THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

Analysis of the Evolution of the MoxR ATPases

Published as part of The Journal of Physical Chemistry virtual special issue "Protein Folding and Dynamics—An Overview on the Occasion of Harold Scheraga's 100th Birthday".

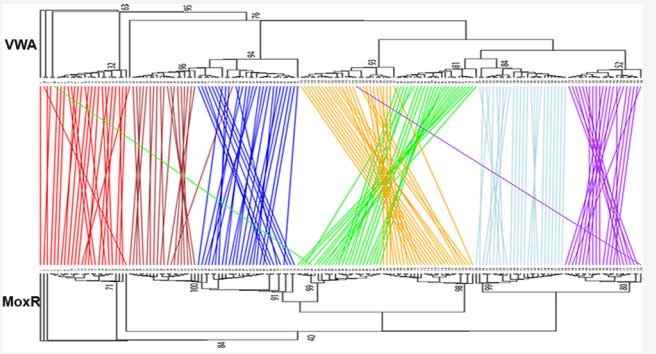
Vaibhav Bhandari, David A. J. Van Ommen, Keith S. Wong, and Walid A. Houry*

Cite This: *J. Phys. Chem. A* 2022, 126, 4734–4746

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: MoxR proteins comprise a family of ATPases Associated with diverse cellular Activities (AAA+). These proteins are widespread and found across the diversity of prokaryotic species. Despite their ubiquity, members of the group remain poorly characterized. Only a few examples of MoxR proteins have been associated with cellular roles, where they have been shown to perform chaperone-like functions. A characteristic feature of MoxR proteins is their association with proteins containing the von Willebrand factor type A (VWA) domain. In an effort to understand the spread and diversity of the MoxR family, an evolutionary approach was undertaken. Phylogenetic techniques were used to define nine major subfamilies within the MoxR family. A combination of phylogenetic and genomic approaches was utilized to explore the extent of the partnership between the MoxR and VWA domain containing proteins (VWA proteins). These analyses led to the clarification of genetic linkages between MoxR and VWA proteins. A significant partnership is described here, as seven of nine MoxR subfamilies were found to be linked to VWA proteins. Available genomic data were also used to assess the intraprotein diversification of MoxR and VWA protein sequences. Data clearly indicated that, in MoxR proteins, the ATPase domain is maintained with high conservation while the remaining protein sequence evolves at a faster rate; a similar pattern was observed for the VWA domain in VWA proteins. Overall, our data present insights into the modular evolution of MoxR ATPases.



INTRODUCTION

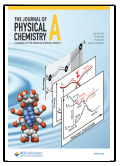
ATPases Associated with various cellular Activities (AAA+) (ATP = adenosine triphosphate) are a superfamily of P-loop NTPases (NTP = nucleoside triphosphate).[1,2] Members of this superfamily are characterized by the presence of the AAA+ module.[2,3] This module is 200-250 amino acids long, containing a few conserved sequence motifs that bind and hydrolyze ATP.[2,3] AAA+ proteins utilize the energy generated from ATP hydrolysis for the molecular remodeling and manipulation of target substrates. These AAA+ proteins are highly diverse and are involved in a variety of cellular processes. Some well-known examples include dynein motor protein, the DnaA protein involved in DNA replication, the protein unfolding ClpX chaperone, and the Lon protease.[4]

Within the AAA+ superfamily lies the MoxR family. Many bacteria, including the model, pathogenic species such as *Pseudomonas aeruginosa*, *Mycoplasma tuberculosis*, and *Escherichia coli* contain multiple such proteins.[4] Despite the prevalence of MoxR throughout prokaryotic species, few proteins from this family have been characterized, and a general understanding of their functional roles is lacking.[4,5]

In trying to gain insights into this obscure group of ATPases, we focused on a noticeable feature of the MoxR proteins, their genetic proximity to proteins containing a von Willebrand Factor A (VWA) domain.[6] Despite the limited experimental information on the MoxR family, this genetic linkage between the MoxR and VWA led to an inference of a functional association.[5,7] The VWA domain is a known metal-binding domain involved in mediating protein-protein interactions.[8,9] Thus, the extent of a partnership between the MoxR ATPases to VWA domain-containing proteins (VWA proteins) was explored.

On the basis of known examples, the binary relationship between MoxR ATPases and VWA domain-containing proteins appears to follow a common blueprint. The relationship is often expressed as an adaptor and effector duality. The VWA-containing protein acts as an adaptor, helping localize the MoxR protein to a substrate. The MoxR protein uses ATP hydrolysis to power its effector function on the substrate. Some examples that

Received: April 13, 2022
Revised: July 2, 2022
Published: July 19, 2022



https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734–4746

follow this blueprint include: the rubisco activase CbbQ and its VWA domain-containing partner CbbO, characterized in *Acidithiobacillus ferrooxidans*;[10] the Nitrate reductase maturation protein NorQ and its partner NorD in *Paracoccus denitrificans*;[11] and the RavA-ViaA pair in *E. coli*.[12]

Here, phylogenetic techniques are employed to assess the relationship between MoxR proteins and their VWA domain partners on a wider evolutionary scale. The extent of this protein partnership across species is explored using protein and species phylogenies. Additionally, the RavA subfamily of the MoxR family and the ViaA group of VWA-containing proteins are used here to examine intraprotein evolutionary trends. Research in our laboratory has structurally and functionally explored the RavA and ViaA protein partnership.[6,12−14] In *E. coli*, the respective genes of the two proteins are on the same operon. Functionally, RavA-ViaA are linked to acid stress and respiratory processes.[6,12−14] ViaA has been shown to act as an adaptor in recruiting RavA to its respiratory enzymatic substrates.[12] A knowledge of these proteins was utilized in assessing sequence variation across bacterial orthologues. Together, the data presented here provide novel insights into the understudied MoxR-VWA relationship.

MATERIALS AND METHODS

Phylogenetic Analysis for MoxR and VWA Trees. Using previous categorizations of the MoxR group, 22 MoxR proteins, representing each subfamily, were used as queries to identify similar sequences in the nonredundant protein sequence database (Table 1).[5,15] This was performed with PSI-BLAST[16] with two iterations. The sequences ranked by e-values from every query were combined and screened with the UniProtKB server[17] to remove redundant sequences. A total of 4760 unique MoxR protein sequences were clustered into UniRef50 groups.[18] Proteins in the same UniRef50 group share a minimum of 50% sequence identity and an 80% lengthwise overlap with the longest sequence in a cluster. The clusters are collectively represented by the member protein with the longest sequence.[18] This yielded 421 unique UniRef50 sequences that were used in subsequent phylogenetic analyses.

Table 1. List of Query Sequences Used for Compilation of MoxR and VWA Proteins

MoxR subgroup	query sequences for AAA+ (given as UniProt accession numbers)	query sequences for VWA related to AAA+ (given as UniProt accession numbers)
CoxD	Q51326, P71922, Q9Y9R8	Q9KX26, P71923, Q9Y9R7
CGN	Q51481, Q51858, Q9HI16	Q51484, Q3J137, Q8PST9
MRP	P30621, Q1MDJ0, Q2A5K6	C5AQA2, C5AQA0, Q1MDI8, Q2A5K5, Q2A5K3, Q2A5K2
PA2707	Q9I0D5, Q7VVY8, Q63VZ7	Q7W524, Q82N54, Q3KGF5
RavA	P31473, Q3 V4Q1, Q58222	P0ADN0, Q3 V4Q4, Q58221
TM0930	Q97HZ8, Q9 × 028, Q7UG34	Q97HZ9, G4FF61, Q7UG35
YehL	P33348, Q9RJY7, Q7US48	P33352, Q9RJY6, Q7US47

For construction of the phylogenetic trees, the 421 UniRef50 sequences were aligned with the MUSCLE alignment tool[19,20] included in the MEGA6 program package.[19,20] The alignment was manually edited, whereby only the AAA+ domain of each MoxR protein was retained. A second MUSCLE alignment was then performed on the AAA+ domain sequences and was used for a phylogenetic analysis.[19] A phylogenetic tree was constructed via the maximum-likelihood method based on the Whelan-Goldman + frequency model.[21] For the construction of the tree, only the amino acid positions present in 50% or more of the sequences of a multiple sequence alignment were included - a total of 236 residue positions. Gamma distribution (four categories) was used to model evolutionary rate differences among sites. The tree with the highest log-likelihood is presented.

A similar approach was taken for our analysis of the VWA proteins, in which 3605 unique VWA sequences were clustered into 745 UniRef50 groups, and their respective representative proteins were utilized for the construction of a VWA tree. From the multiple sequence alignment of the VWA domains, the residue positions present in 50% or more of the aligned sequences were included for the tree construction. A total of 156 aligned positions of the VWA domain was used.

Genomic Association Analysis. Gene coordinates for both the MoxR and the VWA proteins, available through the National Center for Biotechnology Information (NCBI) genome database for download, were utilized in our analysis.[22,23] A cutoff of 20 open reading frames (ORFs) was used, such that a VWA-encoding gene that is positioned within 20 ORFs relative to the MoxR-encoding gene, upstream or downstream and on either one of the DNA strands, was considered proximal. This yielded 2493 unique MoxR-VWA pairs that were used in our comparative analysis on the branching of trees. Pairings were ignored if either the VWA or MoxR protein of a pair was not part of the phylogenetic analysis or if MoxR and VWA proteins were not in proximity to each other. Supporting Information Table 1 provides the names of the MoxR and VWA pairings.

Construction of the 16S rRNA Phylogenetic Tree. A total of 618 species belonging to the gamma-proteobacteria had complete genomic sequences available at the time of analysis. 16S ribosomal RNA (rRNA) sequences were obtained for each of these species. To retrieve their sequences, the *E. coli* 16S rRNA sequence (NCBI reference sequence NR_114042.1) was used as a query. Using default parameters, the query was used in the blastN suite of the BLAST program to search and obtain the sequences from the rRNA/ITS database and search filter.[24] Additionally, 16S rRNA from three β-proteobacteria species were obtained to root phylogenetic trees. These are the *Neisseria meningitidis*, *Nitrosomonas halophila*, and *Bordetella pertussis*.

Alignment of the 621 sequences was performed using the ClustalW algorithm provided in the MEGA X package.[25,26] The PhyML program was used to construct a maximum likelihood tree with 100 bootstrap replicates.[27]

Assembly of Cophyloplot Trees. The pairing of MoxR ATPases and VWA-containing partners found in the same organism was identified through genomic analysis as detailed above. The MoxR subfamilies were linked to VWA protein groupings through phylogenetic and genetic links. The MoxR-VWA pairings include the NorQ/CbbQ-NorD/CbbO, CoxD-CoxE, MRP-DP0636, MRP-MxaL, MRP-MxaC, PA2707-BPP3480, RavA-ViaA, TM0930-TM0929, and YehL-YehP. For each of these groups, the MoxR protein sequences and their genetically linked VWA protein sequences were obtained.[17]

Cophyloplots were constructed for each of the nine MoxR-VWA pairs. First, individual Maximum Likelihood trees were constructed for each MoxR protein subfamily and for the corresponding VWA protein subfamily. The R programing language was utilized for the cophyloplot analysis.[28] Specifically, the Cophyloplot tool found within the APE package was used.[29]

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

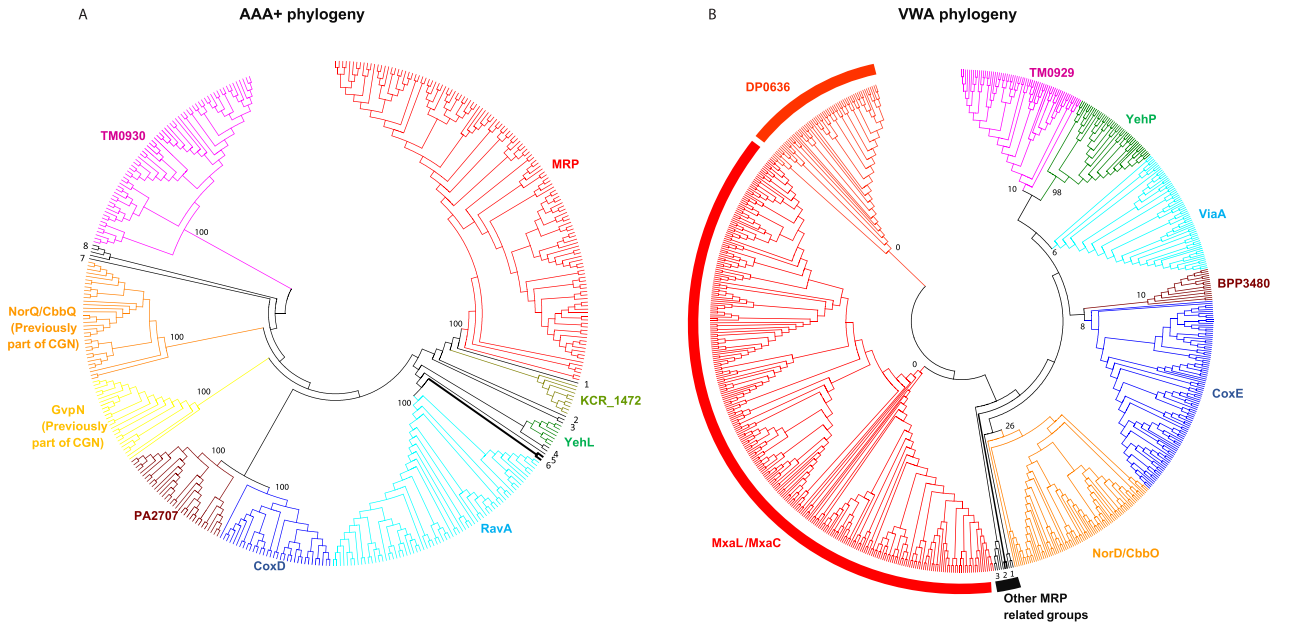


Figure 1. Phylogenetic trees for the MoxR AAA+ ATPases and related VWA domain-containing proteins. (A) Topological presentation of 421 clusters, which represent 4760 MoxR AAA+ ATPases. Members of the MoxR family are shown divided into nine named subfamilies and eight smaller numbered clades. The branch representative of the two viral proteins of MoxR is indicated with a thick line. (B) The resolved maximum likelihood tree depicts the grouping of clustered UniRef50 representatives for various VWA domain containing proteins. Analogous to the AAA+ MoxR family proteins, these proteins can be observed to mostly branch along divisions of the major subgroups of the MoxR family ATPases. The MRP-related VWA proteins can be divided into two large clades, named DP0636 and MxaL/MxaC group, and three smaller ones numbered 1, 2, and 3. The two viral VWA-containing proteins are highlighted using thicker lines. The clades of MRP-related groups are indicated with colored, curved bars. Bootstrap values for major clades are indicated.

A cophyloplot shows two trees corresponding to a MoxR-VWA partnership with connections indicating the position of each species across the trees for the two partner proteins.

In addition to the cophylogenetic trees, pairwise distance matrix-based scatterplots were constructed for each MoxR and VWA-containing protein partnership.[25] A pairwise distance matrix was obtained for each tree in a cophyloplot using the APE package, Cophyloplot tool.[29] For each MoxR-VWA partnership, 10 distinct species were selected as the root or reference. These root species were set at the origin (0, 0). The relative distances of other species were plotted in reference to the root. The slope and coefficient of determination for a linear fit were calculated for each scatterplot. One of the 10 scatterplots for each MoxR-VWA relationship is shown as a reference in Figure 3 - the MoxR distances on the *X*-axis, the VWA protein distances on the *Y*-axis. The average slope and R^2 calculated from the 10 scatterplots are noted for each MoxR-VWA relationship in Figure 3.

A representative cophyloplot was also prepared for the seven MoxR groups and their VWA partners together. A sum of 176 protein sequences, representing the different MoxR and VWA groupings, were selected for each tree. As the MRP group is associated with multiple VWA groups, only the DP0636 was selected for use here.

Sequence Conservation Analysis. Full-length sequences for both RavA and ViaA proteins were identified in 225 species of the gamma-proteobacteria group. For both RavA and ViaA, the sequences from these species were retrieved and aligned using ClustalW as described above.[26] The alignment of sequences for each protein were then used as input to evaluate conservation of residues. The *E. coli* RavA or ViaA sequences were used as a reference when referring to the residue positions. The variability for each amino acid was measured using the Shannon entropy method available through the Protein Variability Server.[30,31]

Detecting the Presence of RavA-ViaA across Sequenced Prokaryotic Genomes. The representation of orthologous sequences for RavA and ViaA proteins was inspected across prokaryotic species. The protein sequences for the *E. coli* RavA (RefSeq ID: WP_088209852.1) and ViaA (RefSeq ID: WP_089580935.1) were obtained from the NCBI database.[32] These were used as query sequences in the standard protein BLAST (blastp) suite to search for similar sequences across selected species.[24] The nonredundant protein sequences database was searched using default settings. Protein sequences sharing similarity over 70% of the total length of the query sequence and having an e-value less than 1×10^{-5} were identified as homologous. The queries were searched against a total of 3135 prokaryotic species, for which complete genome sequences were available. The species represent over 25 bacterial and 3 archaeal phyla.

In addition, if protein sequences were identified to be homologous to RavA or share similarity across the AAA+ domain using BLAST searches, pairwise alignment was performed with the identified protein against RavA. The megaBLAST algorithm and default parameters were used to assess the region of sequence similarity between the two proteins.[33] A similar assessment was performed with sequences identified to be similar to the *E. coli* ViaA protein.

Modeling of a Protein Structure De Novo. The de novo structure model for ViaA was calculated by using AlphaFold2.[34,35] The program was made available through the Google Colab platform, from the following Web site: https://colab.research.google.com/drive/1LVPSOf4L502F21RWBmYJJYYLDlOU2NTL#scrollTo=UGUBLzB3C6WN The program was accessed on 2021-07-19.

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

Table 2. Summary of the Associations between MoxR AAA+ Proteins and VWA Domain Containing Proteins Observed in Genetic Proximity and Corroborated through Phylogeny

MoxR group	Number of distinct proteins observed to interact with VWA domain proteins	Number of interactions with VWA	Corresponding VWA group	Number (and percentage) of VWA domain proteins in corresponding phylogenetic group that are genetically linked to a different MoxR subgroup
NorQ/CbbQ	245	248	NorD/CbbO	3 of 248 (1.21%)
CoxD	266	268	CoxE	5 of 268 (1.87%)
MRP	518	1036	MRP-related	6 of 1036 (0.58%)
PA2707	294	295	BPP3480	1 of 295 (0.34%)
RavA	319	319	ViaA	2 of 319 (0.63%)
TM0930	109	113	TM0929	7 of 113 (6.2%)
YehL	212	214	YehP	3 of 214 (1.4%)
Total	**1963**	**2493**		**27 of 2493 (1.08%)**

The *E. coli* ViaA protein sequence (RefSeq ID: WP_000956642.1) was used as input. The best among five default models is presented in Figure 6C.

RESULTS AND DISCUSSION

Phylogenetic Analysis on MoxR AAA+ ATPases and Their Associated VWA Proteins. The MoxR AAA+ ATPases were previously categorized into seven defined subfamilies based on the sequence similarity of their AAA+ domains and local genetic structures.[5] The phylogenetic analysis performed at that time was based only on 596 MoxR proteins. Given the increase in sequenced genomes and functional annotations on proteins made available over the years, we decided to perform a more comprehensive analysis on the MoxR proteins as well as their associated VWA proteins. Using PSI-BLAST, 4760 proteins were identified from the use of 22 query proteins. The queries represent each subgroup of the MoxR family that had been previously characterized (Table 1).[5] To limit the number of similar protein sequences in the subsequent phylogenetic analysis, the 4760 MoxR proteins were grouped into 421 representative UniRef50 clusters. The representative sequences for each cluster were analyzed and used in the construction of a maximum likelihood tree shown in Figure 1A.

Gene annotations and phylogenetic branching were used in the identification of MoxR subfamilies. Consistent with our previous analysis, the subfamilies CGN (CbbQ-NorQ-GvpN), CoxD (previously APE2220), MRP, PA2707, RavA, TM0930, and YehL are all present in our new MoxR tree and form distinct branches. However, proteins of the APE0892 group, which were postulated to be an eighth subfamily of MoxR,[5] were observed to branch within the CoxD subfamily in the tree. Furthermore, a previously unidentified group is shown here to form a distinct branch composed of 21 proteins (Figure 1A). This branch is named Kcr_1472, based on the deepest branching member. Eight other smaller clusters of proteins are also observed in the tree (Figure 1A; numbered 1 to 8), which were not easily grouped into any known MoxR subfamilies. These proteins may represent divergent subfamilies, or they may be distantly associated with the known subfamilies. However, more data are required before such conclusions can be drawn. Of note are the two viral MoxR proteins (cluster 6) that branch next to the RavA subfamily (Figure 1A). This proximal branching agrees with the work by Scheele and co-workers on one of these two viral proteins, p618, from the archaeal two-tailed virus (RefSeq ID: YP_319897.1, UniRef ID: Q3V4Q1). The authors found p618 to be structurally similar to *E. coli* RavA.[36,37]

Among the branches formed by the major, known subfamilies, it was interesting to note that, though easily differentiated from other MoxR subfamilies, the CGN subfamily did not form a singular branch. Proteins characterized as CbbQ, NorQ, and GvpN were previously considered part of a singular group. Here, while the CbbQ and NorQ proteins could not be distinguished from each other, they are seen to form a cluster distinct from the GvpN proteins (Figure 1A). This distinction between the GvpN proteins and NorQ/CbbQ is further highlighted by their association with nearby VWA proteins. While not always a defining factor, VWA proteins are often found near NorQ and CbbQ proteins, but they were not detected with GvpN proteins.[5] Therefore, the CGN subfamily is redefined here as two subfamilies: NorQ/CbbQ and GvpN (Figure 1A).

Given that a substantial number of MoxR proteins were previously associated with VWA proteins,[5] a similar phylogenetic analysis was also performed on 3605 unique VWA proteins. The phylogenetic tree for the VWA proteins is shown in Figure 1B. The bootstrap support for VWA proteins is not as robust as for the MoxR subfamily. This is likely due to the large number of sequences used in the tree and being limited to the use of only the length of the VWA domain. Additionally, unlike the AAA domain, which contains several conserved motifs across its length, the VWA domain is primarily characterized by its short, conserved, metal ion-dependent adhesion site (MIDAS) motif.[8] Because of these factors, the individual branching within the tree may be difficult to replicate. The tree in Figure 1B provides the tree with the highest likelihood based on the input criteria. An important observation is that the organization of the VWA tree (Figure 1B) is highly similar to the MoxR tree, such that VWA subfamilies can be readily identified and paired with their respective MoxR subfamilies. These include CoxE (CoxD-related), NorD/CbbO (NorQ/CbbQ-related), DP0636 (MRP-related), MxaL/MxaC (MRP-related), BPP3480 (PA2707-related), ViaA (RavA-related), TM0929 (TM0930-related), and YehP (YehL-related).

VWA Domain-Containing Proteins Can Often Be Found in Close Genomic Proximity to MoxR. The similar branching of the VWA subfamilies to the MoxR subfamilies, in their respective trees, is suggestive of association between the two protein groups. To investigate the linkage implied in Figure 1, a genomic pairing analysis was performed. For each MoxR protein included in the MoxR tree (Figure 1A), all VWA proteins in its proximity were identified using current gene annotations and position data. Identified pairs of VWA and MoxR proteins are listed in the Supporting Information Table 1. We note that MoxR proteins that are not associated with any VWA proteins (i.e., Kcr_1472 and GvpN members) were not used in this analysis and that many MoxR-VWA associations may not have been identified here due to incomplete or even

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

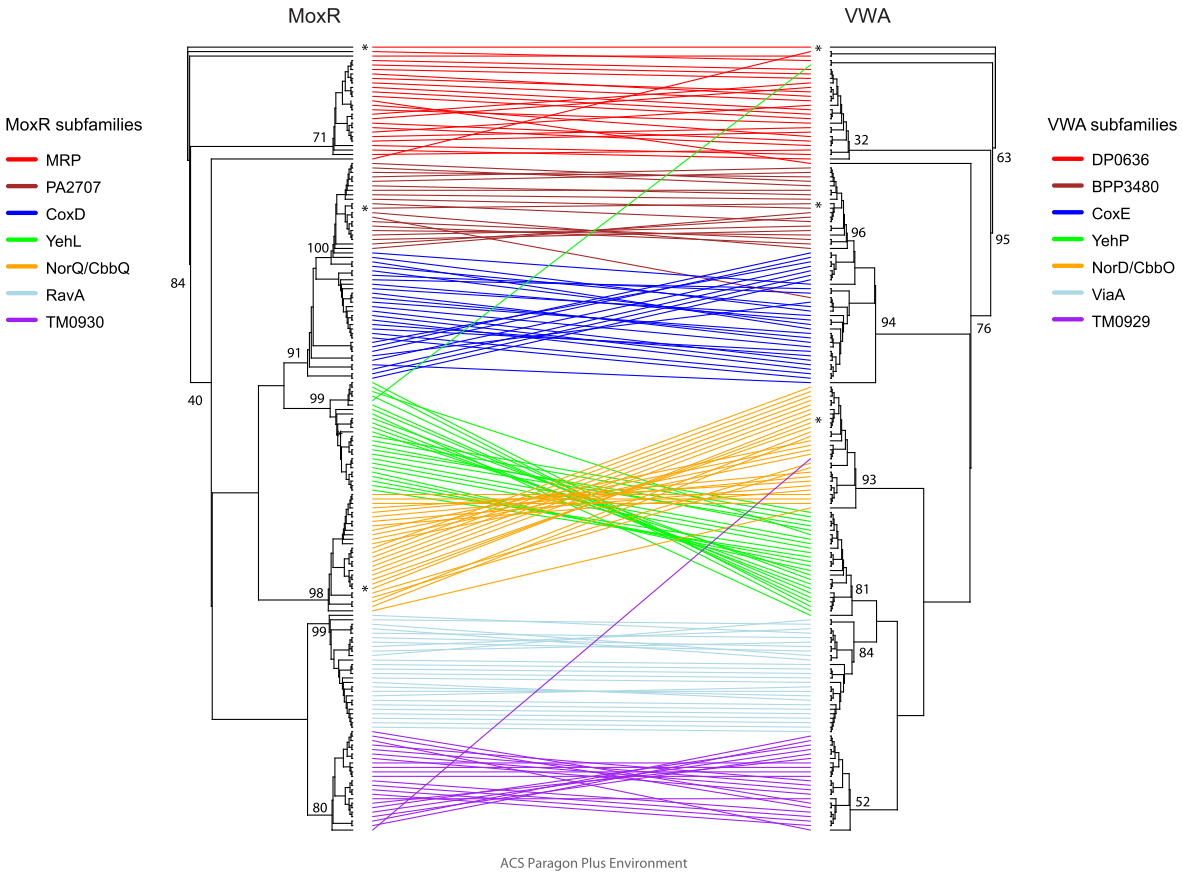


Figure 2. Cophyloplot presents the parallel evolutionary trend for MoxR-VWA. 176 representative sequences for genetically linked MoxR and VWA proteins were used to construct independent phylogenetic trees. The positions of genetically linked proteins across the two trees are connected. The lines are colored based on the subgrouping for the MoxR and VWA in Figure 1. Numbers next to branches provide references to the protein sequence used. Bootstrap values for major clades are indicated. (*) The position of proteins from *P. aeruginosa*.

inaccurate gene annotations. The results are summarized in Table 2.

A total of 2493 unique MoxR-VWA pairings were identified with our methods. Importantly, there are very few exceptions observed (less than 2% in most cases) where a VWA protein does not belong to the phylogenetic group associated with its genetically linked MoxR protein (Table 2). In other words, the similar branching of the VWA and MoxR trees is validated through genetic proximity. This is consistent with the idea that the similar branching of the MoxR and VWA trees is reflective of the close association of the MoxR subfamilies and the respective VWA subfamilies. This association is true not only for the few examples of functionally characterized MoxR ATPases but is widespread across this family of AAA+ ATPases.

An interesting case arises for the MRP subfamily of MoxR, where one MRP member can be associated with up to four VWA proteins (Supporting Information Table 1). Approximately 80% of the MRP proteins identified in this analysis are associated with more than one VWA protein (Supporting Information Table 1). Our phylogenetic analysis revealed two large subfamilies and three small clusters of MRP-related VWA proteins (Figure 1B). Of the large subfamilies, the MxaL/MxaC subfamily is named after MxaL and MxaC from *Methylobacterium extorquens*, which were among the first MRP-related VWA proteins to be characterized.[38] MxaL and MxaC could not be delineated based on their branching within the phylogenetic tree. However, in the species *Methylobacterium extorquens*, *mxaL* and *mxaC* are distinguished as two protein encoding genes in the *mxa* cluster.[38] Thus, MxaL and MxaC are here considered to be separate subgroups among the VWA proteins. The second subfamily, DP0636, is named after a representative member of the clade.

Presentation of Evolutionary Correlation through the Cophyloplot Analysis. To visualize the coevolving partnership shared by MoxR and VWA proteins, a cophyloplot analysis was performed. A representative subset of MoxR proteins was selected along with corresponding VWA proteins in genetic proximity. Also, since the MRP group is associated with multiple VWA groups, only the DP0636 was selected for clarity. To perform the analysis, the boundaries of the AAA+ domain for each of the MoxR proteins were identified and used in the sequence alignment. Similarly, only the VWA domain from VWA domain-containing proteins were aligned. Then, individual trees were constructed for the obtained set of AAA+ and VWA domain sequences. The nodes corresponding to the genetically proximal proteins from the same organism were connected. This is shown in Figure 2.

The two trees represent seven MoxR and VWA families each. Both trees contain 176 protein sequences from 98 distinct prokaryotic species. Among these, multiple MoxR and VWA proteins were included from 49 species. The results show that the domains from the MoxR proteins separate into clades, which replicate recognizable MoxR subgroups from Figure 1. With a few exceptions, the relative positions of the VWA-containing proteins reflect that of their MoxR counterparts.

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

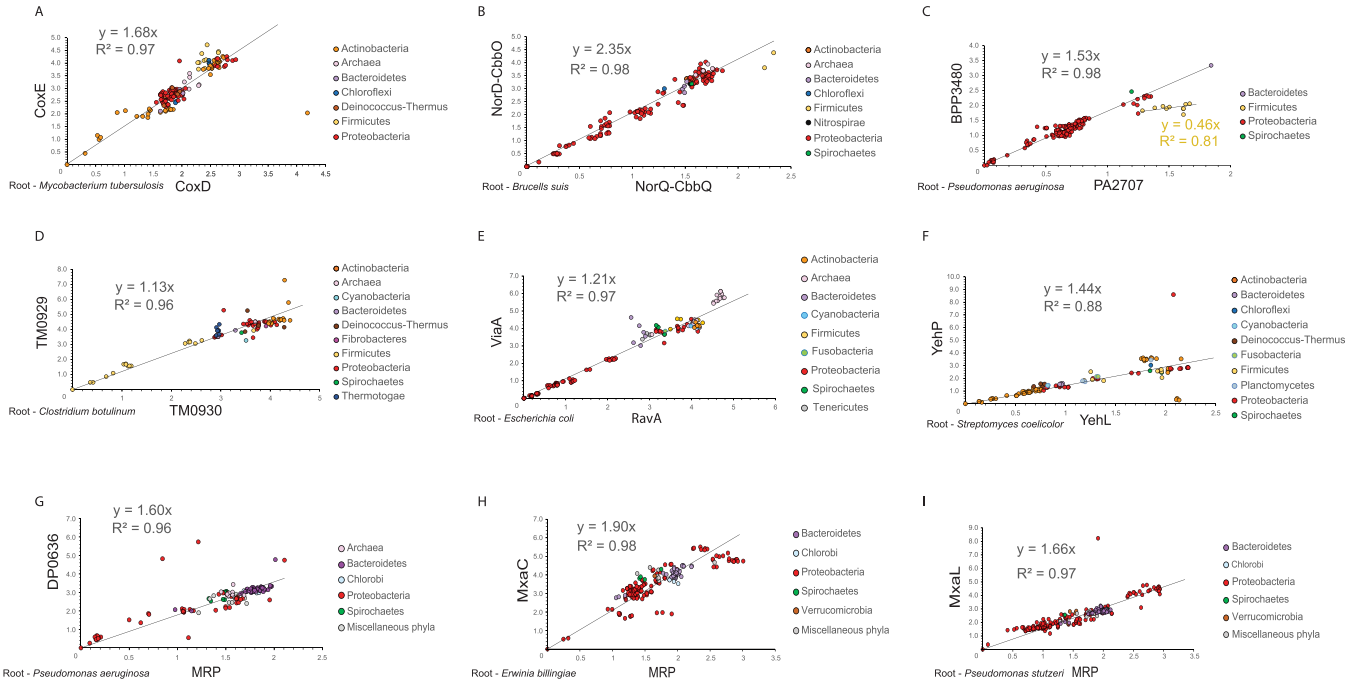


Figure 3. The relative sequence diversification for MoxR-VWA partners. Shown are scatterplots for the sequence diversity for each MoxR-VWA pairing (A–I). In each plot, the sequence diversity for the MoxR group is presented on the *X*-axis, and the VWA group is on the *Y*-axis. Based on the pairwise distance matrices, a total of 10 scatterplots for each MoxR-VWA partnership were constructed, one of which is shown as a representative. The standard deviation from the slope is less than 10% for each MoxR-VWA pair. Protein sequences from one organism were used as a reference for each scatterplot. This is referred to as the root. A colored dot represents diversity of sequence for the protein pair of a single species in relation to the root. The colors of the dots indicate the phylum for the organism. The average slope and fit from the 10 scatterplots are given for each MoxR-VWA pair. In (C), two lines are presented. These highlight the different relationship shared by the BPP3480 and PA2707 in different taxonomic groups.

The cophyloplot provides an opportunity to observe the branching of protein domains obtained from the same species. It is notable that, when multiple AAA+ domains were used from different proteins encoded by the same organism, they most often branch within their MoxR subgrouping and not with other proteins belonging to the same species. Importantly, this is reflected in the branching of their VWA partner. Proteins from *P. aeruginosa* provide a pertinent example (Figure 2). Three proteins from the species are present in the MoxR tree: one representative each from the MoxR, PA2707, and NorQ/CbbQ groups. The proteins are indicated with (*) in the cophyloplot (Figure 2). Though the three proteins are encoded by the same species, each of them branches into clades with other proteins belonging to the same MoxR subfamily. Such branching reflects the paralogous relationship between the subfamilies of MoxR proteins. This relationship is reflected in the positions of the VWA proteins that the MoxR proteins are genetically linked to (Figure 2, right tree). Together, the branching of MoxR and VWA genetic partners in the cophyloplot points to their partnered coevolution.

Considering the MoxR subfamilies and the VWA groups, the positioning of genetically related groups is generally similar in the MoxR and VWA trees. One notable difference across the two trees is seen with the swapping of positions of the YehP and NorD/CbbO groups of the VWA in relation to their MoxR partners. It is crucial to note that the groups maintain their integrity and distinction from each other, with high support expressed in bootstrap values. The YehP proteins form a separate clade from the NorD/CbbO group. The observed swapping of positions across the two trees indicates a difference in the rate of sequence diversification. On the basis of data discussed below (see Figure 3), the NorD/CbbO group of VWA proteins has a high rate of sequence diversification in relation to their MoxR partners. Thus, despite the positional disparity, the cohesiveness of the cladal boundaries shows the close relationship shared by the VWA proteins and their MoxR partners.

The consistent presence of VWA domain proteins with MoxR proteins is seen with a genetic proximity analysis (Table 2, Supporting Information Table 1). This points toward an evolutionary relationship between these two types of proteins. This is bolstered by the cophyloplot analysis. In the cophyloplots, two independent trees establish the subgrouping among the MoxR and VWA proteins, respectively. The mirroring of these relationships across the two trees, presented in the form of the linked lines (Figure 2), establishes the coevolving partnership shared by MoxR and VWA.

The VWA Sequences Show Greater Diversity than Their MoxR Counterparts. The phylogenetic and genomic analyses show that VWA proteins are associated with MoxR subfamilies (see Figure 1, Table 2). Namely, the MoxR subfamilies are the NorQ/CbbQ, CoxD, MRP, PA2707, RavA, TM0930, and YehL. Among these, the MRP subfamily has been genetically linked to three VWA protein groups (Figure 1, Table 2). For each of these groups and their corresponding VWA protein groupings, a cophyloplot analysis was performed. The cophyloplot is used here as a tool to compare the rates of sequence diversification for two genetically linked protein groups across different species. For the analysis, separate species trees were constructed for each MoxR group and its corresponding VWA group. The position for the same species in the two protein trees was linked with connecting lines (Supporting Information Figure 1). Across the two trees of a cophyloplot, each organism is seen to branch at similar

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734–4746

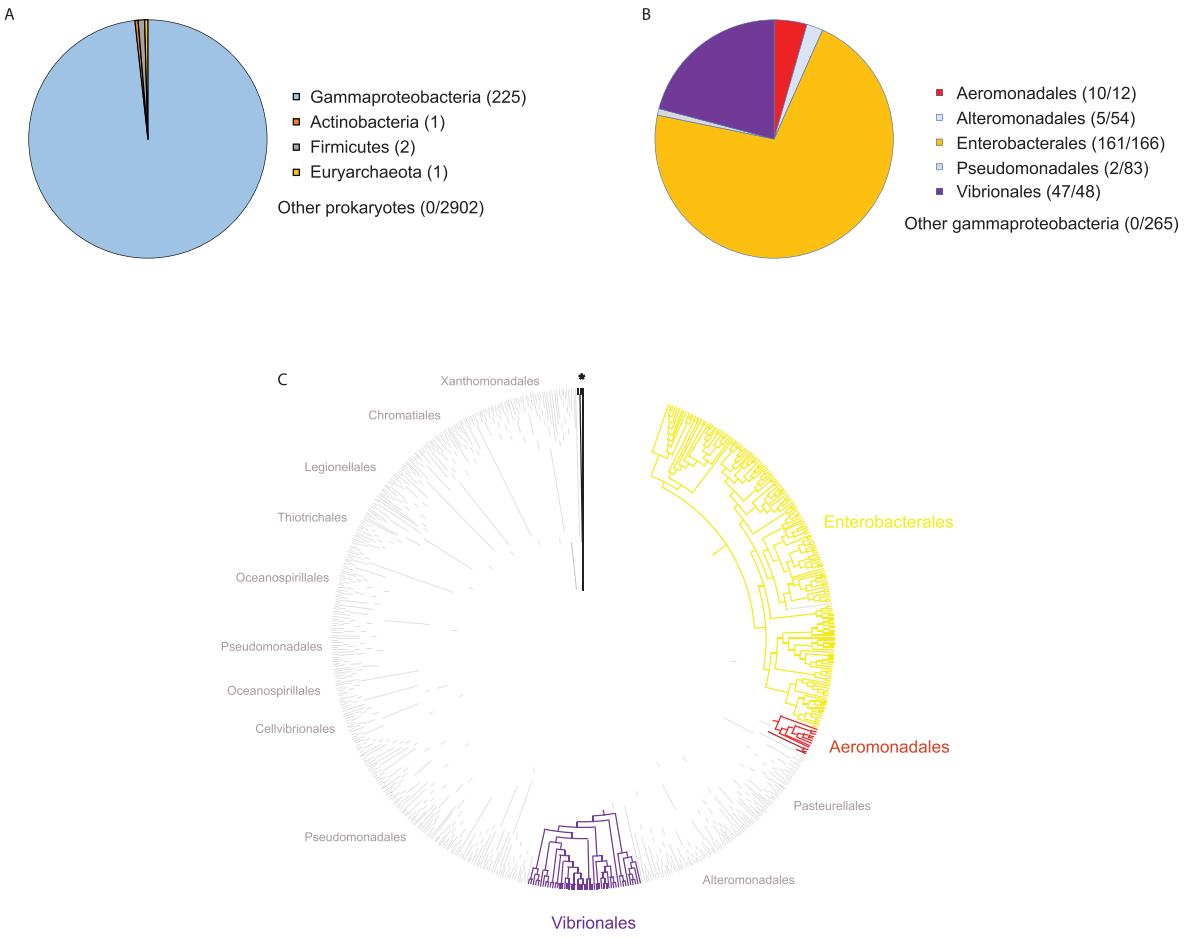


Figure 4. Presence of RavA-ViaA across prokaryotic species. (A) Pie chart depicting the abundance of species that encode for the RavA-ViaA pair in phyla across the prokaryotes. (B) The presence of RavA-ViaA across the order-level taxonomic groups of the Gamma-proteobacteria. (C) The 16S rRNA phylogenetic tree for the Gamma-proteobacteria highlights the three subgroups where RavA-ViaA are prevalent. The outlier (Beta-proteobacteria) species were used to root the tree and are indicated with (*).

positions. Additionally, the organisms tend to group together with other species of the same phylum.

Scatterplot representations of the pairwise cophyloplot data are shown in Figure 3. The axis values shown on the scatterplots are correlated to the sequence diversification and are a numerical representation of the cophylogenetic trees shown in Supporting Information Figure 1. For each MoxR-VWA partnership a representative scatterplot is provided in Figure 3. A root (or reference) species is placed at the origin (0,0) in the graph. The phylogenetic distances in the tree were plotted for all other species relative to the root. The MoxR diversification values are placed on the *X*-axis, and the values for VWA are on the *Y*-axis. While only a single plot is shown, 10 such scatterplots were constructed for each pairing.

There are some observations made salient by the graphs not evident in the cophyloplot. Primary among these is the linear evolutionary progression of MoxR-VWA protein partners. The coefficient of determination values for linear regression (R^2) for the plots are shown (Figure 3). The plots show a high-level fit with linear regression. This can be contrasted with cophyloplots of RavA or ViaA proteins in relation to 16S rRNA (Supporting Information Figure 2). 16S rRNA sequences were used as an example of a non-interacting sequence. When plotted with RavA or ViaA, the data show a poor fit to linear regression. As clarified by the contrast, the diversification of an MoxR protein sequence is reflected in the diversification of its VWA partner across the different species. This indicates a direct, linear sequence evolution for the MoxR and VWA partners across prokaryotes.

The linear associative pattern is consistently observed across the MoxR-VWA pairings. In addition to the *R*-squared values, the slopes for the linear line of regression are also listed (Figure 3). Many of the calculated slopes are greater than 1. This consistency across the MoxR groups is suggestive of the idea that the VWA proteins (*y*-axis) diversify at a greater rate than those of their MoxR partners (*x*-axis). Previous studies have often shown the VWA domain-containing proteins to act as adaptors in mediating protein−protein interactions for their MoxR partners.[10−12] It is intriguing to surmise that this adaptor role of VWA-containing proteins is the cause of the relatively rapid sequence diversity, and, hence, the VWA proteins adapt faster to the evolutionary pressures to mediate and support the function of MoxR proteins. It is notable that the slope and R^2 values are not drastically influenced based on the choice of the reference species. The standard deviation falls within 10% of the stated values for each.

Considering the CoxD-CoxE scatterplot as an illustrative example, *Mycobacterium tuberculosis* was set as the root species and is placed at the origin of the plot (Figure 3A). The plot fits a linear regression with a slope of ∼1.68 units. The slope shows that the average rate of sequence diversification in the CoxE (VWA) protein group is ∼68% greater than its CoxD (MoxR) partner group.

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

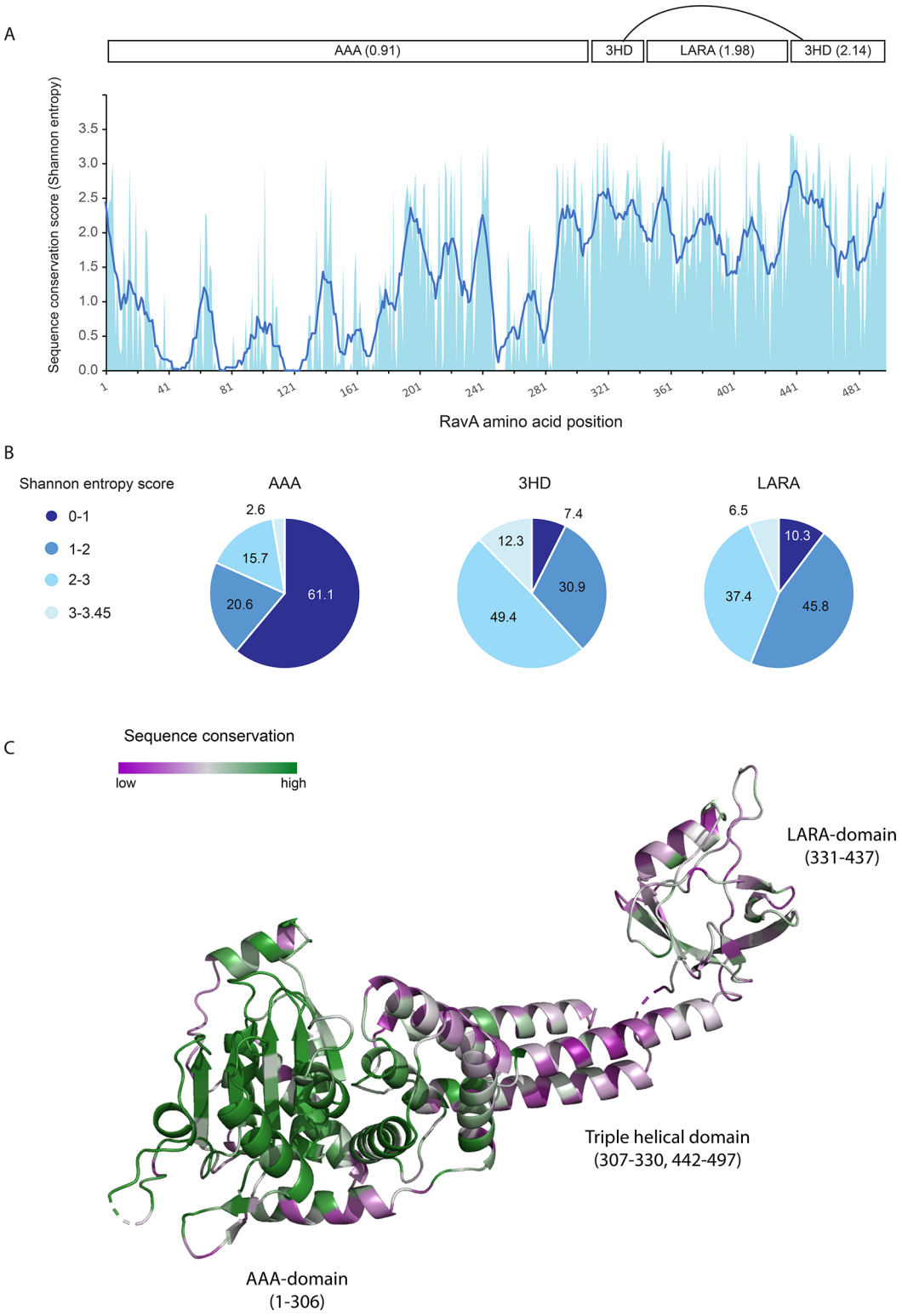


Figure 5. Variability and conservation along the RavA protein sequence. (A) A graph provides the variability (Shannon entropy) of each residue across the length of the RavA protein sequence. Residue positions are indicated in reference to the *E. coli* RavA protein. The domain layout is provided above the graph for reference. Shannon entropy scores are inversely related to conservation of the residue across species. (B) The division of variability scores for the residues belonging to the three domains of RavA are shown in the form of a pie chart. (C) The crystal structure of *E. coli* RavA protomer (PDB ID: 3NBX) is overlaid with variability scores from (A), providing a context for sequence conservation across the protein structure.

An interesting deviation from the general trend is seen in the PA2707-BPP3480 plot (Figure 3C). The relationship between the PA2707 group of MoxR proteins and the BPP3480 VWA proteins appears to be evolving at different rates in different species groups (Figure 3C). Apart from a few isolated exceptions, the protein pair is primarily found in species of the Proteobacteria and Firmicute phyla. In Proteobacteria, the phylogenetic slope is 1.53 units. Comparatively, the slope for

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

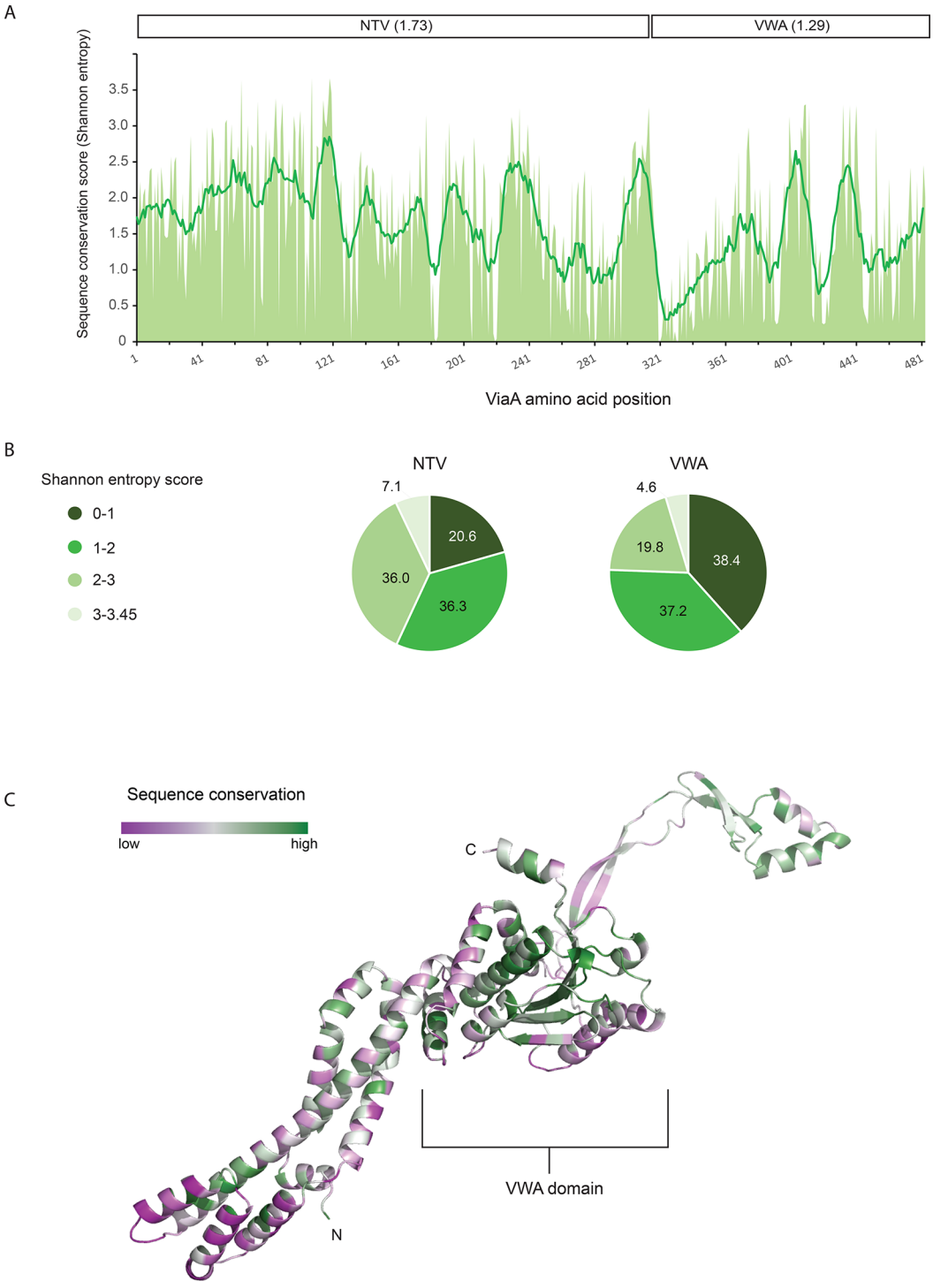


Figure 6. Variability along the ViaA protein sequence. (A) The variability of each residue across the length of the ViaA protein sequence. Residue numbers are indicated in reference to the *E. coli* ViaA protein. The domain layout is provided above the graph for reference. (B) Pie charts show the division of residues in different variability categories for the two domains of ViaA. (C) A *de novo* structure prediction of ViaA using Alphafold is overlaid with variability scores from (A).

Firmicute proteins is 0.46 units. This indicates that the diversification of the BPP3480 VWA protein relative to its MoxR partner PA2707 is noticeably lower in Firmicutes. Hence, there is a differentiation in the relationship between the BPP3480 and PA2707 as they evolve in different species groups. Perhaps, this could be an indication of the different functional roles they might play or a difference in the requirement for BPP3480 in the two phyla. Such deviations from expectations are instructive and provide a useful source for further analysis.

Determining the Presence of RavA-ViaA across Prokaryotic Species. The genomic proximity analysis (Table 2) and the cophyloplot in Figure 2 present a trend of coevolution for the MoxR and VWA-containing proteins. The sequence diversification data derived from the cophyloplot branch lengths, depicted through scatterplots in Figure 3, show

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

the VWA as diversifying faster than their MoxR counterparts. These analyses provide the trends defining the MoxR-VWA relationship. Next, an insight into the rate of evolution across protein sequences of MoxR and VWA partners was sought.

Because of the structural and functional insights available on the RavA and ViaA proteins from our group and others,[6,12−14,39] the RavA subfamily and the associated ViaA group of proteins were used here as representatives of MoxR and VWA, respectively. The conservation of RavA and ViaA sequences across prokaryotes was investigated. The protein sequences for RavA and ViaA from *E. coli* were used as queries. Protein homologues to each query were searched for in all prokaryotic species with publicly available complete genomic sequences. The AAA+ domain is conserved across the MoxR family and not exclusive to RavA proteins; a similar case is true for the VWA domain-containing proteins. Thus, a sequence similarity threshold of 50% across 70% of the sequence length was set as the cutoff when searching for orthologues.

A total of 233 sequences similar to RavA were identified and 239 orthologues for ViaA (Supporting Information Table 2). Expectedly, RavA- and ViaA-like sequences were found together across the species with few exceptions. A total of 225 species was identified with both proteins. The species spread for the ViaA and RavA pair is presented in Figure 4A,B and Supporting Information Table 2.

It is notable that more than 95% of the species harboring the RavA-ViaA duo belong to three order-level groupings of the Gram-negative gamma-proteobacteria: the Enterobacterales, Vibrionales, and Aeromonadales. The presence of the two proteins is almost universal among the Enterobacterales and the Vibrionales groups. Of the 166 species Enterobacterales examined, 162 had RavA and ViaA orthologues. Similarly, among the Vibrionales, 47 of 48 species encode the RavA-ViaA pair. Among the Aeromonadales, 10 of 12 species were identified with sequences for RavA-ViaA. Among all other gamma-proteobacteria, only 7 of 392 species code for RavA-ViaA. Outside the gamma-proteobacteria, RavA and ViaA are found in only 6 and 14 other organisms, respectively.

A 16S rRNA-based phylogenetic tree for the gamma-proteobacteria is presented in Figure 4C. The tree contains the 618 gamma-proteobacteria for which genomic data were available. The branching in the tree is representative of the evolutionary relationships among the species. Notably, the RavA-ViaA coding groups—the Enterobacterales, the Vibrionales, and the Aeromonadales—branch in proximity to each other. Thus, the RavA-ViaA proteins appear to be a feature of this niche within the gamma-proteobacteria.

Diversification of RavA and ViaA Protein Sequences across Prokaryotes. In Figure 1, several species outside of gamma-proteobacteria were shown to carry proteins belonging to the RavA subgroup of MoxR and ViaA subgroup of VWA proteins. The subsequent sequence similarity searches, presented in Figure 4, show that RavA and ViaA are conserved primarily across three closely related taxa in the gamma-proteobacteria. The difference in characterization is a consequence of the methodologies used. The criteria used for Figure 1 relied on phylogenetic analyses. The subsequent analysis performed for Figure 4 and Supporting Information Table 2 is based on a more stringent selection criteria; it was focused on the conservation of the sequences beyond the AAA+ and VWA domains of RavA and ViaA, respectively. As a result, the two methods highlight an interesting aspect of divergent evolution for our proteins of interest. Utilization of the stringent methodology showed that the proteins similar to *E. coli* RavA and ViaA are limited to the Enterobacterales, Vibrionales, and Aeromonadales (Figure 4B). When similarity is limited to the AAA+ domain for RavA and the VWA domain for ViaA, a wider set of organisms presents positive hits (Supporting Information Table 2).

To contextualize the diversification of RavA and ViaA sequences, a protein variability analysis was performed. For this, 225 homologous RavA (Figure 5) or ViaA (Figure 6) sequences were aligned. The alignments were subjected to the Shannon entropy analysis that calculated the diversity of each amino acid in the multiple sequence alignment (see Methods). A lower value denotes less variation and, thus, higher conservation. Values range from 0, complete conservation, to 4.32 if all residues are equally represented at the position. The average residue conservation score for each domain of RavA and ViaA is shown (Figure 5A and Figure 6A, respectively). The relative conservation of residues in each domain is shown in the accompanying pie chart (Figures 5B and 6B).

Considering RavA proteins in Figure 5, the AAA+ domain contains many regions of high sequence conservation. The AAA + domain is considerably less varied than the triple helical domain (3HD) or the LARA domain (Figure 5). This is highlighted by the variability scores overlaid upon the solved RavA structure (Figure 5C). For ViaA proteins (Figure 6), the VWA domain is the most conserved, with greater variability seen for the N-terminal domain.

It is plausible that the functional roles for the proteins evolve according to the divergence of their sequence. This is consistent with the view of an AAA+ domain as a conserved modular segment, providing the protein with the ability to harness the energy obtained from ATP-hydrolysis. In such a case, the remainder of the protein's sequence can evolve to endow the protein with a functionality required within its cellular environment. A similar assessment can be made for the ViaA protein with its VWA domain as the most conserved region within its sequence. This implicates genetic drift as a candidate in inducing the diversity of MoxR-VWA proteins. Genetic drift is a term used to describe the emergence of gene variants produced due to chance. Over a long enough period, or other causal factors, the accumulation of subtle shifts in a gene can manifest into drastic changes. It is perhaps such forces that have produced the currently observed conservation of RavA-ViaA: where the entire protein sequence for RavA-ViaA is conserved in a subset of gamma-proteobacteria while the RavA- and ViaA-like proteins in distant species retain similarity only across a portion of the protein sequence.

Another point of interest in this regard is the scattered presence of RavA-ViaA-like proteins in disparately linked organisms. For example, from a total of over 478 species of the Actinobacteria whose genomes were available, only *M. tuberculosis* was detected to encode for RavA and ViaA similar to the *E. coli* protein. Similarly, among the large Firmicutes group, only *Staphylococcus aureus* and *Streptococcus pneumoniae* contained these proteins. This observation is especially intriguing when considering that other species belonging to the *Staphylococcus*, *Streptococcus*, or *Mycobacterium* genera lack the genes for these proteins. The maintenance of RavA and ViaA in these pathogenic species points to the proteins as evolutionarily advantageous. However, a lack of functional knowledge on RavA-ViaA outside of *E. coli* limits our ability to infer functional utility of these proteins as their sequences diverge. The arrangement of evolutionary dispersal observed

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

here, along with the deviations against the trend, provide fertile ground for future studies on the function and evolution of RavA-ViaA pair.

CONCLUSION

MoxR proteins are well-represented in prokaryotic species yet remain a poorly characterized group. Previous microbiological and biochemical studies had identified a functional link between the MoxR ATPases and VWA domain-containing proteins. The partnerships that these protein groups share have been presented as an effector-adaptor relationship. However, such experimental data are limited to only a few examples. To expand on the available information, we employed bioinformatics approaches. Using protein sequences and genomic data, we have identified that the MoxR ATPases are widely linked to VWA domain-containing proteins. Many of these protein groups show genetic and phylogenetic linkages across varied and numerous prokaryotic species.

The identified MoxR proteins are delineated into nine major groupings or subfamilies (Figure 1). While the association between VWA domain-containing proteins and MoxR proteins is not universal, members from seven of the nine MoxR subfamilies were shown here to be linked to VWA domain-containing proteins. This association was identified phylogenetically, verified by genetic proximity and the cophyloplot analysis (Figures 1 and 2; Table 2, Supporting Information Table 1).

Cophyloplot trees and pairwise distance graphs provide further insight into the relationship between the proteins (Figure 3). They show a linearity in sequence evolution that might be expected. Interestingly, they show that the VWA-containing proteins appear to diversify at a greater rate than their MoxR counterparts (Figure 3). This rate, or diversity, is consistent with an interpretation of the binary MoxR-VWA relationship as an "effector and adaptor" pairing. The model posits that the VWA domain-containing proteins act as adaptors. VWA proteins assist in mediating protein interactions of their MoxR partner with the substrate on which they act. On the basis of this model and their observed evolutionary relationships shown here, it may be inferred that the VWA domain-containing proteins diversify in order to adapt the MoxR-VWA system to their functional roles in different organisms. As VWA domain proteins have been observed to function as mediators of protein−protein interactions, the faster rate of sequence divergence across VWA protein homologues may have allowed for their MoxR partners to adapt their functional roles in the different species as they diverge from a common ancestor.

As mentioned above, seven of nine MoxR subfamilies are genetically linked to VWA proteins. Currently, representatives from four of these groups, namely, the NorQ/CbbQ, CoxD, RavA, and MRP subfamilies of MoxR, have been functionally related to VWA proteins.[6,10−12,14,38,40,41] Yet, considering the size of the MoxR and their links to their VWA partners, it is remarkable that little functional data have been explored. Further research in this area should provide compelling insights into the functional importance of the novel group of ATPases and the partnership that they share with VWA domain-containing proteins.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.2c02554.

Supporting Information Figure 1. Comparative branching of proteins in each MoxR-VWA grouping. Cophyloplot representations for each pair of linked MoxR and ViaA subgroup are shown as indicated (A-I). Lines connect the position of proteins from the same species in the two phylogenetic trees that compose a cophyloplot. The lines are colored to represent the phylum of the species from which the protein sequence was obtained. The legend next to each plot identifies the phylum and its color. (PDF)

Supporting Information Figure 2. Comparative branching of 16S rRNA in relation to ViaA or RavA. The cophyloplot comparing species branching in a ViaA protein tree and 16S rRNA tree is shown in (A), a cophyloplot for RavA proteins and 16S rRNA is shown in (B). Lines connect the same species across the two phylogenetic trees of a cophyloplot. Pairwise distance matrices obtained from the cophyloplots are represented as scatterplots in (C) and (D). For the scatterplots, the sequence diversity for the ViaA or RavA is presented on the *Y*-axis and the 16S rRNA on the *X*-axis. A total of 10 scatterplots for each cophyloplot partnership were constructed, one of these is shown here as a representative. Protein sequences from one organism were used as a reference for the calculation of each scatterplot. This is referred to as the root. Each circular dot represents a single species in relation to the root. The colors of the dots indicate the taxonomic group of the organism. The average slope and average linear fit, from the 10 constructed scatterplots, are noted in the representative plot. (PDF)

Supporting Information Table 1. Genomic pairing of MoxR and VWA proteins. The complete list of identified genetic pairs for the MoxR and VWA is presented. The data identifies the proteins MoxR and VWA found next to each other. The UniProt identifier is provided. The subfamily grouping for the MoxR and VWA is indicated. The grouping is based on the branching of the identified protein in the phylogenetic tree (Figure 1). The data presented here is summarized in Table 2. (XLSX)

Supporting Information Table 2. Summary of homologous sequences discovered for RavA and ViaA. Two segments of the table highlight the presence of RavA- and ViaA-like sequences across the genomes. The section on the left lists the species containing proteins sharing sequence similarity to RavA or ViaA across 50% of their sequence. The section on the right is more stringent in its identification of similarity; it identifies the number of proteins sharing similarity to RavA or ViaA across more than 70% of their sequence. (XLSX)

AUTHOR INFORMATION

Corresponding Author

Walid A. Houry − *Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1M1, Canada; Department of Chemistry, University of Toronto, Toronto, Ontario M5S 3H6, Canada;* orcid.org/0000-0002-1861-3441; Phone: (416) 946-7141; Email: walid.houry@utoronto.ca; Fax: (416) 978-8548

Authors

Vaibhav Bhandari − *Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1M1, Canada*

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746

David A. J. Van Ommen − *Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1M1, Canada*
Keith S. Wong − *Department of Biochemistry, University of Toronto, Toronto, Ontario M5G 1M1, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.2c02554

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

V.B. is the recipient of Jaro Sodek Award−Ontario Student Opportunity Trust Fund fellowship from the Department of Biochemistry at the University of Toronto, Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Scholarship-Doctoral award, and Ontario Graduate Scholarship. D.A.J.V.O. is the recipient of NSERC Undergraduate Student Research Award. This work was funded by the Canadian Institutes of Health Research project grant (MOP-130374) to W.A.H.

■ REFERENCES

(1) Kunau, W. H.; Beyer, A.; Franken, T.; Gotte, K.; Marzioch, M.; Saidowsky, J.; Skaletz-Rorowski, A.; Wiebel, F. F. Two complementary approaches to study peroxisome biogenesis in Saccharomyces cerevisiae: forward and reversed genetics. *Biochimie* **1993**, *75* (3−4), 209−224.
(2) Neuwald, A. F.; Aravind, L.; Spouge, J. L.; Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **1999**, *9* (1), 27−43.
(3) Ogura, T.; Wilkinson, A. J. AAA+ superfamily ATPases: common structure–diverse function. *Genes Cells* **2001**, *6* (7), 575−597.
(4) Iyer, L. M.; Leipe, D. D.; Koonin, E. V.; Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct Biol.* **2004**, *146* (1−2), 11−31.
(5) Snider, J.; Houry, W. A. MoxR AAA+ ATPases: a novel family of molecular chaperones? *J. Struct Biol.* **2006**, *156* (1), 200−209.
(6) Snider, J.; Gutsche, I.; Lin, M.; Baby, S.; Cox, B.; Butland, G.; Greenblatt, J.; Emili, A.; Houry, W. A. Formation of a distinctive complex between the inducible bacterial lysine decarboxylase and a novel AAA+ ATPase. *J. Biol. Chem.* **2006**, *281* (3), 1532−1546.
(7) Yanai, I.; Mellor, J. C.; DeLisi, C. Identifying functional links between genes using conserved chromosomal proximity. *Trends Genet* **2002**, *18* (4), 176−179.
(8) Whittaker, C. A.; Hynes, R. O. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol. Biol. Cell* **2002**, *13* (10), 3369−3387.
(9) Xiong, J. P.; Stehle, T.; Zhang, R.; Joachimiak, A.; Frech, M.; Goodman, S. L.; Arnaout, M. A. Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* **2002**, *296* (5565), 151−155.
(10) Tsai, Y. C.; Ye, F.; Liew, L.; Liu, D.; Bhushan, S.; Gao, Y. G.; Mueller-Cajar, O. Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (1), 381−387.
(11) Kahle, M.; Ter Beek, J.; Hosler, J. P.; Adelroth, P. The insertion of the non-heme FeB cofactor into nitric oxide reductase from P. denitrificans depends on NorQ and NorD accessory proteins. *Biochim Biophys Acta Bioenerg* **2018**, *1859* (10), 1051−1058.
(12) Wong, K. S.; Bhandari, V.; Janga, S. C.; Houry, W. A. The RavA-ViaA Chaperone-Like System Interacts with and Modulates the Activity of the Fumarate Reductase Respiratory Complex. *J. Mol. Biol.* **2017**, *429* (2), 324−344.
(13) El Bakkouri, M.; Gutsche, I.; Kanjee, U.; Zhao, B.; Yu, M.; Goret, G.; Schoehn, G.; Burmeister, W. P.; Houry, W. A. Structure of RavA MoxR AAA+ protein reveals the design principles of a molecular cage modulating the inducible lysine decarboxylase activity. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (52), 22499−22504.
(14) Wong, K. S.; Snider, J. D.; Graham, C.; Greenblatt, J. F.; Emili, A.; Babu, M.; Houry, W. A. The MoxR ATPase RavA and its cofactor ViaA interact with the NADH:ubiquinone oxidoreductase I in Escherichia coli. *PLoS One* **2014**, *9* (1), No. e85529.
(15) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2004**, *33*, D501−504.
(16) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389−3402.
(17) Bateman, A.; et al. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49* (D1), D480−D489.
(18) Suzek, B. E.; Huang, H.; McGarvey, P.; Mazumder, R.; Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **2007**, *23* (10), 1282−1288.
(19) Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32* (5), 1792−1797.
(20) Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30* (12), 2725−2729.
(21) Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18* (5), 691−699.
(22) Tatusova, T. A.; Karsch-Mizrachi, I.; Ostell, J. A. Complete genomes in WWW Entrez: data representation and analysis. *Bioinformatics* **1999**, *15* (7−8), 536−543.
(23) Tatusova, T.; Ciufo, S.; Federhen, S.; Fedorov, B.; McVeigh, R.; O'Neill, K.; Tolstoy, I.; Zaslavsky, L. Update on RefSeq microbial genomes resources. *Nucleic Acids Res.* **2015**, *43*, D599−D605.
(24) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215* (3), 403−410.
(25) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35* (6), 1547−1549.
(26) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22* (22), 4673−4680.
(27) Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59* (3), 307−321.
(28) R Core Team. R: A language and environment for statistical computing. *R Project;* R Foundation for Statistical Computing: Vienna, Austria, 2013.
(29) Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **2004**, *20* (2), 289−290.
(30) Garcia-Boronat, M.; Diez-Rivero, C. M.; Reinherz, E. L.; Reche, P. A. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.* **2008**, *36*, W35−W41.
(31) Shannon, C. E. A mathematical theory of communication. *Bell system technical journal* **1948**, *27* (3), 379−423.
(32) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44* (D1), D733−745.
(33) Morgulis, A.; Coulouris, G.; Raytselis, Y.; Madden, T. L.; Agarwala, R.; Schaffer, A. A. Database indexing for production MegaBLAST searches. *Bioinformatics* **2008**, *24* (16), 1757−1764.

(34) Mirdita, M.; Steinegger, M.; Soding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **2019**, *35* (16), 2856−2858.
(35) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(36) Scheele, U.; Erdmann, S.; Ungewickell, E. J.; Felisberto-Rodrigues, C.; Ortiz-Lombardia, M.; Garrett, R. A. Chaperone role for proteins p618 and p892 in the extracellular tail development of Acidianus two-tailed virus. *J. Virol* **2011**, *85* (10), 4812−4821.
(37) Happonen, L. J.; Erdmann, S.; Garrett, R. A.; Butcher, S. J. Adenosine triphosphatases of thermophilic archaeal double-stranded DNA viruses. *Cell Biosci* **2014**, *4*, 37.
(38) Morris, C. J.; Kim, Y. M.; Perkins, K. E.; Lidstrom, M. E. Identification and nucleotide sequences of mxaA, mxaC, mxaK, mxaL, and mxaD genes from Methylobacterium extorquens AM1. *J. Bacteriol.* **1995**, *177* (23), 6825−6831.
(39) Malet, H.; Liu, K.; El Bakkouri, M.; Chan, S. W.; Effantin, G.; Bacia, M.; Houry, W. A.; Gutsche, I. Assembly principles of a unique cage formed by hexameric and decameric E. coli proteins. *Elife* **2014**, *3*, No. e03653.
(40) Pelzmann, A. M.; Mickoleit, F.; Meyer, O. Insights into the posttranslational assembly of the Mo-, S- and Cu-containing cluster in the active site of CO dehydrogenase of Oligotropha carboxidovorans. *J. Biol. Inorg. Chem.* **2014**, *19* (8), 1399−1414.
(41) Maisel, T.; Joseph, S.; Mielke, T.; Bürger, J.; Schwarzinger, S.; Meyer, O. The CoxD protein, a novel AAA+ ATPase involved in metal cluster assembly: hydrolysis of nucleotide-triphosphates and oligomerization. *PLoS One* **2012**, *7* (10), No. e47424.

https://doi.org/10.1021/acs.jpca.2c02554
J. Phys. Chem. A 2022, 126, 4734−4746