. Microbiol. Mol. Biol. Rev.: MMBR 67, 550–573.
Pandit, S.B., Skolnick, J., 2008. Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. BMC Bioinformatics 9, 531.
Park, S.G., Ewalt, K.L., Kim, S., 2005. Functional expansion of aminoacyl-tRNA synthetases and their interacting factors: new perspectives on housekeepers. Trends Biochem. Sci. 30, 569–574.
Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. Nucleic Acids Res. 40, 29.
Qin, X., Hao, Z., Tian, Q., Zhang, Z., Zhou, C., Xie, W., 2014. Cocrystal structures of glycyl-tRNA synthetase in complex with tRNA suggest multiple conformational states in glycylation. J. Biol. Chem. 289, 20359–20369.
Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., Yao, X., 2009. DOG 1.0: illustrator of protein domain structures. Cell Res. 19, 271–273.
Shiba, K., 2005. Glycyl-tRNA synthetases. In: The Aminoacyl-tRNA Synthetases. Landess Bioscience, Georgetown, Texas, USA, pp. 125–134.
Sillitoe, I., Cuff, A.L., Dessailly, B.H., Dawson, N.L., Furnham, N., Lee, D., Lees, J.G., Lewis, T.E., Studer, R.A., Rentzsch, R., Yeats, C., Thornton, J.M., Orengo, C.A., 2013. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res. 41, D490–D498.
Soding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248.
Söll, D., Schimmel, P.R., 1974. Aminoacyl-tRNA synthetases. In: Paul, D.B. (Ed.), The Enzymes. Academic Press, pp. 489–538.
Veith, T., Martin, R., Wurm, J.P., Weis, B.L., Duchardt-Ferner, E., Safferthal, C., Hennig, R., Mirus, O., Bohnsack, M.T., Wohnert, J., Schleiff, E., 2012. Structural and functional analysis of the archaeal endonuclease Nob1. Nucleic Acids Res. 40, 3259–3274.
Woese, C.R., Olsen, G.J., Ibba, M., Soll, D., 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev.: MMBR 64, 202–236.
Wolf, Y.I., Koonin, E.V., 2001. Origin of an animal mitochondrial DNA polymerase subunit via lineage-specific acquisition of a glycyl-tRNA synthetase from bacteria of the *Thermus-Deinococcus* group. Trends Genet.: TIG 17, 431–433.
Wu, Y.H., Chang, C.P., Chien, C.I., Tseng, Y.K., Wang, C.C., 2013. An insertion peptide in yeast glycyl-tRNA synthetase facilitates both productive docking and catalysis of cognate tRNAs. Mol. Cell. Biol. 33, 3515–3523.
Yao, P., Fox, P.L., 2013. Aminoacyl-tRNA synthetases in medicine and disease. EMBO Mol. Med. 5, 332–343.
Zhang, Y., Skolnick, J., 2005. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 33, 2302–2309.
Zhou, X.L., Zhu, B., Wang, E.D., 2008. The CP2 domain of leucyl-tRNA synthetase is crucial for amino acid activation and post-transfer editing. J. Biol. Chem. 283, 36608–36616.