Received: 12 June 2023 | Revised: 14 October 2023 | Accepted: 24 October 2023

DOI: 10.1002/prot.26632

RESEARCH ARTICLE

PROTEINS WILEY

Identification of sequence determinants for the ABHD14 enzymes

Kaveri Vaidya[1] | Golding Rodrigues[1] | Sonali Gupta[1] | Archit Devarajan[2] | Mihika Yeolekar[1] | M. S. Madhusudhan[1] | Siddhesh S. Kamat[1]

[1]Department of Biology, Indian Institute of Science Education and Research Pune, Pune, Maharashtra, India

[2]Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, Madhya Pradesh, India

Correspondence
Siddhesh S. Kamat and M. S. Madhusudhan, Department of Biology, Indian Institute of Science Education and Research Pune, Dr. Homi Bhabha Road, Pashan, Pune 411008, Maharashtra, India.
Email: siddhesh@iiserpune.ac.in and madhusudhan@iiserpune.ac.in

Funding information
Department of Biotechnology, Ministry of Science and Technology, India; European Molecular Biology Organization

Abstract

Over the course of evolution, enzymes have developed remarkable functional diversity in catalyzing important chemical reactions across various organisms, and understanding how new enzyme functions might have evolved remains an important question in modern enzymology. To systematically annotate functions, based on their protein sequences and available biochemical studies, enzymes with similar catalytic mechanisms have been clustered together into an enzyme superfamily. Typically, enzymes within a superfamily have similar overall three-dimensional structures, conserved catalytic residues, but large variations in substrate recognition sites and residues to accommodate the diverse biochemical reactions that are catalyzed within the superfamily. The serine hydrolases are an excellent example of such an enzyme superfamily. Based on known enzymatic activities and protein sequences, they are split almost equally into the serine proteases and metabolic serine hydrolases. Within the metabolic serine hydrolases, there are two outlying members, ABHD14A and ABHD14B, that have high sequence similarity, but their biological functions remained cryptic till recently. While ABHD14A still lacks any functional annotation to date, we recently showed that ABHD14B functions as a lysine deacetylase in mammals. Given their high sequence similarity, automated databases often wrongly assign ABHD14A and ABHD14B as the same enzyme, and therefore, annotating functions to them in various organisms has been problematic. In this article, we present a bioinformatics study coupled with biochemical experiments, which identifies key sequence determinants for both ABHD14A and ABHD14B, and enable better classification for them. In addition, we map these enzymes on an evolutionary timescale and provide a much-wanted resource for studying these interesting enzymes in different organisms.

KEYWORDS
ABHD14A, ABHD14B, bioinformatics, lysine deacetylase, phylogenetic analysis, serine hydrolase, structural analysis

Kaveri Vaidya and Golding Rodrigues contributed equally to this study.

1 | INTRODUCTION

Enzymes are remarkable biological catalysts that greatly facilitate rates of biological reactions,[1,2] ranging from simple hydrolytic reactions to highly complex radical transformations across diverse organisms. By regulating the rates of biological reactions, enzymes serve as critical metabolic lynchpins in numerous physiological processes. While the repertoire of biochemical reactions catalyzed by enzymes is quite vast and continues to grow steeply, interestingly enough, in comparison, the three-dimensional protein scaffolds on which enzyme active sites are built remain limited.[3–6] Given this fact, it has been hypothesized that over millions of years of evolution, as the number of catalytic reactions expanded with the complexity in multicellular organisms, correspondingly most enzymes also evolved most likely from a common ancestral protein.[7–10] Such an ancestral enzyme presumably having a defined three-dimensional scaffold, during evolution, allowed the development of diverse active sites within its architecture to catalyze emerging chemical reactions.[7–10] Over the past three decades, there has been an explosion in the publicly available genome sequences and three-dimensional protein structures from different organisms. This information coupled with high-throughput biochemical studies toward annotating function to enzymes has identified several homologous enzymes from diverse organisms that catalyze different catalytic reactions, but have similar biochemical mechanisms and/or aspects of catalysis. Such related enzymes have since been grouped together in what is now popularly known as an "Enzyme Superfamily".[4–7,11]

Given the tremendous efforts of numerous research labs, thanks largely to the formation of several global consortia, to date, distinct enzyme superfamilies have been clearly defined. Collective efforts of these consortia have mapped unique biological functions to a significant number of enzymes for a majority of these enzyme superfamilies. Among such enzyme superfamilies lies the *Serine Hydrolase* enzyme superfamily, which is split almost equally into two distinct classes: serine proteases and metabolic serine hydrolases (mSHs).[12–14] The mSHs class within this superfamily comprises ~125 members in humans.[12] The mSHs catalyze an array of hydrolytic reactions with a consensus ping-pong catalytic mechanism, and dysregulation in activities is associated with various pathological conditions and diseases. In humans, these diseases range from genetic diseases (e.g., autoimmune conditions and neurological disorders) to metabolic disorders (e.g., dyslipidemia and diabetes) to cancers.[12,13] Of note, most members of the mSH family possess the canonical α/β-hydrolase domain (ABHD)-fold.[15–18] Interestingly, while classifying the mSHs, as several members did not have any known function assigned to it at that time, they were monikered as ABHDn (n = number) protein.

Based on a comparative sequence phylogenetic analysis,[12,19] at the periphery of the mSH class lie two closely related enzymes, ABHD14A and ABHD14B. Based on their protein sequence, both these enzymes are hypothesized to possess the ABHD-fold, and a conserved catalytic triad (Ser–His–Asp). As compared to other mSHs, the invariant nucleophilic active site serine residue falls within a non-canonical SxSxS motif (canonical motif: GxSxG) (x = any amino acid) for both these enzymes.[12,19] This distinction makes both ABHD14A and ABHD14B outlying members of the mSH family. Interestingly, unlike any other mSH, both ABHD14A and ABHD14B were discovered from high-throughput screens looking to identify protein interactors of transcription factors. ABHD14A (also known as Dorz1) was found to be a genetic interactor of the transcription factor Zic1, and in mammals, ABHD14A is predicted to have a role in embryonic development of the cerebellum.[20] On the other hand, a yeast two-hybrid screen identified ABHD14B (also known as CCG1-interacting protein B) as a protein interactor of the histone acetyl-transferase domain of the general transcription factor TFIID, and was tentatively assigned a role in transcriptional regulation.[21] The same study also reported the three-dimensional structure of human ABHD14B, and showed that this enzyme possessed the ABHD-fold.[21] It is important to note that both these enzymes were discovered over two decades ago, and yet no biochemical function has been assigned to either one of them, till recently.

In an effort to functionally annotate these enzymes, we initiated studies on human ABHD14B. Through a series of biochemical and cellular assays, we found that ABHD14B functions as a novel lysine deacetylase (KDAC).[22] We showed that this enzyme transfers an acetyl-group from a post-translationally modified protein lysine residue to a molecule of Coenzyme-A (CoA) to make acetyl-CoA.[22] In a follow-up study, we found that depleting ABHD14B in mammalian cells and liver of mice results in altered (systemic) glucose metabolism and cellular energetics.[23] While the exact physiological substrates and mechanisms of action for this altered metabolic phenotype remain cryptic, our studies have successfully established a catalytic mechanism for ABHD14B. On the other hand, ABHD14A still remains elusive with regards to any biochemical function.

Given their relatively high sequence similarity, we found while studying these enzymes that across various protein databases, despite being distinct proteins, ABHD14A and ABHD14B were often misassigned as one another. Therefore, to understand what are minimal sequence determinants for categorizing these enzymes, in this article, we perform a thorough bioinformatics study on both ABHD14A and ABHD14B. Next, we couple this in silico analysis with a few biochemical studies, and present a resource for assigning a given protein sequence as either ABHD14A or ABHD14B. Further, we assess the presence of both these enzymes on the evolutionary time scale, and identify protein sequences in various organisms that correspond to either ABHD14A and ABHD14B. Our studies thus pave the way for a better classification of ABHD14A and ABHD14B in an effort toward assigning physiologically relevant functions to them in different organisms in the coming years.

2 | MATERIALS AND METHODS

2.1 | Bioinformatics searching and analysis

To determine the prevalence of ABHD14A and ABHD14B protein sequences (and its homologues) across all organisms, PSI-BLAST

searches[24] were carried out on reference sequences of human ABHD14A (RefSeq: NP_056222.2, Uniprot: Q9BUJ0) and human ABHD14B (RefSeq: NP_001139786.1, Uniprot: Q96IU4), for three iterations, against standard nonredundant databases with an expect threshold of 0.00005 and a maximum return of 5000 hits. To assess the presence or absence of any transmembrane domain(s), all protein sequences from the initial search were submitted to two different transmembrane region prediction software: CCTOP[25] and TMHMM.[26] For additional curation of the data, a home-made script was written to verify the presence of the conserved catalytic triad conserved within the searched motifs across the length of the sequence, by carrying out pairwise global alignments using the Needleman–Wunsh algorithm.[27] For the final resulting of the datasets, a single protein sequence was taken from each representative organism, and subjected to a multiple sequence global alignment using MAFFT.[28] The resulting alignments were further assessed in MEGA-X[29] to build maximum-likelihood trees, and the trees were visualized and annotated in iTol.[30] The fully conserved residues for both ABHD14A and ABHD14B were identified using pymsaviz (pypi.org/project/pymsaviz/) and Bio.Align (biopython.org/docs/1.76/api/Bio.Align.html) packages in Python. The functionally conserved residues for both enzymes were identified by manual inspection of filtered protein sequences.

2.2 | Cloning, expression, and purification of ABHD14B mutants

The human *abhd14b* gene[22] was synthesized as a codon-optimized construct (from Genscript) for expression in *Escherichia coli* and, further subcloned into the pET-45b(+) vector (Millipore) with a N-terminal $(His)_6$-tag. For generating single-point mutants, site-directed mutagenesis was performed using Pfu DNA polymerase as per manufacturer's instructions (Promega, Cat. # M7741). The resulting sequence verified plasmids having the desired mutation were individually transformed into *E. coli* BL21(DE3) competent cells. A single colony was inoculated into 5 mL of Luria Bertani medium containing ampicillin (final concentration in media was 100 μg/mL) and grown overnight (12–14 h) in an incubator with a temperature at 37°C with constant shaking. This primary inoculum was transferred to 1 L of the same medium, and grown in an incubator with a temperature at 37°C with constant shaking until the absorbance at 600 nm reached ~0.6. At this time point, the temperature of the incubator was reduced to 18°C, and ABHD14B expression was induced by adding 500 μM of isopropyl-β-D-1-thiogalactopyranoside to the culture. Upon induction of protein expression, the bacterial culture was grown for an additional 16–18 h at 18°C. The resulting cells were pelleted by centrifugation at 6000*g* for 20 min and stored at −80°C until further use. The ABHD14B overexpression was assessed by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis and only upon confirming protein overexpression, the resulting cell pellets were used for further processing.

The ABHD14B mutants were purified by Ni-NTA (Nickel nitriloacetic acid) affinity chromatography using a His-Trap high-performance column (GE Heathcare, Cat #17-5248-02) as per manufacturer's instructions. Briefly, a cell pellet (with confirmed overexpression of the ABHD14B mutant) was thawed and re-suspended in 45 mL of binding buffer (50 mM Tris, 10 mM imidazole at pH 8) at 4°C. Cells were lyzed using a Vibra Cell VCX 130 probe sonicator (Sonics) at 60% amplitude for 30 min with 2-s "ON" and 3-s "OFF" cycles on ice. The resulting homogenate was centrifuged for 30 min at 30000*g* at 4°C to separate the soluble and insoluble fractions, and the soluble fraction (that contains ABHD14B mutants) was applied to a pre-equilibrated His-Trap high-performance column. The desired ABHD14B mutant was eluted using an increasing imidazole gradient (50–500 mM) to the binding buffer, and the purity of the various collected fractions of the eluted ABHD14B mutant was confirmed by SDS-PAGE analysis. The fractions containing the ABHD14B mutant (>95% pure) were pooled, concentrated in 50-mM Tris (pH 8) using a 10-kDa molecular weight cutoff filter (Millipore), and ~4 mL of this concentrated ABHD14B mutant was loaded onto a pre-equilibrated HiLoad™16/60 Superdex200 (GE Healthcare) gel filtration column. The fractions collected from this size exclusion chromatography column were assessed for the purity of the desired ABHD14B mutant by SDS-PAGE analysis. The fractions containing the desired pure ABHD14B mutant (>95%) were pooled, concentrated as described earlier, aliquoted (10 μL aliquots), flash-frozen using liquid nitrogen, and subsequently stored at −80°C till further use.

2.3 | Gel-based activity-based protein profiling experiments

All gel-based activity-based protein profiling (ABPP) assays were performed as reported earlier.[22,31,32] In these assays, the total reaction volume was 50 μL, while the final protein and activity probe (fluorophosphonate-rhodamine, FP-rhodamine) concentrations were 5 μM each. All samples were resolved on a 12.5% SDS-PAGE gel, and the activities of the mutants were visualized using a iBright1500 (Invitrogen) gel documentation system.

2.4 | Colorimetric substrate hydrolysis assays

All colorimetric substrate hydrolysis assays with *p*-nitrophenyl-acetate (pNP-acetate) were performed as reported earlier[22] in 50 mM Tris at pH 8 in a final reaction volume of 250 μL. For measuring the ABHD14B-dependent pNP-acetate turnover, a final concentration of 10 and 100 μM for the ABHD14B mutant and pNP-acetate, respectively, were used in the assay. For assessing the acetyl-transferase reaction catalyzed by ABHD14B using pNP-acetate and CoA, a final concentration of 10 μM, 100 μM, and 1 mM for the ABHD14B mutant, pNP-acetate, and CoA, respectively, was used in the assays. For assessing the relative activities of the ABHD14B mutants, 500 s

time-point was chosen, as this colorimetric assay was found to be linear with a good signal-to-noise ratio at this time point. All assays were done in biological triplicates to ensure reproducibility.

2.5 | Data plotting

All data presented in this article were plotted using the GraphPad Prism 9 (version 9.5.1) software for MacOS-X. Unless otherwise mentioned, all data are presented as mean ± standard deviation for at least three independent experiments (biological replicates).

3 | RESULTS AND DISCUSSION

3.1 | Identification of ABHD14B sequences

The three-dimensional structure of human ABHD14B has shown that this mSH has an ABHD-fold, and possess a conserved catalytic triad, consisting of Ser-111, Asp-162, and His-188 (Figure 1).[21] In addition to this, a bioinformatics survey of various ABHD-proteins from the mSH family have identified two conserved sequence motifs for (mammalian) ABHD14Bs, namely the nucleophilic motif (consisting of a SxSxS motif within the VVISPSLSGMY sequence, x = any amino acid) and an acyltransferase motif (consisting of the HxxxxD motif within the GAGHPCYLDKPE sequence, x = any amino acid) toward the C-terminal end of the protein (Figure 1).[15]

From the 5000 hits obtained from using human ABHD14B (RefSeq: NP_001139786.1, Uniprot: Q96IU4) as a query sequence for search, the ABHD14B nucleophilic motif was identified in 1012 sequences (allowing for up to two mismatches), of which 792 sequences were perfect matches. Among the 5000 hits, the ABHD14B acyltransferase motif was identified in 1265 sequences (allowing for up to two mismatches), with 217 sequences being perfect matches. Of note, the catalytic triad was identified in 2929 sequences. Based on the above information, in our study, a protein sequence was classified as ABHD14B, if it possessed the catalytic triad, the ABHD14B nucleophile motif (allowing for up to two mismatches), and the ABHD14B acyltransferase motif (allowing for up to two mismatches). Based on these filtering criteria, overall, we identified 847 ABHD14B sequences identified from 697 organisms, of which 197 ABHD14B sequences identified from 130 organisms were perfect matches to the aforementioned filtering criteria (Supplementary Table 1).

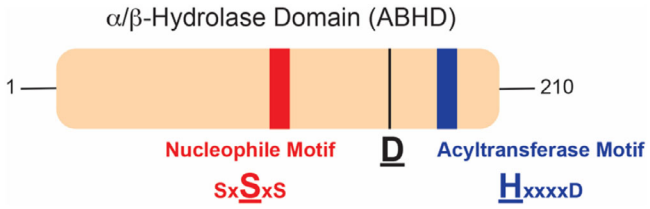


FIGURE 1 Schematic representation of the human ABHD14B structure. Based on available literature, ABHD14B is predicted to contain the catalytic triad consisting of Ser–Asp–His, a nucleophilic motif that contains the active site serine residue as part of a SxSxS sequence, and an acyltransferase motif that contains the HxxxxD sequence, all within an overall ABHD-fold.

3.2 | Phylogenetic classification of ABHD14B sequences

From the filtering of protein sequences obtained from the database searches, we identified 847 ABHD14B sequences and found that these protein sequences came from 697 organisms. Upon manual curation of these data, it was clear that a few organisms (especially from class *Mammalia* and *Aves*) possessed more than one isoform of ABHD14B. Hence, the total number of ABHD14B sequences was significantly greater than organisms identified from our search (Supplementary Table 1). Nonetheless, we chose the longest ABHD14B sequence from any particular organism to perform a phylogenetic (evolutionary) analysis for ABHD14B, and found that ABHD14B was almost exclusively confined to phylum *Chordata* (~99.7%, 695 organisms of the 697 organisms identified) (Figure 2A). Among the phylum *Chordates*, the ABHD14B sequences were most prevalent in *Aves* (Birds) (~45%, 313 organisms of the 697 organisms identified), *Mammalia* (Mammals) (~27%, 189 organisms of the 697 organisms identified), and *Actinopterygii* (Bony Fish) (~22%, 156 organisms of the 697 organisms identified), with a smaller representation seen in *Reptilia* (Reptiles) (~5%, 33 organisms of the 697 organisms identified) (Figure 2B). Based on the phylogenetic analysis, the avian ABHD14B protein sequences were most closely related to the mammalian ABHD14B sequences, while those of class *Actinopterygii* and *Reptilia* were most closely related to one another, and together had a closer sequence relation to mammalian ABHD14Bs than the avian ABHD14Bs (Figure 2A). It was interesting to note that of the 697 organisms possessing ABHD14B sequence, only 263 organisms possessed ABHD14A sequence (~38% of the total organisms) (Figure 2) (discussed in Section 3.6).

3.3 | Conservation of residues within the ABHD14B sequences

To assess the overall conservation in protein sequence across all the 697 ABHD14B sequences identified from different organisms following database searches and subsequent phylogenetic analysis, we performed a multiple sequence alignment analysis on them (Supplementary File 1). From this analysis, we found that across all the 697 ABHD14B sequences, 86 residues were highly conserved (present at a frequency of >90% in all the 697 ABHD14B sequences at the defined position), suggesting that across all organisms, the overall sequence conservation is fairly high (~40%). In addition, upon closer inspection, we found that within a particular class (e.g., *Mammalia*), the extent of sequence conservation is significantly

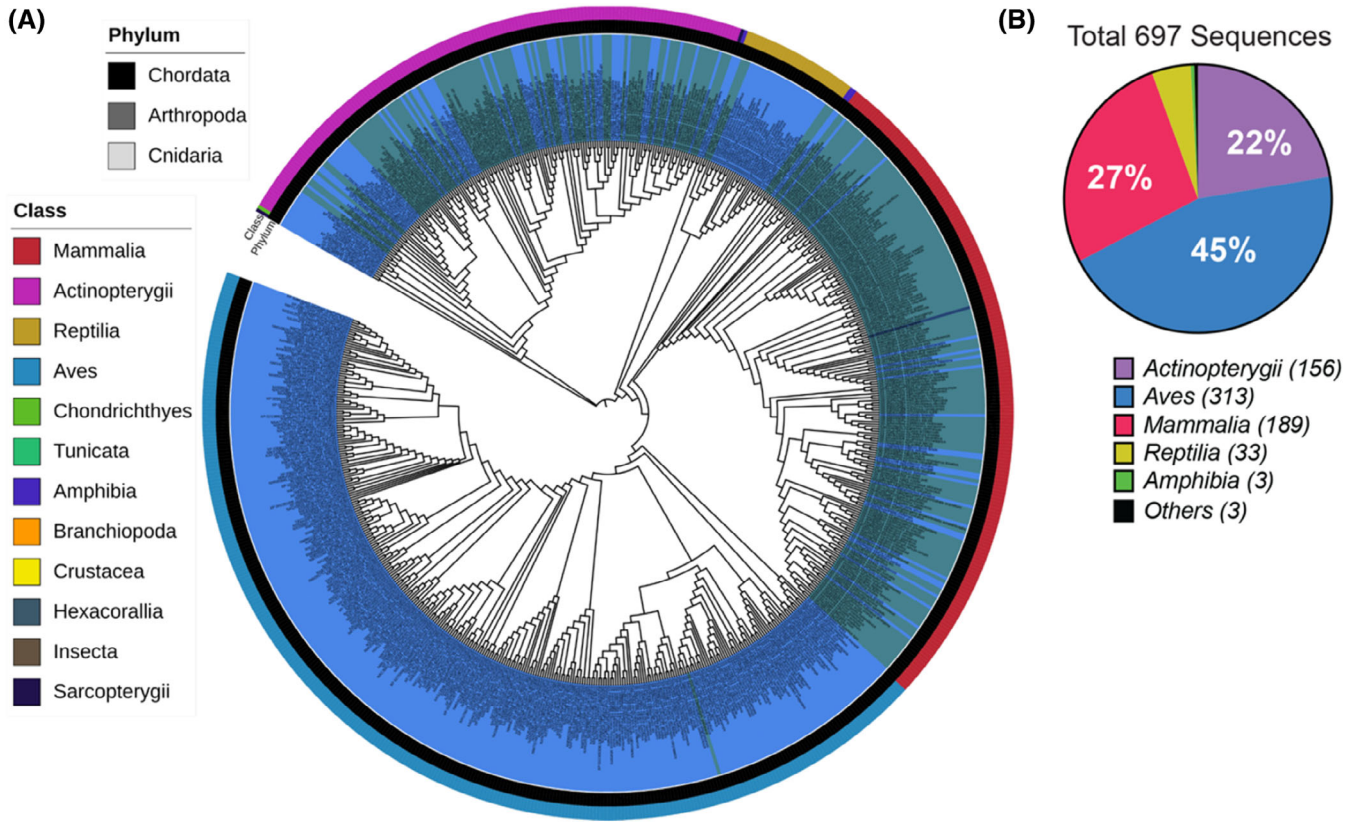
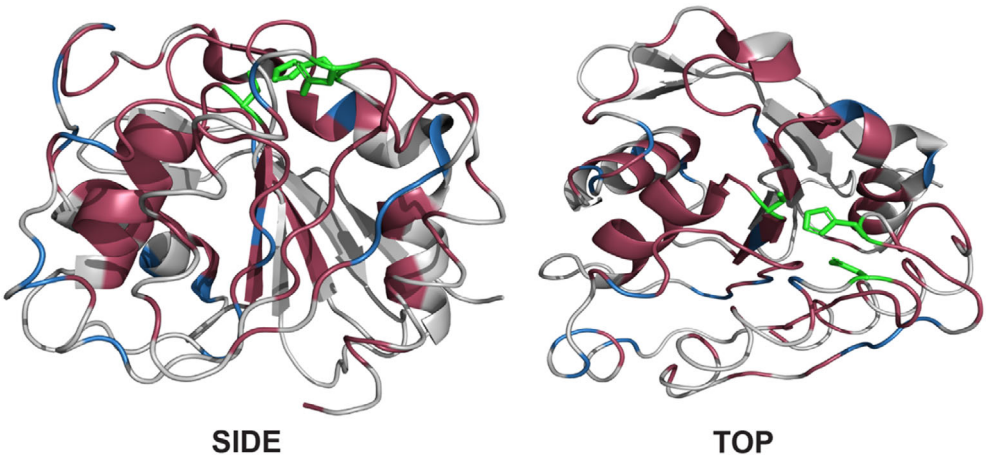


FIGURE 2 Phylogenetic analysis of ABHD14B sequences. (A) Phylogenetic tree representing the identified ABHD14B sequences from 697 organisms. The outermost and middle circular coloring denotes the Class and Phylum to which a sequence belongs (see associated legends within the figure), while the innermost circular coloring (blue or green) denotes whether an organism contains a sequence only for ABHD14B (blue) or for both ABHD14B and ABHD14A (green). (B) Pie-chart analysis representing the data from the phylogenetic tree for the various classes of organisms, mainly from Chordates that contain a ABHD14B sequence. This analysis shows that class Aves, Mammalia, and Actinopterygii contain the most sequences for ABHD14B within the Chordates.

FIGURE 3 Mapping conserved residues on the ABHD14B structure. The residues colored in red are those that are absolutely conserved, while those shown in blue are functionally conserved based on the multiple sequence alignment analysis across all the 697 sequences of ABHD14B from different organisms. The catalytic triad residues that mark the active site are shown in green.



higher (~80%) (Supplementary File 1 and Supplementary Table 1). In addition to the overall conserved residues, we found that an additional 20 residues were functionally conserved (presence of similar amino-acid type at that position; e.g., Glu and Asp are functionally conserved residues) across all the 697 ABHD14B sequences, suggesting that the realistic sequence conservation across all organisms for ABHD14B was ~50% (Supplementary File 1 and Supplementary Table 1).

Next, we mapped all the conserved residues (including functionally conserved residues) on the available three-dimensional structure of ABHD14B (PDB: 1imj), and found not surprisingly that these conserved residues were clustered around the enzyme active site and the

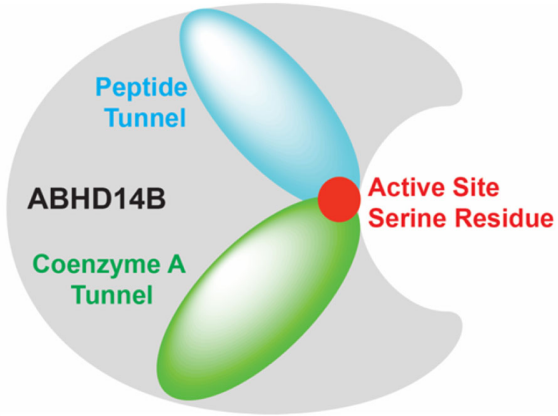


FIGURE 4 Cartoon representation of the substrate tunnels within ABHD14B. The three-dimensional structure of human ABHD14B suggests that this enzyme has two tunnels for the two substrates that converge on the active site serine residue.

cleft adjoining this active site where the putative substrates are predicted to bind (Figure 3). We also found from this structural analysis that the residues present on the flexible loops involved in substrate recognition were also highly conserved. Next, previous modeling studies by us toward understanding the mechanism of this enzyme[22] revealed that ABHD14B has two tunnels that seem to converge on the active site serine residue: (i) the putative peptide (substrate # 1) tunnel and (ii) the CoA (substrate # 2) tunnel (Figure 4). It is interesting to note that the entrance to both these tunnels has highly conserved positively charged residues presumably gating the entrance of substrates to the active site (for the human ABHD14B, Arg42, and Lys141 at the entrance of the CoA and peptide tunnel, respectively). Further, we noticed that the residues that comprise these tunnels also show a high degree of conservation. Finally, of note, we found that the CoA tunnel comprises several hydrophobic residues, presumably to exclude water from the active site, to prevent the hydrolysis of the acetyl-enzyme intermediate.[22,23]

3.4 | Biochemical characterization of ABHD14B mutants

The bioinformatics analysis helped us identify key conserved residues across the different ABHD14B sequences from various organisms. Next, using biochemical assays, we wanted to validate some of these bioinformatics findings and assess the effect of mutating some conserved residues on ABHD14B activity. Toward this, we chose to assay the recombinant human ABHD14B, as we and others have previously studied this enzyme and a three-dimensional structure is available for it.[21–23] For assaying the various human ABHD14B mutants, we chose two previously established biochemical assays, namely the gel-based ABPP assay and pNP-acetate-based colorimetric substrate hydrolysis assay.[22] Using an activity-based probe, FP-rhodamine in this case, the gel-based ABPP assay reports on the nucleophilicity of the active site serine residue, and serves as a proxy for the activity of an enzyme.[22,33] On the contrary, the pNP-acetate-based colorimetric substrate hydrolysis assay serves as a direct readout for the overall catalytic ability of the enzyme to turn over an acetylated substrate surrogate.[22]

First, we made alanine mutants for the residues from the catalytic triad, that is, S111A, D162A, and H188A, and, as expected, found from gel-based ABPP assays that these mutants were unable to react with the activity probe, and show signals on the activity gel (Figure 5A). In addition to this, relative to wild-type (WT) ABHD14B, all three mutants also showed a significant loss in their ability to turn over pNP-acetate (Figure 5B) and perform the acetyltransferase reaction (Figure 5C), with S111A having the most diminished activity in all the assays as expected.

Second, we focused on the two gate-keeping residues of the tunnels, namely R42 (CoA tunnel) and K141 (peptide tunnel), and made alanine mutants for both these residues. Upon assaying them, we found that relative to WT ABHD14B, both these mutants showed comparable activity in the gel-based ABPP assay (Figure 5A). This suggested that the nucleophilicity of the active site serine is unaffected in both mutants. Next, we found that relative to WT ABHD14B, while the R42A mutant did not show any change, mutation to K141 resulted in a significant (>50%) reduction in its ability to turn over pNP-acetate (Figure 5B). Along similar lines, we found that relative to WT ABHD14B, both R42A and K141A mutants had significantly hampered acetyltransferase activity (Figure 5C). As the K141A mutant has a diminished ability to turnover pNP-acetate, it is expected that this mutant will also have hampered acetyltransferase activity (Figure 5B,C). Interestingly, we found that while the R42A mutant had WT ABHD14B equivalent activity for the turnover of pNP-acetate, it had significantly reduced acetyltransferase activity (Figure 5B, C). These results together strongly suggest that as predicted, R42 and K141 play important roles in the proper orientation of the substrates (CoA for R42, and peptide for K141) for optimal catalysis.

Third, we wanted to assess the role that D193 plays during this catalysis, as this residue is part of the invariant HxxxxD acyltransferase motif (x = any amino acid) found in all ABHD14B sequences. Interestingly, when we assayed the D193A mutant, we found that while it had moderate pNP-acetate hydrolysis activity (Figure 5B) (~65% of WT ABHD14B), the nucleophilicity of the active site serine was significantly diminished, as evidenced by its inability to react with the activity probe in the gel-based ABPP assays (Figure 5A). Based on known mechanisms for acyltransferases, by regulating the nucleophilicity of S111, presumably due to its interactions with H188 as part of the HxxxxD acyltransferase motif, D193 might have an important role to play in the transfer of the acetyl-group from the acetyl-enzyme intermediate to CoA.[22,23,34] To test this premise, we tested this mutant in the acetyltransferase reaction assay and found that the D193A mutant had substantially reduced acetyltransferase activity (Figure 5C), thus confirming an important role for this residue in the acetyl-transfer step in the second-half of the enzyme cycle.

Fourth, we wanted to check if mutations to residues within a protein tunnel had any effect on ABHD14B activity, and we chose residues from the CoA tunnel, as previous structural modeling studies from our lab had speculated some roles for different residues within this tunnel.[22] Toward

this, we made alanine mutants for H55, R56, Y191, W198, and H199, which are conserved and predicted to the neutralize the negative charges on phosphates (H55, R56) or π-stack with and stabilize the nucleotide region of CoA (Y191, W198, H199). We found all the single-point alanine mutants of the aforementioned residues (H55A, R56A, Y191A, W198A, and H199A) had comparable activity relative to WT ABHD14B, in the gel-based ABPP assay (Figure 5A) and in their ability to turnover pNP-acetate (Figure 5B). Interestingly, we found that all these aforementioned mutants had substantially diminished acetyltransferase activity (Figure 5C), pointing to an important role for these residues in the binding and orientation of CoA to ABHD14B, and the subsequent acetyl-transfer reaction.

Fifth, and lastly, we found from the bioinformatics and multiple sequence alignment analysis that S75 (in humans) was a universally conserved residue in ABHD14B sequences from all organisms (Supplementary File 1). This was an intriguing finding, as based on the three-dimensional structure of human ABHD14B, S75 resides on a flexible loop that is present ~17 A° away from the active site, and it does not seem to have any particular predicted role in the enzyme catalytic cycle (Figure 6). However, we found that upon mutation of this residue, the S75A mutant had significantly diminished activity in all the assays (Figure 5). Several post-translational modification (PTM) prediction databases/algorithms seem to suggest that ABHD14B might be phosphorylated, and that perhaps this protein phosphorylation might have physiological implications in different biological contexts. Though only speculative yet, given the activity profile seen from the assays done with the S75A mutant, it might be possible that through its PTM, S75 might be allosterically regulating ABHD14B activity.

Overall, the information from the conserved residues has helped confirm known findings and open new avenues for residues that may be tested to probe the enzymatic mechanism of ABHD14B and regulation of its catalytic activity further.

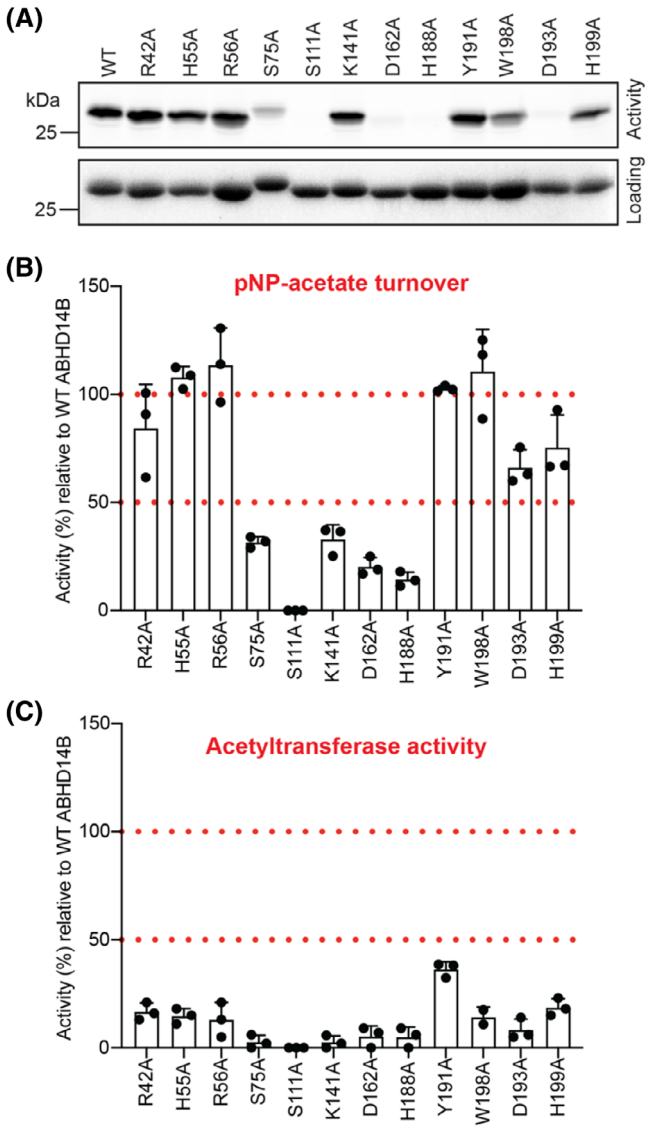


FIGURE 5 Biochemical characterization of ABHD14B mutants. (A) Representative gel showing the activity of the various mutants generated for human ABHD14B in a gel-based ABPP assay. In this assay, the protein and probe concentrations were 5 μM each. The top gel shows in-gel fluorescence, while the bottom gel shows equal protein loading by Coomassie staining. This experiment was repeated three independent times (biological replicates) with reproducible results each time. (B) *p*-nitrophenyl-acetate (pNP-acetate) substrate hydrolysis assay showing the activity of various mutants in this colorimetric assay. The bar data denotes mean ± standard deviation for the individual data points (three biological replicates), relative to wild-type (WT) ABHD14B (control). In this assay, the protein and pNP-acetate concentrations were 10 and 100 μM, respectively. (C) Acetyltransferase assay showing the activity of various mutants in this pNP-acetate-based colorimetric assay. The bar data denotes mean ± standard deviation for the individual data points (three biological replicates), relative to WT ABHD14B (control). In this assay, the protein, pNP-acetate and CoA concentrations were 10 μM, 100 μM, and 1 mM, respectively.

3.5 | Identification of ABHD14A sequences

Owing to its high sequence similarity to ABHD14B, various bioinformatics studies have predicted that the three-dimensional structure of ABHD14A also has an ABHD-fold, and possesses a conserved catalytic triad, consisting of Ser-171, Asp-222, and His-249 for the human ABHD14A sequence (Figure 7). Like ABHD14B, bioinformatics studies with other ABHD-proteins from the mSH family have identified a conserved nucleophilic motif (consisting of a SxSxS motif within the VLVSPSLSGHY sequence, x = any amino acid) (Figure 7).[15] Of note, sequence analysis suggests that ABHD14A lacks the acyltransferase motif seen in ABHD14B, and unlike ABHD14B, surprisingly, has an integral membrane domain consisting of ~30–40 amino acids that form an anchoring α-helical sequence at the N-terminal end of the protein (Figure 7).

Of the 5000 hits obtained from using human ABHD14A (RefSeq: NP_056222.2, Uniprot: Q9BUJ0) as a query sequence for search, the nucleophilic motif for ABHD14A was identified in 733 sequences (allowing for up to two mismatches), of which 134 sequences were perfect matches, while the catalytic triad was identified in 2624 sequences. From the 5000 hits obtained from the initial search, the two transmembrane prediction software, CCTOP and TMHMM,

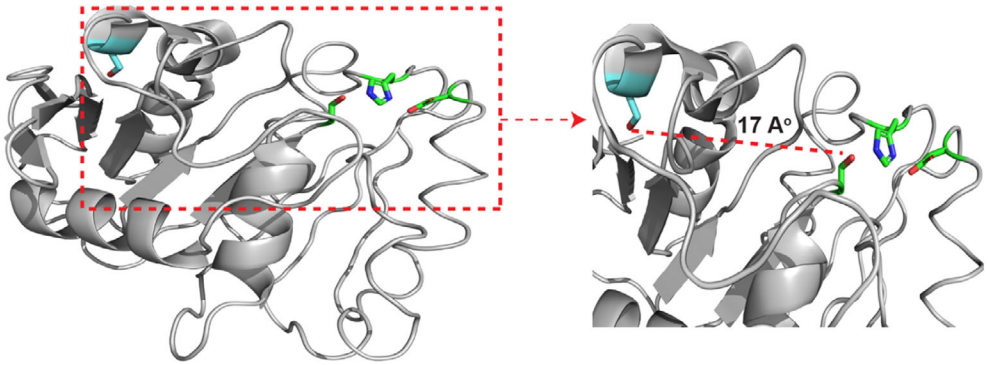


FIGURE 6 Position of S75 relative to the catalytic triad in human ABHD14B. The three-dimensional structure of human ABHD14B showing the position of S75 (shown in blue) relative to the catalytic triad (shown in green). The zoomed image of this, shows that S75 is ~17 A° away from the active site S111 residue on a flexible loop.

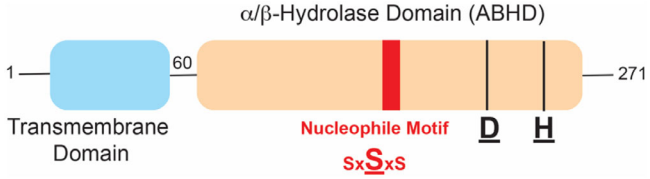


FIGURE 7 Schematic representation of the human ABHD14A structure. Based on available literature, ABHD14A is predicted to contain the catalytic triad consisting of Ser-Asp-His, and a nucleophilic motif that contains the active site serine residue as part of a SxSxS sequence within an overall ABHD-fold. In addition, ABHD14A contains a N-terminal domain that is predicted to anchor it to the cellular membranes.

predicted transmembrane regions in 1069 and 1084 sequences, respectively. Of these, a total of 1010 sequences were common to both prediction software, while 129 were identified by a single software, leading to a complete set of 1139 protein sequences with an identified transmembrane region. Based on the above information, in our study, a protein sequence was classified as ABHD14A, if it possessed the catalytic triad, the ABHD14A nucleophile motif (allowing for up to two mismatches), and had a transmembrane domain. Based on this filtering criteria, overall, we identified 483 ABHD14A sequences identified from 312 organisms, of which 116 ABHD14A sequences identified from 52 organisms were perfect matches to the aforementioned filtering criteria (Supplementary Table 2).

3.6 | Phylogenetic classification of ABHD14A sequences

Just as with ABHD14B, upon applying the aforementioned filtering criteria for the shortlisted ABHD14A sequences obtained from various databases, we identified 483 ABHD14A sequences from 312 organisms. Upon manual curation of these data, it was clear that several organisms (especially from class *Mammalia* and *Actinopterygii*) possessed more than 1 isoform of ABHD14A, and therefore, there was a disconnect between the total ABHD14A sequences and the organisms identified from our search (Supplementary Table 2). Next, we chose the longest ABHD14A sequence from any particular organism to perform a phylogenetic (evolutionary) analysis for ABHD14A, and found that ABHD14A was largely confined to phylum *Chordata* (~97.5%, 304 organisms of the 312 organisms identified), with a smaller representation from the *Arthropoda* and *Cnidaria* phyla (Figure 8A). Among the phylum *Chordates*, the ABHD14A sequences were predominantly found in class *Mammalia* (Mammals) (~58%, 182 organisms of the 312 organisms identified) and *Actinopterygii* (Bony Fish) (~37%, 114 organisms of the 312 organisms identified), with a much smaller representation seen in *Reptilia* (Reptiles) (~2%, 6 organisms of the 312 organisms identified) (Figure 8B). The phylogenetic analysis showed that within the *Chordates*, the ABHD14A protein sequences from class *Actinopterygii* and *Mammalia* were most closely related to one another, and together were related to ABHD14A sequences from the class *Reptilia* (Figure 7). Of the 312 organisms that we identified possessing a ABHD14A sequence, 263 organisms also possessed an ABHD14B sequence, an overall fraction (~84% of the total organisms) far greater than ~38% observed from the earlier ABHD14B analysis (Figure 8C and Supplementary Tables 1 and 2).

The most striking and surprising aspect of this phylogenetic analysis for ABHD14A was that based on our filtering criteria, we failed to identify any ABHD14A sequences from class *Aves* within the *Chordates* (Figure 8A, B). This runs counter to the fact that *Aves* possessed the highest number of sequences for ABHD14B, suggesting that there might be three broad possibilities: (i) Perhaps our filtering criteria for identification of ABHD14A were too stringent; and/or (ii) *Aves* did not possess sequences for ABHD14A (lost during evolution); and/or (iii) ABHD14A sequences possessed by *Aves* were very distinct from ABHD14A sequences identified from other organisms. To exclude, possibility (i), we re-searched the 5000 hits, lowering the mismatch threshold to three and four mismatches, and were still unable to identify any ABHD14A sequences from *Aves*. Our current studies therefore, cannot exclude the possibilities (ii) and (iii), and further bioinformatics analysis for ABHD14A sequences specific to class *Aves* will be needed to resolve this preliminary yet interesting finding.

3.7 | Conservation of residues within the ABHD14A sequences

To assess the overall conservation in protein sequence for ABHD14A, we performed a multiple sequence alignment analysis on all the

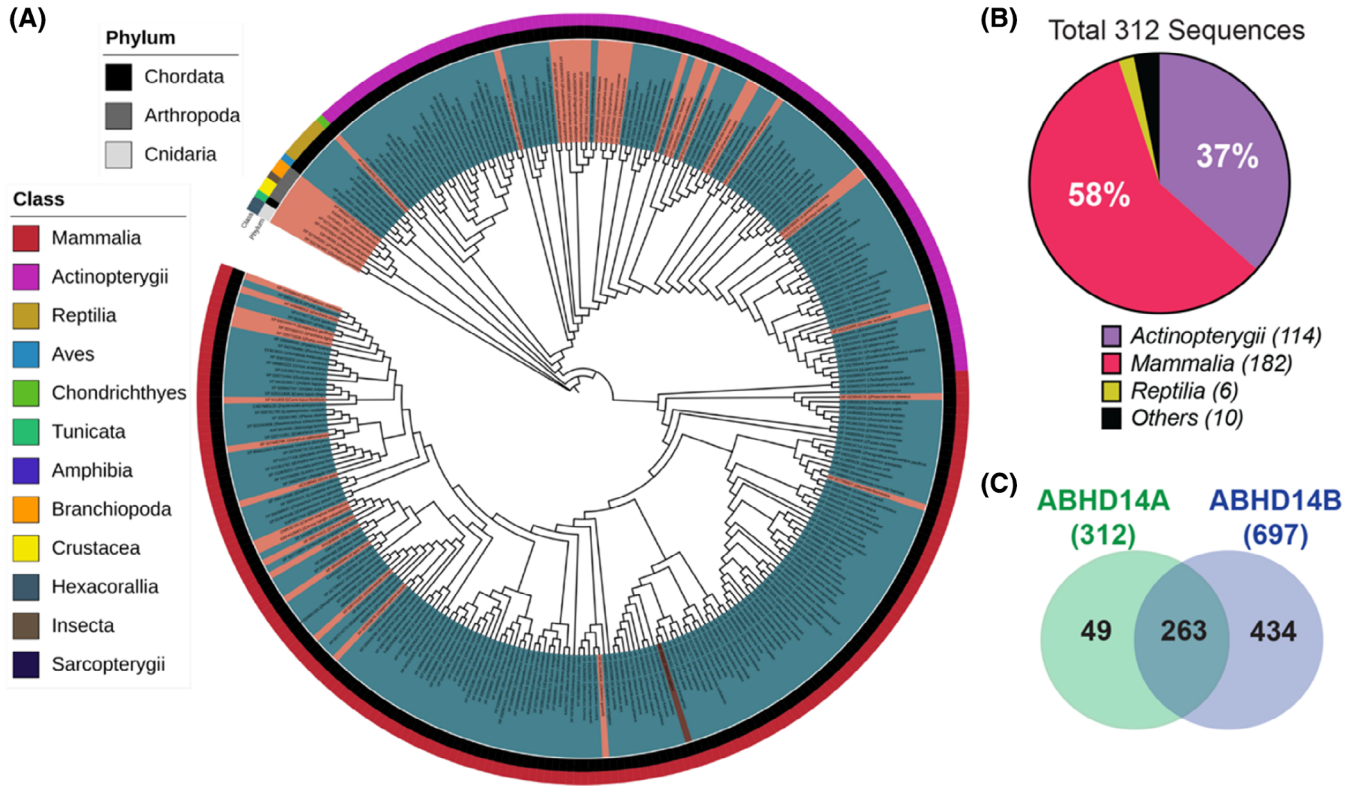


FIGURE 8 Phylogenetic analysis of ABHD14A sequences. (A) Phylogenetic tree representing the identified ABHD14A sequences from 312 organisms. The outermost and middle circular coloring denotes the Class and Phylum to which an ABHD14A sequence belongs (see associated legends within the figure), while the innermost circular coloring (orange or teal) denotes whether an organism contains a sequence only for ABHD14A (orange) or for both ABHD14B and ABHD14A (teal). (B) Pie-chart analysis representing the data from the phylogenetic tree for the various classes of organisms, mainly from *Chordates* that contain an ABHD14A sequence. This analysis shows that class *Mammalia* and *Actinopterygii* contain the most sequences for ABHD14A within the *Chordates*. (C) Venn diagram showing the overlapping number of organisms that contain ABHD14A and ABHD14B sequences among those identified from the bioinformatics analysis.

312 sequences identified for ABHD14A from different organisms (Supplementary File 2). Here, we found that from the 312 ABHD14A sequences, 99 residues were highly conserved (present at a frequency of >90% in all the 312 ABHD14A sequences at the defined position), suggesting that across all organisms, the overall sequence conservation is fairly high (~40%). In addition, upon closer inspection, for ABHD14A sequences, we found that within a particular class (e.g., *Mammalia*), the extent of sequence conservation was significantly higher (~80%) (Supplementary File 2 and Supplementary Table 2). Besides the conserved residues, we found that 50 residues were functionally conserved for the 312 ABHD14A sequences, suggesting that the overall sequence conservation across all organisms for ABHD14A was realistically ~60% (Supplementary File 2 and Supplementary Table 2).

Next, we wanted to map the conserved residues identified for ABHD14A to a three-dimensional protein structure. However, in the absence of any available ABHD14A structure, we used an AlphaFold Protein Structure Database predicted structure of human ABHD14A (Uniprot ID: Q9BUJ0).[35,36] To ensure that this structure was useful for further analysis, we first overlaid this AlphaFold-generated structure of human ABHD14A, with the crystal structure of human ABHD14B (PDB: 1imj). From this structural analysis, we found an almost perfect overlay of the ABHD-fold portion of ABHD14A with the overall structure of ABHD14B (Figure 9). In fact, the catalytic triad that comprises the active site of both these enzymes showed near identical orientations (Figure 9), and gave us confidence that the overall predicted structure of human ABHD14A would be useful for our structural analysis. Interestingly, the N-terminal region of the predicted human ABHD14A structure showed both an extended disordered stretch and an α-helical region that presumably anchors this enzyme to a cellular membrane (Figure 9). Finally, we mapped all the conserved residues on the AlphaFold predicted structure of human ABHD14A, and found that these conserved residues (including functionally conserved residues) were clustered around the enzyme active site (catalytic triad) and the cleft adjoining this active site where the putative substrates of this enzyme are predicted to bind (Figure 10). Further, we also found from this structural analysis that the residues present on the α-helical region at the N-terminus of the protein that is predicted to be involved in membrane anchoring of ABHD14A were also highly conserved (Figure 10).

This is interesting, because unlike the cellular localization of ABHD14B (cytosolic and nuclear),[22] ABHD14A might be localized to the plasma membrane or microsomal components (e.g., Golgi,

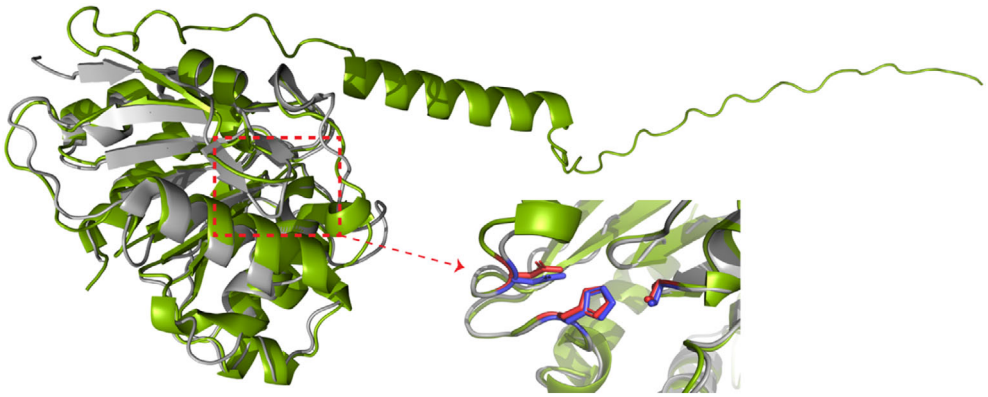


FIGURE 9 Overlay of the human ABHD14A and human ABHD14B structures. Superimposition of the AlphaFold predicted structure of human ABHD14A (Uniprot ID: Q9BUJ0) (green structure) with the three-dimensional crystal structure of ABHD14B (PDB: 1 imj) (gray structure) showing an almost perfect overlay of the ABHD-fold region of both proteins. A zoomed image of the catalytic triad also shows that these residues are perfectly aligned in both structures (blue residues for ABHD14B and red residues for ABHD14A). The ABHD14A predicted structure shows an additional N-terminal region that comprises of an α-helix, that is responsible for membrane anchoring of this enzyme.

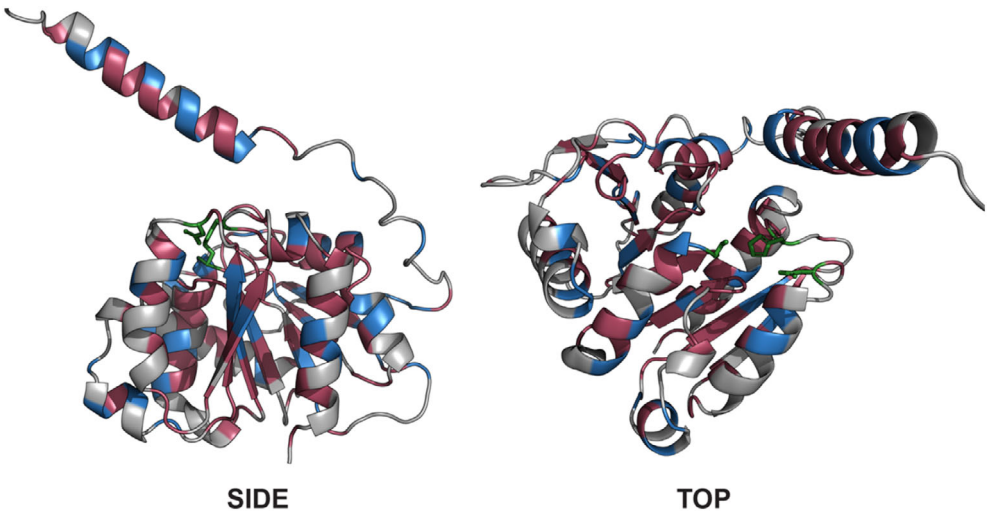


FIGURE 10 Mapping conserved residues on the ABHD14A structure. The residues colored in red are those that are absolutely conserved, while those shown in blue are functionally conserved based on the multiple sequence alignment analysis across all the 312 sequences of ABHD14A from different organisms. The catalytic triad residues that mark the active site are shown in green.

endoplasmic reticulum, peroxisome), and may have functions that are quite distinct from ABHD14B. While only speculative based on the bioinformatics data, this premise however needs to be experimentally validated. Also of note and biological importance is that despite the high overall sequence similarity (∼40%–50%) of ABHD14A and ABHD14B within a particular organism that possesses both enzymes (e.g., humans), it is interesting to note that ABHD14A lacks the acyltransferase motif. This implies that unlike the ABHD14B, perhaps ABHD14A performs a function distinct from the known acyltransferase-type activity of ABHD14B, and in the coming years, finding a biological activity/function of ABHD14A might be interesting from an evolutionary and phylogenetic standpoint.

4 | CONCLUSION

The advent of genome sequencing technologies has resulted in a deluge of available protein sequences, and mis-annotations of protein functions across public databases are increasingly propagating due to the over-reliance of automated tools of protein function assignment.[11] To overcome this, several efforts have been undertaken to tie in high-throughput bioinformatics analysis to experimental outputs to ensure proper validation of biochemical functions.[37] Realizing that this is an important problem, especially in an era of high data volumes and AI tools for interpreting them, here, we focused on development of a framework to identify and classify two outlying enzymes from the mSH family,[12] namely ABHD14A and ABHD14B. Until our recent functional annotation of ABHD14B as a novel KDAC,[22,23] both enzymes lacked any known biological function, despite their discoveries over two decades ago.[20,21] Specifically, in this study, we provide a bioinformatics framework that helps unambiguously identify a sequence as ABHD14A or ABHD14B from various databases across diverse organisms. Further, we provide an updated list of ABHD14A and ABHD14B sequences from various organisms that possess them, along with a detailed phylogenetic analysis for these sequences for researchers interested in studying these cryptic enzymes. As a proof

of concept, we also experimentally (biochemically) validate these findings by performing mutagenesis studies for a few residues of human ABHD14B. Here, we find that there are residues distal to the active site (e.g., S75 for human ABHD14B), that might regulate catalysis by possible allosteric mechanisms, a fact previously unknown for these enzymes. The knowledge and resources generated from this study will also be of biomedical importance, as it is now evident from several recent population-wide genetic association studies that dysregulation in ABHD14B expression has strong associations with different types of cancers,[38–43] while anomaly in ABHD14A expression is found to be associated with neurodegenerative conditions,[44–48] and even some cancers.[49–52]

AUTHOR CONTRIBUTIONS

Kaveri Vaidya: Methodology; investigation; writing – original draft; formal analysis; validation; writing – review and editing. **Golding Rodrigues:** Writing – original draft; investigation; methodology; validation; formal analysis; writing – review and editing. **Sonali Gupta:** Investigation; methodology; writing – review and editing. **Archit Devarajan:** Validation; investigation; writing – review and editing. **Mihika Yeolekar:** Investigation; writing – review and editing. **M. S. Madhusudhan:** Conceptualization; funding acquisition; writing – original draft; investigation; formal analysis; writing – review and editing; supervision. **Siddhesh S. Kamat:** Conceptualization; funding acquisition; writing – original draft; supervision; formal analysis; writing – review and editing.

ACKNOWLEDGMENTS

This work was supported by an EMBO Young Investigator Award and intramural funding from IISER Pune (to S.S.K.), and a Department of Biotechnology, Government of India grant to IISER Pune under the BICB Centre Scheme (BT/PR40262/BTIS/137/38/2022). K.V. and S.G. were supported by a Graduate Student Fellowship from IISER Pune and the Prime Minister's Research Fellowship, respectively. A.D. was supported by a DST-INSPIRE Scholarship.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/prot.26632.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Siddhesh S. Kamat https://orcid.org/0000-0001-6132-7574

REFERENCES

1. Wolfenden R, Snider MJ. The depth of chemical time and the power of enzymes as catalysts. *Acc Chem Res*. 2001;34:938-945.
2. Wolfenden R. Transition state analog inhibitors and enzyme catalysis. *Annu Rev Biophys Bioeng*. 1976;5:271-306.
3. Gerlt JA, Babbitt PC. Enzyme (re)design: lessons from natural evolution and computation. *Curr Opin Chem Biol*. 2009;13:10-18.
4. Gerlt JA, Babbitt PC. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem*. 2001;70:209-246.
5. Babbitt PC, Gerlt JA. New functions from old scaffolds: how nature reengineers enzymes for new functions. *Adv Protein Chem*. 2000;55:1-28.
6. Babbitt PC, Gerlt JA. Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities. *J Biol Chem*. 1997;272:30591-30594.
7. Glasner ME, Gerlt JA, Babbitt PC. Evolution of enzyme superfamilies. *Curr Opin Chem Biol*. 2006;10:492-497.
8. Glasner ME, Gerlt JA, Babbitt PC. Mechanisms of protein evolution and their application to protein engineering. *Adv Enzymol Relat Areas Mol Biol*. 2007;75:193-239. xii-xiii.
9. Gerlt JA, Babbitt PC. Barrels in pieces? *Nat Struct Biol*. 2001;8:5-7.
10. Gerlt JA, Babbitt PC. Can sequence determine function? *Genome Biol*. 2000;1:REVIEWS0005.
11. Gerlt JA, Allen KN, Almo SC, et al. The enzyme function initiative. *Biochemistry*. 2011;50:9950-9962.
12. Long JZ, Cravatt BF. The metabolic serine hydrolases and their functions in mammalian physiology and disease. *Chem Rev*. 2011;111:6022-6063.
13. Bachovchin DA, Cravatt BF. The pharmacological landscape and therapeutic potential of the serine hydrolases. *Nat Rev Drug Discov*. 2012;11:52-68.
14. Botos I, Wlodawer A. The expanding diversity of serine hydrolases. *Curr Opin Struct Biol*. 2007;17:683-690.
15. Lord CC, Thomas G, Brown JM. Mammalian alpha beta hydrolase domain (ABHD) proteins: lipid metabolizing enzymes at the interface of cell signaling and energy metabolism. *Biochim Biophys Acta*. 2013;1831:792-802.
16. Holmquist M. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr Protein Pept Sci*. 2000;1:209-235.
17. Nardini M, Dijkstra BW. Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol*. 1999;9:732-737.
18. Ollis DL, Cheah E, Cygler M, et al. The alpha/beta hydrolase fold. *Protein Eng*. 1992;5:197-211.
19. Simon GM, Cravatt BF. Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. *J Biol Chem*. 2010;285:11051-11055.
20. Hoshino J, Aruga J, Ishiguro A, Mikoshiba K. Dorz1, a novel gene expressed in differentiating cerebellar granule neurons, is downregulated in Zic1-deficient mouse. *Brain Res Mol Brain Res*. 2003;120:57-64.
21. Padmanabhan B, Kuzuhara T, Adachi N, Horikoshi M. The crystal structure of CCG1/TAF(II)250-interacting factor B (CIB). *J Biol Chem*. 2004;279:9615-9624.
22. Rajendran A, Vaidya K, Mendoza J, Bridwell-Rabb J, Kamat SS. Functional annotation of ABHD14B, an orphan serine hydrolase enzyme. *Biochemistry*. 2020;59:183-196.
23. Rajendran A, Soory A, Khandelwal N, Ratnaparkhi G, Kamat SS. A multiomics analysis reveals that the lysine deacetylase ABHD14B influences glucose metabolism in mammals. *J Biol Chem*. 2022;298:102128.
24. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389-3402.
25. Dobson L, Remenyi I, Tusnady GE. CCTOP: a consensus constrained TOPology prediction web server. *Nucleic Acids Res*. 2015;43:W408-W412.
26. Moller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 2001;17:646-653.

27. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970;48:443-453.
28. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059-3066.
29. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547-1549.
30. Letunic I, Bork P. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*. 2007;23:127-128.
31. Kumar K, Mhetre A, Ratnaparkhi GS, Kamat SS. A superfamily-wide activity atlas of serine hydrolases in Drosophila melanogaster. *Biochemistry*. 2021;60:1312-1324.
32. Kelkar DS, Ravikumar G, Mehendale N, et al. A chemical-genetic screen identifies ABHD12 as an oxidized-phosphatidylserine lipase. *Nat Chem Biol*. 2019;15:169-178.
33. Liu Y, Patricelli MP, Cravatt BF. Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci USA*. 1999;96:14694-14699.
34. Liu G, Huang L, Lian J. Alcohol acyltransferases for the biosynthesis of esters. *Biotechnol Biofuel Bioprod*. 2023;16:93.
35. Varadi M, Anyango S, Deshpande M, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res*. 2022;50:D439-D444.
36. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583-589.
37. Gerlt JA, Bouvier JT, Davidson DB, et al. Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim Biophys Acta*. 2015;1854:1019-1037.
38. Buchman AS, Yu L, Klein HU, et al. Proteome-wide discovery of cortical proteins that may provide motor resilience to offset the negative effects of pathologies in older adults. *J Gerontol A Biol Sci Med Sci*. 2023;78:494-503.
39. Kong SH, Lee JH, Bae JM, et al. In-depth proteomic signature of parathyroid carcinoma. *Eur J Endocrinol*. 2023;188:385-394.
40. Aushev VN, Gopalakrishnan K, Teitelbaum SL, et al. Tumor expression of environmental chemical-responsive genes and breast cancer mortality. *Endocr Relat Cancer*. 2019;26:843-851.
41. Posorski N, Kaemmerer D, Ernst G, et al. Localization of sporadic neuroendocrine tumors by gene expression analysis of their metastases. *Clin Exp Metastas*. 2011;28:637-647.
42. Kaemmerer D, Posorski N, von Eggeling F, et al. The search for the primary tumor in metastasized gastroenteropancreatic neuroendocrine neoplasm. *Clin Exp Metastasis*. 2014;31:817-827.
43. Namani A, Cui QQ, Wu YH, Wang HY, Wang XJ, Tang XW. NRF2-regulated metabolic gene signature as a prognostic biomarker in non-small cell lung cancer. *Oncotarget*. 2017;8:69847-69862.
44. Casey JP, Magalhaes T, Conroy JM, et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum Genet*. 2012;131:565-579.
45. Wu XH, Chen WD, Lin FQ, et al. DNA methylation profile is a quantitative measure of biological aging in children. *Aging*. 2019;11:10031-10051.
46. Cukier HN, Dueker ND, Slifer SH, et al. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism*. 2014;5:1.
47. Preeprem T, Gibson G. SDS, a structural disruption score for assessment of missense variant deleteriousness. *Front Genet*. 2014;5:82.
48. Henrichsen CN, Csardi G, Zabot MT, et al. Using transcription modules to identify expression clusters perturbed in Williams-Beuren syndrome. *PLoS Comput Biol*. 2011;7:e1001054.
49. Cava C, Armaos A, Lang B, Tartaglia GG, Castiglioni I. Identification of long non-coding RNAs and RNA binding proteins in breast cancer subtypes. *Sci Rep*. 2022;12:693.
50. Savci-Heijink CD, Halfwerk H, Koster J, Horlings HM, van de Vijver MJ. A specific gene expression signature for visceral organ metastasis in breast cancer. *BMC Cancer*. 2019;19:333.
51. He Y, Wei C, Sun Z, et al. Genome-wide methylation profiling reveals differentially methylated genes in blood DNA of small-cell lung cancer patients. *Precis Clin Med*. 2022;5:pbac017.
52. Chang YS, Tu SJ, Chiang HS, et al. Genome-wide analysis of prognostic alternative splicing signature and splicing factors in lung adenocarcinoma. *Genes*. 2020;11:1300.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Vaidya K, Rodrigues G, Gupta S, et al. Identification of sequence determinants for the ABHD14 enzymes. *Proteins*. 2025;93(1):255-266. doi:10.1002/prot.26632