BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

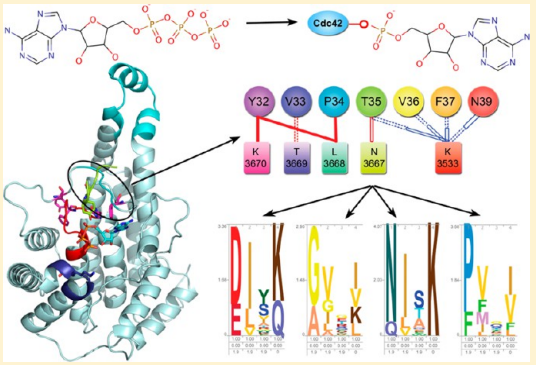
Deciphering the Molecular Basis of Functional Divergence in AMPylating Enzymes by Molecular Dynamics Simulations and Structure Guided Phylogeny

Shradha Khater and Debasisa Mohanty*

Bioinformatics Center, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi 110067, India

Supporting Information

ABSTRACT: The Fic domain was recently shown to catalyze AMPylation—the transfer of AMP from ATP to hydroxyl side chains of diverse eukaryotic proteins, ranging from RhoGTPases to chaperon BiP. We have carried out a series of explicit solvent molecular dynamics (MD) simulations up to 1 μs duration on the apo, holo, and substrate/product bound IbpA Fic domain (IbpAFic2). Simulations on holo-IbpAFic2 revealed that binding of Mg^{2+} to α and β phosphates is crucial for preserving catalytically important contacts involving ATP. Comparative analysis of the MD trajectories demonstrated how binding of ATP allosterically induces conformational changes in the distal switch II binding region of Fic domains thereby aiding in substrate recognition. Our simulations have also identified crucial aromatic−aromatic interactions which stabilize the orientation of the catalytic histidine for inline nucleophilic attack during AMPylation, thus providing a structural basis for the evolutionary conservation of these aromatic residue pairs in Fic domains. On the basis of analysis of interacting interface residue pairs that persist over the microsecond trajectory, we identified a tetrapeptide stretch involved in substrate recognition. The structure-based genome-wide search revealed a distinct conservation pattern for this segment in different Fic subfamilies, further supporting its proposed role in substrate recognition. In addition, combined use of simulations and phylogenetic analysis has helped in the discovery of a new subfamily of Fic proteins that harbor a conserved Lys/Arg in place of the inhibitory Glu of the regulatory helix. We propose the novel possibility of auto-enhancement of AMPylation activity in this new subfamily via the movement of regulatory helix, in contrast to auto-inhibition seen in most Fic proteins.



AMPylation, also known as adenylylation, is a novel post-translational modification (PTM) involving the covalent transfer of the AMP moiety from ATP to the Thr/Tyr side chain of a substrate protein.[1] Though AMPylation of bacterial proteins has been known for four decades,[2] AMPylation of eukaryotic proteins was only recently discovered.[3,4] AMPylation of the switch I region of RhoGTPases by IbpA and VopS Fic (filamentation induced by cAMP) domains, which disrupt its downstream signaling causing actin cytoskeleton collapse, has been implicated in pathogenesis of mammalian hosts.[3−5] Though the transfer of the AMP group by other Fic domains has also been implicated in pathogenesis,[4,6,7] their biological role is not limited to pathogenesis alone.[8] The Fic family covers a large number of proteins (Pfam database 6829) and is characterized by the presence of HXFX(D/E)(A/G)N(G/K)R motif.[9,10] Even though Fic domains share little sequence similarity outside the conserved motif,[10] they share a conserved alpha helical fold.[11] Death on curing (Doc) proteins, the toxin part of toxin−antitoxin module of *Bacteriophage P1* and avirulence protein B (AvrB), the effector protein of *Pseudomonas syringae* type III also share the same α-helical fold as Fic domains (Supplementary Figure S1).[12,13] While Doc proteins share sequence similarity with the Fic family and have been classified together in the Pfam database,[9] AvrB does not share any significant sequence similarity with Fic/Doc proteins.[11] Given the unique α-helical topology shared by Fic, Doc, and AvrB proteins, they have been grouped together as a single superfamily called FiDo.[11]

Structural and biochemical analysis has helped in the elucidation of the catalytic mechanism of Fic domain containing proteins and has provided valuable insights into the ligand and substrate binding by Fic domains.[6,7,13−18] Fic domains have been shown to catalyze diverse functions like AMPylation, phosphorylation, and phosphocholination (Supplementary Figure S2). The high sequence divergence of Fic domains is reflected in the diversity of its catalytic function as well as plasticity in substrate recognition.[10] Though the majority of *in vivo* characterized substrates of the Fic domain are RhoGTPase, other proteins like RIPK and vimentin have also been shown to be its targets by direct *in vitro* assays (Supplementary Table S1). AvrB too has been implicated in phosphorylating Rin4, a non-GTPase protein.[19,20] IbpA and

Received: April 2, 2015
Revised: August 1, 2015
Published: August 7, 2015

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

AvrB are the only two Fido domain containing proteins, where structures with their targets/substrates have been solved. IbpA, which consists of 4095 residues, contains two Fic domains (Fic1 and Fic2) in its C-terminal. The structure of the second Fic domain of IbpA (IbpAFic2), corresponding to residues 3482−3797, has been elucidated with and without AMPylated Cdc42.[18] The N-terminal helices α1-α5 form a unique arm domain, whereas α8-α14 forms the core Fic fold (Figure 1). A

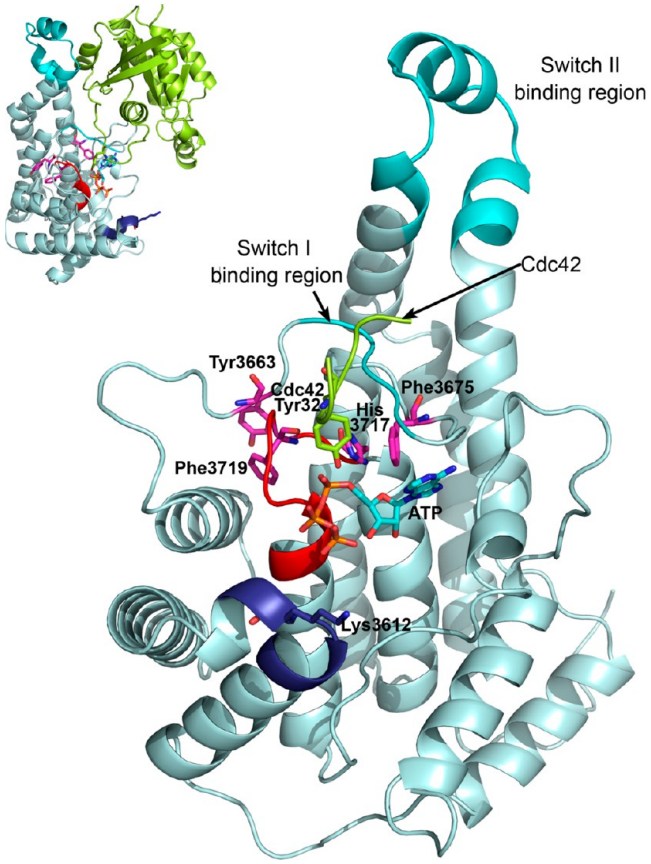


Figure 1. IbpAFic2 in complex with Cdc42. Cdc42 has been represented in green (for clarity only the switch I region is shown). Switch I and switch II regions and ATP (transposed from VbhT crystal structure 3ZCB) have been marked in cyan. Conserved Fic motif has been highlighted in red. Two π-interactions between a pair of residues are in pink and are conserved across Fido proteins. The modified inhibitory motif of IbpAFic2 is in deep blue. H3717 is mutated to alanine in crystal structure 3N3V. Hence, for depiction the histidine was transformed from crystal structure 3N3U. The inset shows IbpAFic2 (light blue) in complex with Cdc42.

similar arm domain is known to be present only in VopS and AvrB proteins.[12,17] The IbpAFic2-Cdc42 structure (PDB ID: 3N3V, superseded by 4ITR) shows conserved switch I and switch II regions of Cdc42 to be the major contributors to the Fic−Cdc42 interfaces (Figure 1). The N-terminal arm domain interacts with the switch II region, whereas the switch I region interacts with the α9−α10 loop of the Fic domain. This loop is conserved as a β-hairpin loop in other Fic domain containing proteins, except Doc,[11] and is also known to interact with various ligands.[6,16] Even AvrB, which shares no sequence similarity with Fic proteins, has the β hairpin loop conserved, and it interacts with the nucleotide ADP.[12] This β hairpin loop forms the interface of AvrB and Rin4 peptide. To help recognize other substrates of the Fido domain, a deeper understanding of the substrate−enzyme interaction through the β-hairpin loop is required. These structures of Fido domain with their targets can be used to understand the interactions in greater detail. Recently, it was shown that the activity of Fic proteins is regulated by inter- or intradomain helices.[6] These helices contain a conserved inhibitory glutamate residue that can obstruct the ATP binding and hence can abrogate the AMPylation reaction (Supplementary Figure S3). This inhibitory glutamate is usually present in an N-terminal or a C-terminal helix of Fic domain or on a protein encoded by a gene in the same genomic neighborhood. Analysis by novPTMenzy web server indicates that IbpA and VopS are the only two Fic domain containing virulent factors lacking the inhibitory glutamate on the regulatory helix.[21] The absence of inhibitory glutamate makes Vops and IbpA highly lethal. Hence, the structures of IbpAFic2 can also be used to analyze its unregulated catalytic efficiency.

In this study, we have carried out detailed structural analysis of substrate-free and substrate-bound crystal structures of IbpAFic2 domain to identify the functionally important residues that play a crucial role in catalysis and substrate recognition. In order to understand the role of liagands and cofactors in catalysis, we have modeled holo structures of IbpAFic2 in complex with ATP and Mg^{2+} using docking and carried out explicit solvent molecular dynamics (MD) simulations on apo, holo, and substrate bound structures of IbpAFic2. Our simulations have not only identified the specificity determining residues in switch I binding region of Fic domain, but also revealed novel ligand-induced allosteric conformational changes in switch II binding region, another substrate interaction site. On the basis of these structural studies, we have also analyzed how various subfamilies of AMPylation domains like Fic, Doc, AvrB, and ankyrin repeat-containing protein X (AnkX) catalyze a diverse set of reactions (Supplementary Figure S2) using the same Fido fold. Even though identification of functionally important residues was based on simulations on a representative Fic structure, IbpAFic2, we have carried out phylogenetic analysis to explore the conservation profile of these residues across various organisms. Our structure guided phylogenetic analysis has established the relevance of these functionally important residues for Fic domains in general. Our analysis revealed that all interacting aromatic−cyclic residue pairs are coevolving, suggesting they might play a crucial role in Fido domain catalysis. On the basis of our structural and evolutionary analysis, we have also identified a novel class of Fic domains which have positively charged arginine or lysine residue in place of inhibitory glutamate on the regulatory helix.

MATERIALS AND METHODS

Modeling of Holo-AMPylation Domain and Starting Structures for MD Simulations. Crystal structures of AMPylation domain of IbpA (IbpAFic2) were available for apo enzyme (PDB ID: 3N3U, residues 3488−3786) and enzyme substrate complex (PDB ID: 3N3V superseded by 4ITR). The enzyme substrate complex consisted of a product bound form where IbpAFic2 was bound to Cdc42 AMPylated at Tyr32. Therefore, the structure for ATP and Mg^{2+} bound holo form of AMPylation domain had to be modeled using the apo form of IbpAFic2 (IbpA_A), i.e., 3N3U. Three different models were built. For the first model (IbpA_H1) only ATP was docked using AutoDock4[22] onto apo structure 3N3U, keeping active site residues Glu3721, Asn 3723, and Arg 3725

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

flexible during docking. Blind docking was carried out using a docking grid which encompassed the entire structure of IbpA_A, and a representative structure from the highly populated low energy cluster was chosen as the initial structure for the IbpA_H1 system. For the second model (IbpA_H2), magnesium ion (Mg^{2+}) was manually docked onto IbpA_H1 such that β phosphate, γ phosphate, and Glu3721 were at an appropriate distance to coordinate the Mg^{2+} ion. After the availability of the crystal structure (PDB ID: 3ZCB) for the holo form of VbhA Fic domain, the third model (IbpA_H3) was built by transforming coordinates of ATP and Mg^{2+} from 3ZCB to 3N3U after optimum superposition of both the AMPylation domains. The major difference between the models IbpA_H2 and IbpA_H3 was that the magnesium ion was coordinated by β and γ phosphates in the former, while it was coordinated by α and β phosphates in the latter. In both cases, Glu3721 also coordinated the Mg^{2+} ion. Coordinates for the residue stretch 3668−3672 in the switch I binding loop were not resolved for the structure 3N3U, while the coordinates for this segment was available in the product bound complex structure 3N3V. Hence, the coordinates of the loop were transformed from structure 3N3V to all three holo models as well as the apo structure IbpA_A. The structures IbpA_A, IbpA_H1, IbpA_H2, and IbpA_H3 were used as starting structures for MD simulations on apo and holo forms of the AMPylation domain. Chain A and D of the 2:2 heterotetrameric complex of the IbpAFic2−Cdc42 (PDB ID: 3N3V) crystal structure was used as the starting structure for MD simulations on the enzyme−substrate complex (IbpA:Cdc42). The AMPylated Cdc42 in IbpA:Cdc42 was also bound to GDP.

Molecular Dynamics Simulation and Analysis of MD Trajectories. In order to carry out explicit solvent MD simulations on apo, holo, and product bound forms of AMPylation domain, IbpA_A, IbpA_H1, IbpA_H2, IbpA_H3, and IbpA:Cdc42 were solvated in a rectangular box of TIP3P water molecules.[23] The box boundaries were set at 10 Å from the outermost atoms of the protein in all three directions. AMBER11 package[24] and ff03 force field[25] were used to carry out MD simulations. Force field parameters for ATP, GDP, and Mg^{2+} were based on previous studies[26,27] and were obtained from the AMBER Parameter Database maintained at http://www.pharmacy.manchester.ac.uk/bryce/amber/. AMPylated Tyr parameters were generated using InsightII software and CVFF force field.[28,29] SANDER module of the AMBER11 package was used to minimize the initial crystal or modeled structures using the steepest descent approach to an RMS gradient of 0.001 kcal/mol/Å. MD simulations were carried out on the minimized structure by the PMEMD module. A time step of 1 fs was used, and SHAKE[30] was used to constrain bonds involving hydrogen atoms. The particle mesh Ewald (PME)[31] approach was employed to compute long-range electrostatic interactions, and a cutoff of 10 Å was used in the direct space for nonbonded interactions. The system was equilibrated to a temperature of 300 K and density of g/cm^3 in two stages before starting the production run of MD. The system was simulated in NVT conditions increasing the temperature gradually from 100 K over 5 ps to the final temperature. To constrain the temperature of the system at a specified value, Langevin dynamics temperature coupling scheme was used with a collision frequency of 5 ps^{-1}. The second stage of equilibration process was done for 15 ns after the temperature of system stabilized at 300 K. The density of water was then adjusted to 1 g/cm^3. After the stabilization of the system, production runs of MD simulation were carried out in NPT ensemble using parallel CUDA version of particle mesh Ewald molecular dynamics (PMEMD) for durations ranging from 350 ns to 1 μs. MD simulations on the apo and holo structures IbpA_A and IbpA_H3 were carried out for 1 μs each, while the holo structures IbpA_H1 and IbpA_H2 were simulated for 350 ns each. The IbpA:Cdc42 complex was simulated for 500 ns (Table 1).

Table 1. Systems Chosen for Molecular Dynamics

simulation system	name of the system	description	time length of simulation (ns)
Apo IbpAFic2	IbpA_A	PDB ID: 3N3U	1000
Holo IbpAFic2	IbpA_H1	3N3U + docked ATP	350
	IbpA_H2	3N3U + docked ATP + manually docked Mg^{2+} on β and γ phosphate of ATP	350
	IbpA_H3	3N3U + ATP along with Mg^{2+} ion on $\alpha\beta$ phosphate transformed from VbhT structure 3ZCB	1000
IbpAFic2 in complex with Cdc42	IbpA:Cdc42	chain A and D of crystal structure 3N3V	500

Various measures like root-mean-square deviation (RMSD) with respect to the starting structure, residue wise theoretical B-factor, and the distance between functionally important residues were calculated using the ptraj module of AMBER 11. For these calculations snapshots were retrieved at 5 ps intervals from the entire MD trajectories. The B-factor was calculated from root mean square fluctuation (RMSF)[32] using the equation:

$$\text{B-factor} = \frac{8\pi^2}{3}\text{RMSF}^2$$

To calculate the shortest distance between any two residues or chemical moieties, an in-house Perl script was used to calculate the minimum distance between any two atoms belonging to these residues/moieties. In various figures depicting time evolution curves for RMSD and distance between interacting residues, running averages over 200 frames or 1 ns were computed and plotted to improve clarity. Running or moving average was calculated using the zoo package of R.[33] The original time evolution curves have also been provided as Supporting Information. Principal component analysis (PCA) was carried out for conformers sampled over different MD trajectories to study the major modes of movement in these structures. The eigenvectors from PCA analysis was visualized using IED plugin of VMD,[34] whereas other plots were visualized using the ggplot2 implementation of R.[35]

Clustering of Conformers Sampled during MD Simulations and Analysis of Interface Interactions. Clustering of conformers was carried out to understand the effect of dynamic fluctuations in structures on the interaction between IbpAFic2 and Cdc42. Snapshots at every 500 ps were extracted from the 500 ns trajectory of IbpA:Cdc42 using the ptraj module (1000 snapshots in total). These snapshots were clustered using the kclust module of the MMTSB tool set[36] based on their coordinate RMSD values. Thirty-two clusters

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

were obtained by fixing the clustering radius at 2 Å. Out of these 32 clusters, minor clusters containing less than 20 snapshots were removed. Conformers corresponding to centroids of remaining (16) clusters were analyzed for interface contacts. Interface contacts were computed using default parameters of Protein Interactions Calculator (PIC) server.[37] Distances between interacting residue pairs existing in more than five clusters were monitored over entire 500 ns of simulation. Those contacting residue pairs that persisted over 60% of the simulation time were identified.

Analysis of Genome-Wide Conservation of Crucial Residues Involved in Catalysis and Substrate Recognition. Even though sequence based methods like BLAST or more sensitive profile based methods like PSI-BLAST are critical for functional annotations of newly sequenced genes, these methods do not work well in the twilight region of sequence similarity. Therefore, incorporation of structural information, which is more conserved than sequences, can help in increasing the sensitivity of these sequence searches. HHPred[38] uses a structure based profile−profile alignments to search for remote homologues. Use of profiles for both query and target helps in incorporating evolutionary information and thus further increasing the sensitivity of alignment. In addition to remote homology searches, these structure based profile−profile alignments can also be used to align protein sequences with little sequence similarity or loop regions, which are usually not very well conserved. Therefore, for Fido domain proteins, which show a very high degree of sequence divergence, structure-based profile−profile alignment was used in the search for genome-wide conservation of crucial functional residues. A database of profiles using sequences of available structures of Fic domain (2F6S, 4BET,[14] 3DD7,[13] 3HSG,[6] 3N3V,[18] 3LET,[17] 3EQX,[15] 3CUC, 2VZA,[7] 2G03) was built. These ten profiles with structural information from the crystal structures served as the targets for profile−profile search. Conservation of various functional residues was analyzed across all Fic proteins sequences from Pfam Fic family (Pfam ID: PF02661).[9] Profiles were built for each Fic domain sequences using HHblits,[39] and structural information was incorporated in the profiles using predicted secondary structures.[40] Profiles for all Fic sequences were then aligned to structure profiles using HHSearch.[41] Corresponding to each of the sequence, 10 alignments were produced. These alignments were then probed for various functional residues.

Using the above-mentioned protocol the conservation of Asn3667−Lys3670 was studied across all Fic proteins sequences.[9] The HMM logo representing the conserved motif for this stretch was created using Skylign.[42] On the basis of the first residue of the motif, sequences were manually clustered into seven subfamilies. Using a similar structure based profile−profile comparison, alignments of all 6829 Fic proteins from the Pfam database[9] were used to analyze coevolution of H3717−F3675 and F3719−Y3663 pairs across various genomes. The conservation of these residues was represented as Venn diagrams.

Analysis of Sequence Conservation in Inhibitory Helices of Fic Domain. Glutamate, in SXXXE(G/N) inhibitory motif, is known to inhibit ATP binding in many Fic proteins.[6] To search for variations in the inhibitory motif that might aid in catalysis instead of inhibiting it, structure based profile−profile alignments were used. The protocol followed is same as stated above except templates used were 2F6S, 2G03, 3CUC, and 3EQX, for which the position of inhibitory glutamate is known.[6] The inhibitory glutamate can occur at N-terminal of Fic domains (SoFic −3EQX and BtFic −3CUC) or C-terminal of the Fic domain (NmFic −2G03 and HpFic −2F6S). Fic domains of the former type have been classified as class II and latter type as class III. Sequence stretches corresponding to inhibitory motif were extracted. The inhibitory motif was considered for further analysis only when both alignments of class II or class III gave the same motif. To check for a modified inhibitory motif only sequences with lysine or arginine in place of inhibitory glutamate were considered. The Fic active site motif was also extracted for those sequences where inhibitory glutamate has been replaced by arginine or lysine. These motifs were represented by HMM Logo,[42] and the quantification of their conservation was analyzed using a Venn diagram.

RESULTS

MD Simulations on Apo and Holo Forms of IbpAFic2 Reveal Role of ATP and Mg^{2+} in Catalysis. The structure of the apo form of IbpAFic2 and IbpAFic2 bound to AMPylated Cdc42 has been elucidated by crystallographic studies. However, biochemical studies have revealed that Fic domains require ATP and magnesium ion[3,4] for catalysis of the AMPylation reaction. Also, it has been shown that AnkX, a Fic domain protein catalyzing phosphocholination, requires Mg^{2+} ion for its catalytic activity.[14] In order to understand the role of Mg^{2+} and ATP in catalysis, as mentioned in the Methods section, we modeled various structures of IbpAFic2 and carried out explicit solvent MD simulations on apo (IbpA_A), holo (IbpA_H1, IbpA_H2 and IbpA_H3), and substrate bound (IbpA:Cdc42) forms of IbpAFic2 (Table 1). IbpA_H1 was modeled by blind docking of ATP onto IbpA_A (i.e., apo form crystal structure 3N3U) using a docking grid covering the entire surface. The IbpA_H2 structure was modeled by adding Mg^{2+} to IbpA_H1 such that Mg^{2+} ion was coordinated by Glu3721 and β and γ phosphates of ATP. The coordination of Mg^{2+} in IbpA_H2 was based on a reaction mechanism proposed by Loung et al.[17,18] for VopS Fic domain (Supplementary Figure S4). On the other hand, IbpA_H3 was modeled based on the holo form crystal structure of VbhA Fic domain[16] where the Mg^{2+} ion was coordinated by α and β phosphates of ATP. The aim was to understand if MD simulation studies can discriminate between these two different modes of coordination of the Mg^{2+} ion. Figure 2 shows conformations of ATP and Mg^{2+} in the IbpA_H1, IbpA_H2, and IbpA_H3 system. It is interesting to note that the binding pose of ATP (Figure 2A,B) obtained from docking in IbpA_H1 and IbpA_H2 is similar to the binding pose (Figure 2C) in IbpA_H3 which was modeled based on the VbhA crystal structure. These three structural models for the holo IbpAFic2 were refined by 350 ns explicit solvent MD simulations. Figure 3A shows variation of RMSD with respect to the initial structures over the 350 ns trajectories of IbpA_H1, IbpA_H2, and IbpA_H3. After 50 ns, the RMSD of each of these systems converged to approximately 3 Å, indicating their overall stability. Figure 3B shows computed B-factor plots, obtained from RMSF values, which indicate residue wise fluctuations from the average structure. As can be seen the major difference in the three systems is the flexibility in the switch II binding region. The high flexibility in this segment of IbpA_H3 correlates well with the experimental B-factors in the crystal structure of IbpAFic2 (PDB ID: 3N3U and 3N3V). The B-factor of ATP in the three systems also shows some differences

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

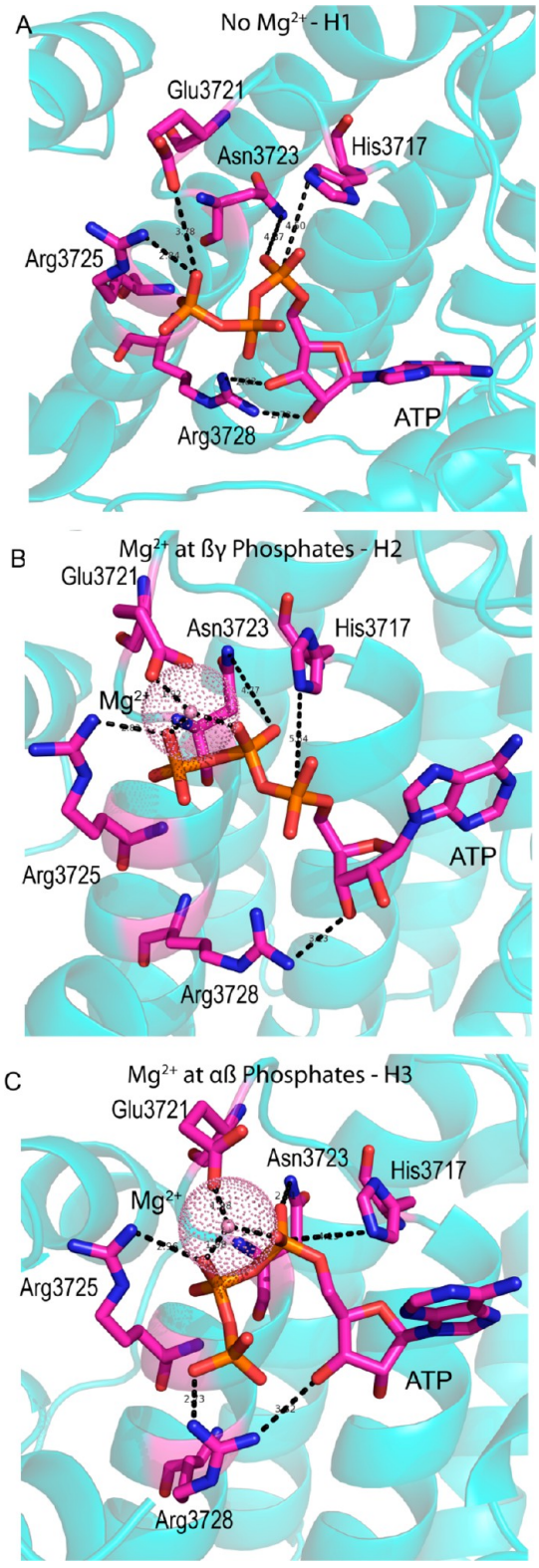


Figure 2. Comparison of the active site pockets of three IbpAFic2 + ATP systems. (A) IbpAFic2 + docked ATP (IbpA_H1). (B) IbpAFic2 + ATP + Mg^{2+} docked on β and γ phosphates of ATP (IbpA_H2). (C) IbpAFic2 + ATP + Mg^{2+} on α and β phosphates of ATP (IbpA_H3). The IbpAFic2 structure has been shown in cartoon representation, active site residues and ATP in stick representation and Mg ion as a nonbonded sphere.

(Figure 3C). High flexibility of ATP ligand is seen in the IbpA_H2 system, while flexibility of ATP in IbpA_H1 and IbpA_H3 is relatively lower. Several residues surrounding the catalytic pocket (Figure 2 and Supplementary Figure S4) have been shown to be important for AMPylation activity in IbpA and other Fic proteins.[17,18] We analyzed the distances of these critical active site residues from ATP over the MD trajectory (Figure 3D). As can be seen, only IbpA_H3 had all the important contacts preserved over the entire length of the simulation, indicating Mg^{2+} coordination by α and β phosphates to be crucial for AMPylation reaction. Therefore, we have used the IbpA_H3 holo system for further analysis.

Essential Dynamics of Apo, Holo, and Substrate Bound IbpAFic2. MD simulations for IbpA_A and IbpA_H3 were extended until 1 μs, while the substrate bound complex IbpA:Cdc42 was simulated for 500 ns. The convergence of these three simulations was monitored through RMSD analysis (Figure 4A), and the flexibilities of different regions were compared using the residue wise computational B-factor plots (Figure 4B). As can be seen, the computed B-factor for the switch II binding region (residues 3527−3560) is very high compared to other regions of the IbpAFic2. Also, the B-factor is the highest for the holo form, IbpA_H3, and lowest for substrate bound form, IbpA:Cdc42. In addition to B-factor plots, PCA was also performed on these three MD trajectories to identify major modes of movements in apo, holo, and substrate bound structures of IbpAFic2 domain. Figure 4C−E depicts the eigenvectors from PCA analysis of the three MD trajectories. Interestingly, PCA of MD trajectories revealed that the movement of the switch II binding region of apo form (IbpA_A) was away from the substrate binding region (Figure 4C), whereas that of the holo form (IbpA_H3) was toward the substrate binding region (Figure 4D). Also, the displacement was higher in the case of the holo form compared to the apo form. The IbpAFic2−Cdc42 complex showed movements of the switch II binding region toward the substrate similar to the holo form (IbpA_H3), but the magnitude of movement was comparatively lower. Contacts of the switch II binding region in the holo form (IbpA_H3) were analyzed in greater detail to understand the molecular basis of flexibility in the arm domain of IbpAFic2. The analysis revealed that Asn3667 of the switch I binding region is in contact with either Ser 3534 or Glu 3526 of the arm domain during the simulation time (Figure 5). Interaction of Asn with Ser causes the conformational change in the arm domain of IbpAFic2 which is reflected in the B-factor plot and PCA. Hence, interactions of Asn, present on the ATP binding loop, brings about the nucleotide-dependent conformational transitions in the arm domain. Thus, principal component and contact analyses of the 1 μs simulations on apo and holo forms of IbpAFic2 indicate that ATP binding induces allosteric changes in the switch II binding region and its enhanced flexibility helps in substrate recognition through an induced fit mechanism.

Fic Subfamilies Have Conserved Substrate Recognition Motif. In the crystal structure of the IbpA:Cdc42 complex the Fic domain mainly makes contact with the switch I and switch II regions of Cdc42.[18] The switch I region is in contact with the catalytic pocket of the AMPylation domain such that Tyr32, the site of AMPylation, is in close proximity to the bound ATP and facilitates transfer of the AMP moiety from ATP to the tyrosine residue. Also the switch II region of Cdc42 makes contact with the arm domain of IbpAFic2. We carried out explicit solvent MD simulations on the IbpA:Cdc42

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

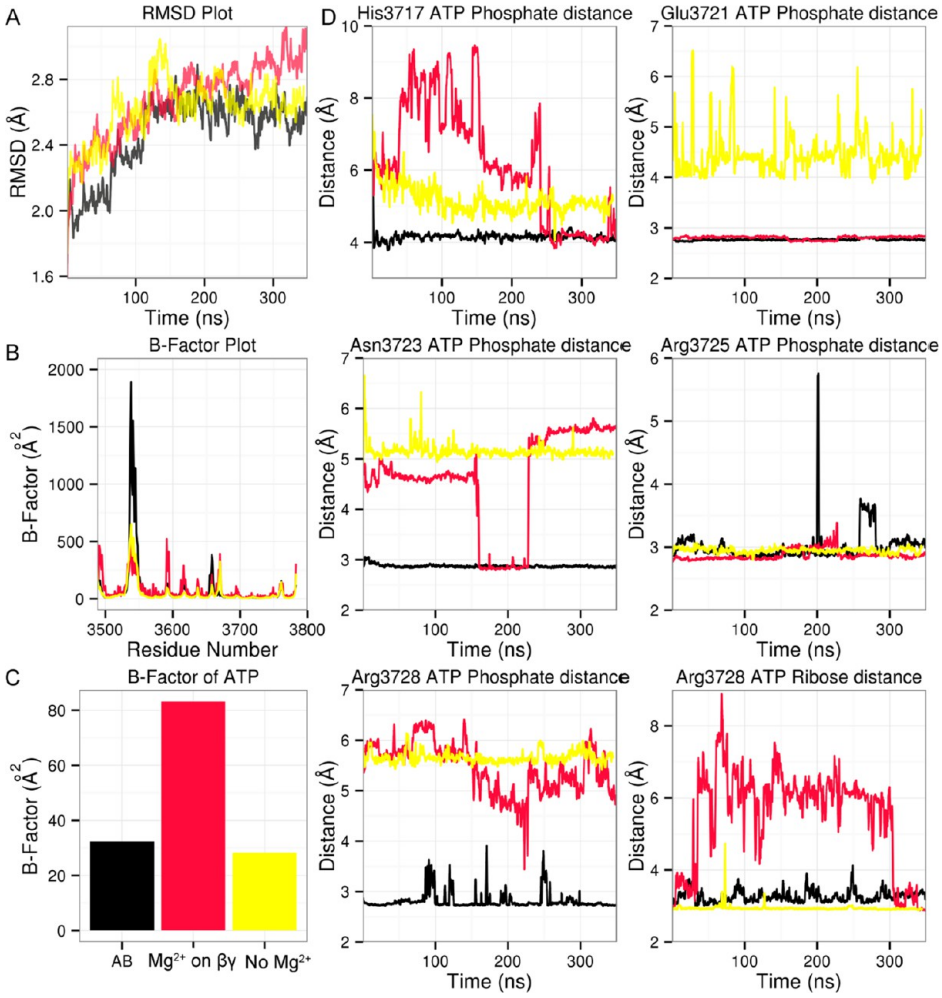


Figure 3. Comparison of the dynamic behavior of three modeled holo IbpAFic2 systems; black lines and bars represent IbpAFic2 + Mg^{2+} docked on $\alpha\beta$ phosphates (IbpA_H2). Red represents IbpAFic2 + Mg^{2+} on $\beta\gamma$ phosphates (IbpA_H3), and yellow represents IbpAFic2 + ATP only (IbpA_H1). (A) RMSD plots (the curve was running average filtered, 1 ns time window) and (B) B-Factor plots of protein depict overall and residue wise flexibility of the protein. (C) B-Factor of ATP of the three systems represents the flexibility of the nucleotide. (D) Interactions of experimentally verified catalytic residues with ATP were studied by plotting the variation of the shortest distances between the residues and ATP (the curves were running average filtered, 1 ns time window).

complex to investigate if these contacts present in the static crystal structure persist upon incorporation of flexibility and if additional contacts formed during the MD simulations define the dynamic interface between IbpAFic2 and Cdc42. The conformations sampled during 500 ns trajectory were analyzed to identify both switch I and switch II interface residues. A representative set of structures were selected from the trajectory by clustering the sampled conformers based on structural similarity,[36] and the interface contacts were identified in these representative set of structures using PIC web server.[37] Analysis of interface contacts which persisted more than 60% of the 500 ns simulation (Supplementary Figure S5 and S6) time revealed that most native contacts present in the starting crystal structure remained intact, while some new contacts were formed in both switch I and switch II interface (Figure 6A and Supplementary Figure S6B). Native contacts that persisted during the simulation have been shown in red, and new contacts that were formed during the simulation and persisted for more than 60% of the simulation time have been shown in blue (Figure 6A and Supplementary Figure S6B). The switch II binding region is specific to IbpAFic2 and is absent in other AMPylation domains, while the switch I binding region, a β hairpin loop between α2 and α3,[11] is usually present in most Fic proteins, except Doc. Since the switch I region of Cdc42 contains the site of AMPylation, the switch I interface might harbor the residues which dictate the specificity of a given AMPylation domain toward its substrates. Therefore, the switch I interface was analyzed in greater detail. Figure 6B shows the schematic representation of the switch I binding pocket of IbpAFic2 which is composed of five residues (represented as square boxes) and recognize a seven-residue linear stretch, C-terminus to the site of AMPylation. Out of the five residues of IbpAFic2 involved in recognition of switch I region K3533 is outside the core Fic domain. K3670, L3668, and N3667 of Fic domains are involved in interactions involving side chains, whereas T3669 is involved in interactions involving main chain contacts. In order to understand the substrate recognition by other Fic domains, these four residues (NLTK corresponding to 3667−3770 of IbpAFic2) were extracted from all Fic proteins, and a conservation profile of this switch I binding stretch was visualized using HMM logo[42] (Figure 6C). Apparently, there was no significant conservation in the substrate binding region. However, the sequences could be clustered into seven smaller subfamilies, and the conservation

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

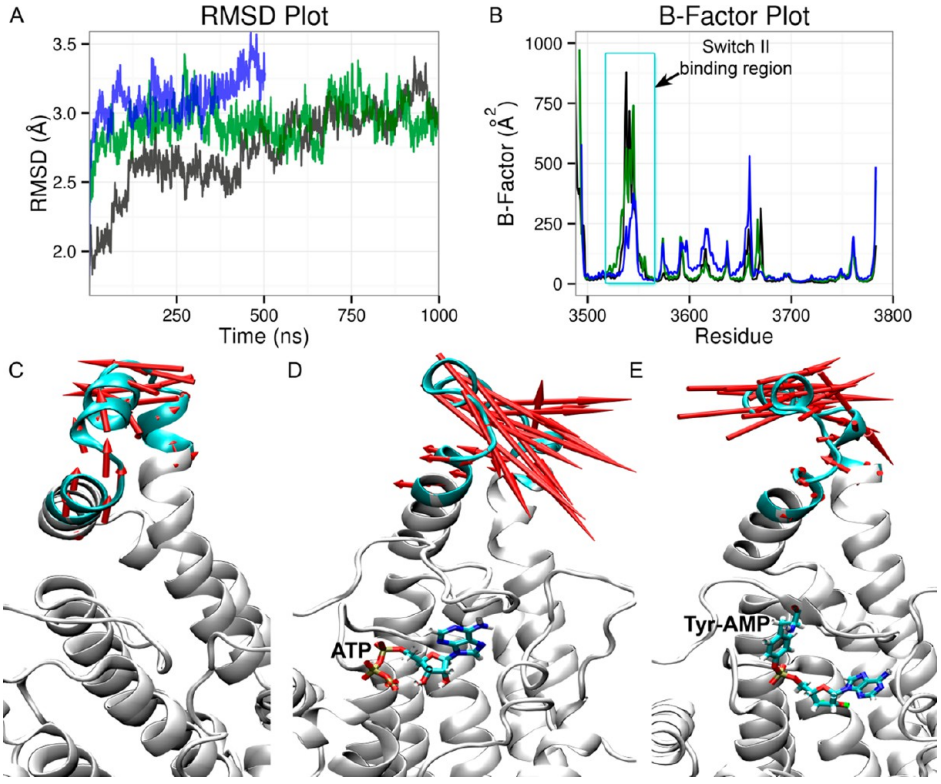


Figure 4. Dynamic behavior of apo and holo IbpAFic2 and IbpAFic2 in complex with AMPylated substrate (A) RMSD plot of the apo, holo, and substrate complexed IbpAFic2. The curve was running average filtered (1 ns time window). (B) B-Factor plot of the three systems showing high flexibility in the switch II binding region (residues 3540−3563, indicated by a cyan box). The color code of (A) and (B) is as follows: black lines represent holo IbpAFic2, where IbpAFic2 is in a complex with ATP and Mg^{2+} ion coordinated by α and β phosphates. Green lines represent apo IbpAFic2 complex, and blue lines represent IbpAFic2 in complex with Cdc42. (C) High flexibility of the switch II binding region (colored cyan) was also confirmed using principal component analysis. Porcupine plots depicting the eigenvectors and projections show a high degree of conformational flexibility in apo IbpAFic2 (C), holo IbpAFic2 (D), and substrate−IbpA complex (E).

of the substrate interacting motif within each subfamily became more prominent (Figure 6D). For instance, the first subfamily (SF 1) has an acidic residue conserved at position 1, a hydrophobic residue at position 2, and lysine or glutamine at position 4. Position 3 shows little conservation not only in SF1 but in other six subfamilies too. Position 3 corresponds to T3669 which is involved in main chain−main chain contacts, and hence most of the amino acids can occur at that position, thus explaining the lack of conservation at this position. Of the 5965 proteins which showed significant alignment with the IbpAFic2 sequence in a structure-based profile−profile search, 975 proteins showed gapped alignment in this region. Of these 975 proteins, a majority (729) are Doc proteins, defined by a degenerate Fic motif of HxFx(D/N)(A/G)NKR[43] and lack of switch I interacting β hairpin. Thus, our structure-based sequence analysis suggests that Fic domains recognize their substrates using distinct sets of conserved motifs present in different subfamilies. Hence, Fic domains belonging to other subfamilies can potentially AMPylate substrates other than GTPases.

Interestingly, a very recent report by Ham et al.[44] which was published after completion of our work has provided experimental evidence for substrate diversity of Fic domains. Ham et al. have shown that substrate of the human and Drosophila Fic domain is the HSP70 domain of chaperon BiP, a non-GTPase Fic substrate. Our structure-guided phylogenetic analysis indicates that most eukaryotic Fic domains (110 out of 238) belong to the SF6 subfamily. Hence, it is possible that all eukaryotic Fic domains might be involved in AMPylation of BiP-like substrates. Apart from eukaryotic Fic domains, the SF6 subfamily in our data set also contains 328 bacterial and 35 archaeal Fic domains. Since interacting proteins in prokaryotic organisms are often encoded by genes in the same genomic neighborhood, we searched for BiP-like proteins (containing HSP70 domain) in the neighborhood of Fic domain containing genes in prokaryotic genomes. Interestingly, an archaeal Fic protein (UniProt Accession: A6UND7_METVS) which is very close to the SF6 subfamily has an HSP70 domain containing protein in the genomic neighborhood. This observation further supports our hypothesis about the involvement of prokaryotic Fic domains belonging to SF6 subfamily in the regulation of BiP like substrates. *Mycobacterium smegmatis* has a Fic domain containing protein belonging to SF6 subfamily, while Fic protein in *M. tuberculosis* belongs to the SF4 subfamily. It would be interesting to explore whether the substrate specificities of these Fic domains are different and whether Fic proteins in *M. tuberculosis* strains are involved in pathogenesis.

Structural Basis of Catalytic Diversity in Fido Family. Doc proteins share sequence and fold level similarity with Fic proteins, whereas AvrB shares only fold level similarity with Fic and Doc proteins. Recent experimental studies have characterized phosphorylation activity for Doc.[43,45] Even though based on some recent findings a model involving phosphorylation of Rin4 peptide by AvrB has been proposed,[19,20] no direct experimental evidence is available for phosphorylation activity of AvrB. Similarly, inverse binding of the nucleotide in

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

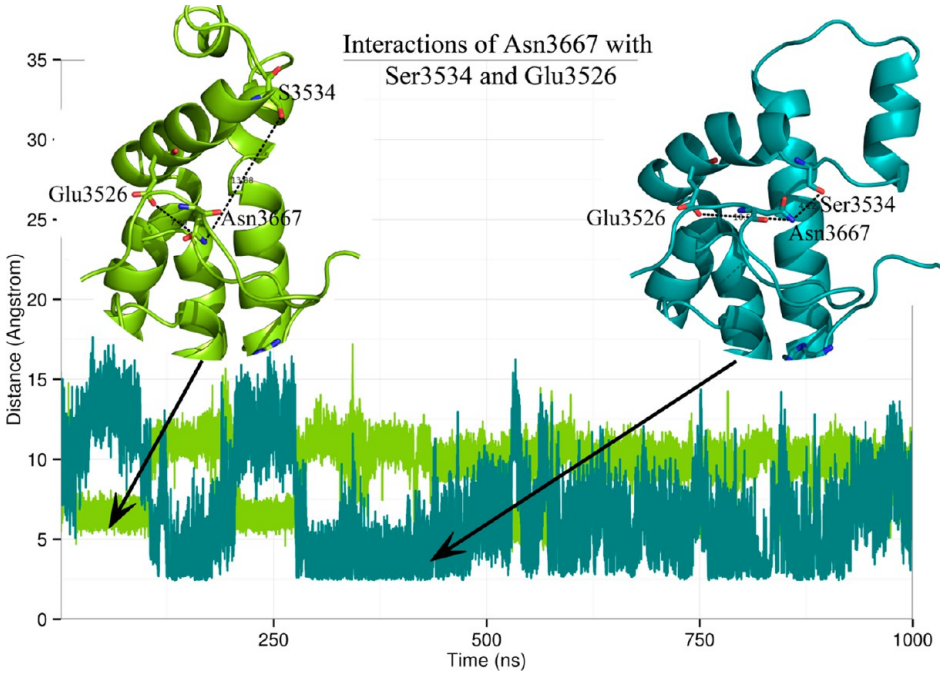


Figure 5. Interactions between substrate binding regions explain the high flexibility. Distance plot of Asn3667 with Glu3526 (blue line) and Asn 3667 and Ser3534 (green line) was plotted to explain the high flexibility in switch II binding region. Asn3667, which lies on the switch I binding region, makes contacts with either Glu3526 or Ser3534. Contact with Ser3534 makes the switch II binding region move significantly as compared to the starting structure where Glu3526 interacts with Asn3667. Insets show snapshots where Asn is in contact with Glu3526 (green) and with Ser3534 (blue).

AnkX has been shown to catalyze the phosphocholination reaction.[14] On the basis of the ATP and substrate binding sites identified by our simulations on IbpAFic2 domain, a comparative analysis of substrate binding sites in Fic, Doc, AvrB, and AnkX domains was carried out to understand how these subfamilies catalyze diverse enzymatic reactions like AMPylation, phosphorylation, and phosphocholination using the same fold. Hence, the crystal structures of IbpAFic2, AnkX,[14] and Doc[46] were aligned to that of AvrB[12] using Dali Server.[47] The Z-score of the IbpAFic2 structure was 9.5 (3.3 Å RMSD over 167 residues), AnkX was 12.9 (3.4 Å over 208 residues), and that of Doc was 6.9 (3.7 Å RMSD over 114 residues). A Z-score of 2 or above is considered to be significant. Structural superposition of crystal structures of IbpA:Cdc42 complex on AvrB:Rin4 complex showed that the incoming nucleophiles (i.e., the site of AMPylation or phosphorylation) from Cdc42 and Rin4 were at the same position, but the adenine moiety of the ADP bound to AvrB had an inverted orientation with respect to the adenine moiety of ATP in IbpAFic2 (Figure 7A). This suggests that, even though substrate binds in a similar orientation in both Fic and AvrB, reverse orientation of nucleotide helps in transfer of AMP in the case of Fic and probably a phosphate group in the case of AvrB.[7] Structural superposition of ADP bound AvrB (PDB ID: 2NUN) onto CDP-phosphocholine bound AnkX (PDB ID: 4BET) revealed that CDP-phosphocholine binds in an orientation similar to ADP in AvrB (Figure 7B), and hence in the case of AnkX too the phosphocholine moiety is transferred to the substrate protein. Since no ATP or ADP bound structures were available for Doc, ATP-binding site on Doc was identified by superposition of ADP bound AvrB (2NUN) onto the crystal structure of Doc in complex with antitoxin PhD. As can be seen from Figure 7C, the ATP binding site overlaps the antitoxin binding site in Doc. It has been hypothesized that PhD inhibits Doc activity by obstruction of the substrate binding site.[48] Also, it is known that the deletion of C-terminal residues of PhD antitoxin abrogates its inhibitory effect on Doc.[49] In a recent study, using NMR chemical shift perturbation data assisted docking Castro-Roa et al. demonstrated that the binding site of nucleotide was occluded by PhD antitoxin.[45] Residues that show maximum chemical shifts on ATP binding have been marked in red in Figure 7D.[45] Transformed ADP from 2NUN interacts with most of the residues supporting the theory that AvrB and Doc share similar nucleotide binding site. Rin4 peptide from the AvrB structure was also transformed on Doc structure. Residues that show maximum chemical shift on EfTu binding have been marked in blue (Figure 7D).[45] The Rin4 peptide too seems to be in the vicinity of residues proposed to be involved in substrate binding. Thus, our structural analysis suggests that Doc and AvrB, which catalyze the phosphorylation reaction, share the donor nucleotide and acceptor substrate binding site. These results are in agreement with a recent report by Garcia-Pino et al.[10]

Conserved Residues Outside Fic Motif Also Help in Fic Catalysis. Despite a common structural fold, sequences of Fic proteins are highly diverged. Nevertheless, Kinch et al.[11] had shown few conserved residues outside the Fic motif that might play a crucial role in the functioning of the Fic domain. We analyzed the conformations sampled during MD simulations on IbpA_A, IbpA_H3, and IbpA:Cdc42 to identify such key residues which might play an important role in catalyzing the AMPylation reaction or maintaining the conserved Fic fold. Our analysis revealed that the aromatic ring of catalytic His (His3717) forms a $\pi-\pi$ stacking interaction with the aromatic ring of Phe3675 (Figure 8A). Similarly, the aromatic rings of Phe3719 (part of the conserved Fic motif) and Tyr3663 were oriented at an angle of ~25° to each other resulting in $\pi-\pi$

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

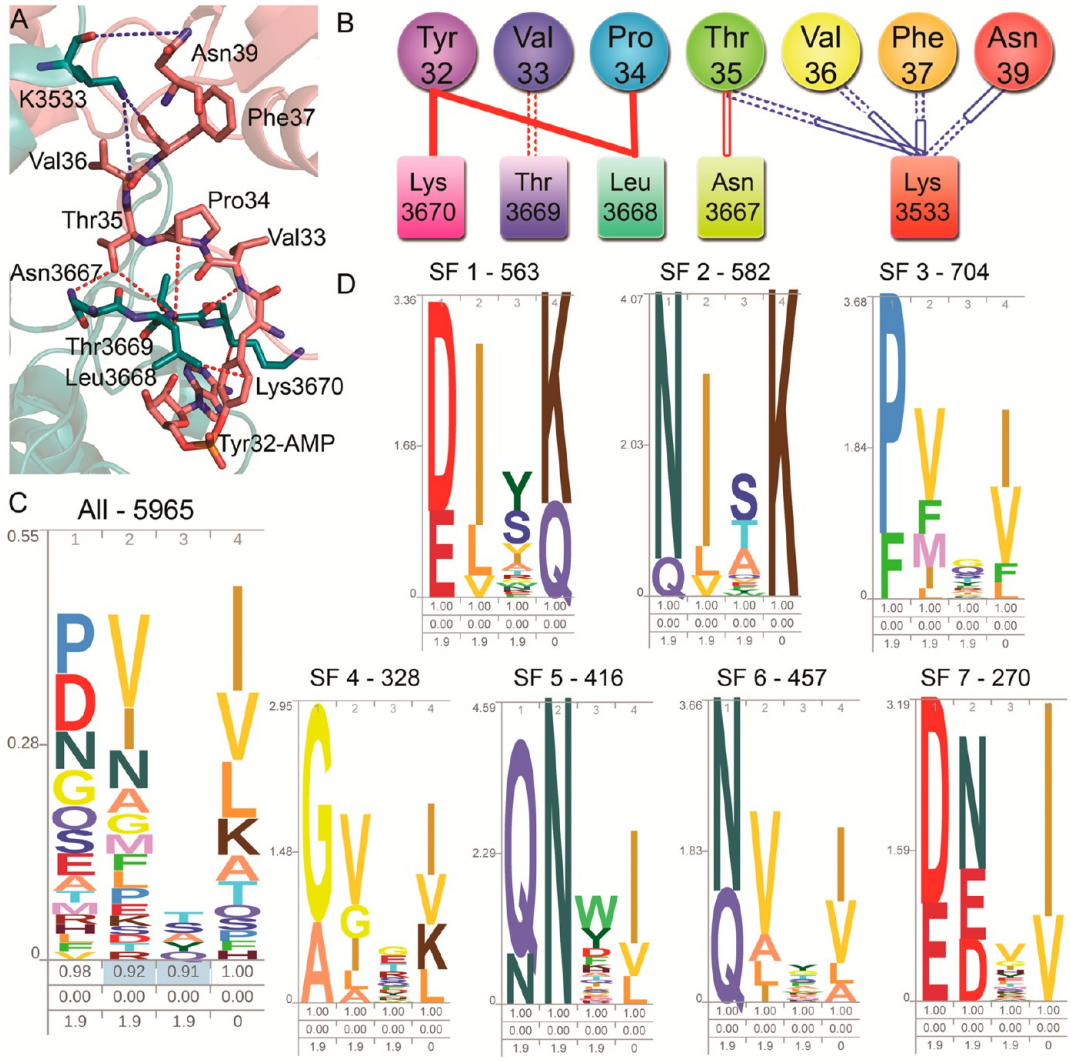


Figure 6. Conservation of binding pockets of IbpAFic2. (A) Interactions of IbpAFic2 with the switch I region of Cdc42. Red lines indicate native contacts that were also present in the crystal structure. Blue lines indicate new contacts that were formed during the molecular dynamics simulations. Only those contacts were considered which were present for 60% or more of the simulation time. (B) Schematic representation of switch I contacts. Circles represent Cdc42 and rectangular boxes represent IbpAFic2 residues. Red and blue lines indicate native and new contacts, solid lines indicate hydrophobic contact, and hollow lines indicate hydrogen bond or ionic interactions. Dotted line represents main chain atoms were involved in the contact, whereas continuous lines indicate the participation of side chain atoms. (C) HMM logo for switch I interacting residues (N3667−K3670) obtained from 5965 different sequences. Though a clear conservation in not visible in the logo, if the Fic family is clustered into several small groups (SF1 to SF7) the conservation becomes much more apparent (D). Numbers below the HMM logs indicate the number of sequences belonging to each group.

stacking interactions (Figure 8B). It may be noted that several studies have reported the role of aromatic interactions in structural stability and catalysis, and the distance range for closest approach of ring carbon atoms at which interactions between the aromatic ring remains favorable has been shown to be between 3.3 and 4.6 Å.[50,51] Therefore, we monitored the variations of distance between these crucial aromatic pairs of residues over the MD trajectories of apo, holo, and product bound structures of the IbpAFic2 AMPylation domain. It is interesting to note that in the holo system IbpA_H3 the closest distance between His3717 and Phe3675 remained within the above-mentioned limit but, for the apo structure IbpA_A the corresponding distance was greater than 5 Å for most part of the 1 μs trajectory indicating a lack of favorable aromatic interactions involving catalytic histidine (Figure 8C,E). Thus, our MD simulations have revealed the novel role of ATP and Mg^{2+} binding in inducing the formation of a π-stacked interaction between His3717 and Phe3675. Detailed analysis of the sampled conformers revealed that the π-stacked interaction between these two rings facilitates proper orientation of catalytic His for inline nucleophilic attack during catalysis (Supplementary Figure S5). In the case of NmFic and HpFic Phe3675Ala mutation reduces the auto-AMPylation activity substantially. The same mutation in IbpA resulted in lowering of enzymatic activity, but the extent of reduction was less compared to NmFic and HpFic.[18] Even though Phe3675 has been shown to be functionally important, the structural basis of the role of Phe3675 in catalysis was unclear. Our simulation results suggest that Phe3675 helps in stabilizing proper orientation of the catalytic histidine residue. The variation of distance between the second pair of aromatic−aromatic rings, i.e., Phe3719 and Tyr3663, over the MD

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

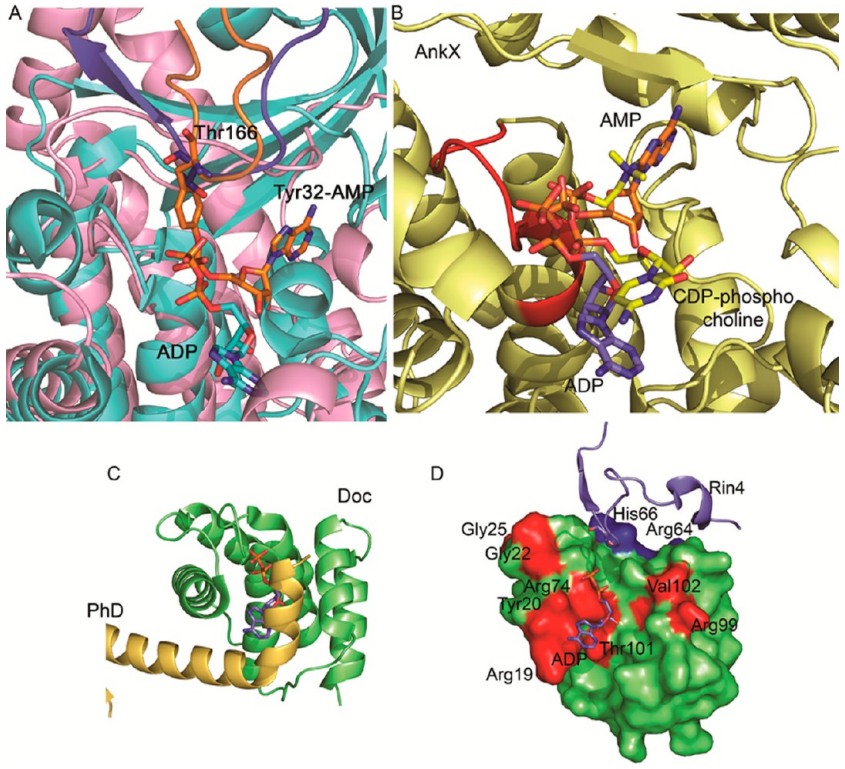


Figure 7. Substrate interacting sites of Fido proteins. (A) Cartoon representation depicting interaction sites of IbpAFic2 (blue) with switch I region of Cdc42 (orange) and AvrB (pink) with Rin4 (deep blue) is similar. (B) Ligand binding site of AnkX (yellow) was compared to that of IbpAFic2 (AMP represented in orange) and AvrB (transposed AvrB represented in deep blue). (C) ADP transposed from AvrB (deep blue) onto Doc crystal structure −3KH2 (green). PhD (yellow) of Doc-PhD shows steric clashes with transposed ADP, obstructing the ligand binding site of Doc. (D) Transposed ADP (deep blue stick representation) interacts with residues that were shown to be important for ATP binding (red colored surface representation). Transposed Rin4 peptide (deep blue, cartoon representation) is in the vicinity of patches deemed important for EfTu interaction of Doc proteins (deep blue, surface presentation), indicating Fido proteins have the same substrate interaction sites.

trajectories revealed that $\pi-\pi$ stacking interactions were maintained throughout the 1 μs trajectories of IbpA_A and IbpA_H3 and also 500 ns trajectory of IbpA:Cdc42 (Figure 8D,F). This suggests that Phe3719−Tyr3663 interaction might be important for stabilizing the fold of Fic proteins.

Since our simulations revealed the role of these aromatic residue pairs in catalysis of AMPylation reaction and stabilization of Fic fold, we wanted to investigate the conservation profile of these residues in other structures of Fic domain and also in sequences of Fic domain in various organisms. Analysis of available structures of Fic domains revealed that, the interacting pair His3717:Phe3675 was conserved in most Fic structures as aromatic ring pairs except SoFic and VopS where Phe3675 was mutated to a proline residue (Supplementary Figure S7A). Proline residues are known to stabilize tertiary structures of proteins forming either hydrophobic or CH/π interactions with aromatic rings.[51,52] Hence, aromatic amino acid and proline comprise a group of cyclic residues that can form π interactions with other aromatic rings. This interaction was also conserved in AnkX, a Fic protein with phosphocholine transferase activity, while the interaction was absent in the Doc structure because of the absence of the β-hairpin that contains Phe3675. Conservation of Phe3675 in AnkX prompted us to check for its conservation in AvrB proteins. The structural alignment revealed His217 of AvrB to be present in place of Phe3675. The histidine residue is conserved in sequences of other AvrB proteins too.[11] A "conservation-of-conservation" of structural features in proteins with the same fold has been shown to have functional importance even though the actual interaction might vary among different proteins.[53] Hence, we predict His217 to play an important role in AvrB. This hypothesis is also supported by experimental studies which have reported abrogation of AvrB induced function upon mutating His217[54] and its proximity to the Thr161 (probable phosphorylation site of Rin4) in peptide bound structure of AvrB (Supplementary Figure S7B).[12] Given its proximity to incoming nucleophile and conservation of the position in Fido domains, we hypothesize that His217 might be the missing catalytic histidine of AvrB. Thus, analysis of the conservation profile of the His3717−Phe3675 pair in Fic structures further supports the proposed role of Phe3675 in catalysis of AMPylation reaction based on our MD simulations. The absence or mutation of the corresponding residue in Doc and AvrB subfamilies also explains why they do not catalyze the AMPylation reaction despite having the conserved Fic fold. Analysis of the conservation profile of the Phe3719−Tyr3663 pair in Fic structures revealed that this interaction was present in most Fic and Doc proteins except SoFic and HpFic (Supplementary Figure S7C).

We have also analyzed Fic/Doc proteins across all genomes to investigate the conservation of these aromatic residue pairs. The His3717−Phe3675 and Phe3719−Tyr3663 pairs are present on two contiguous sequence stretches, one corresponding to the Fic motif around H3717 and the other adjacent to β hairpin insertion around F3675 (Figure 1). Therefore, the corresponding sequence stretches were extracted from all Fic/Doc domains, and HMM logos were generated based on their MSA (Figure 9A). Even though conservation of H3717 and

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

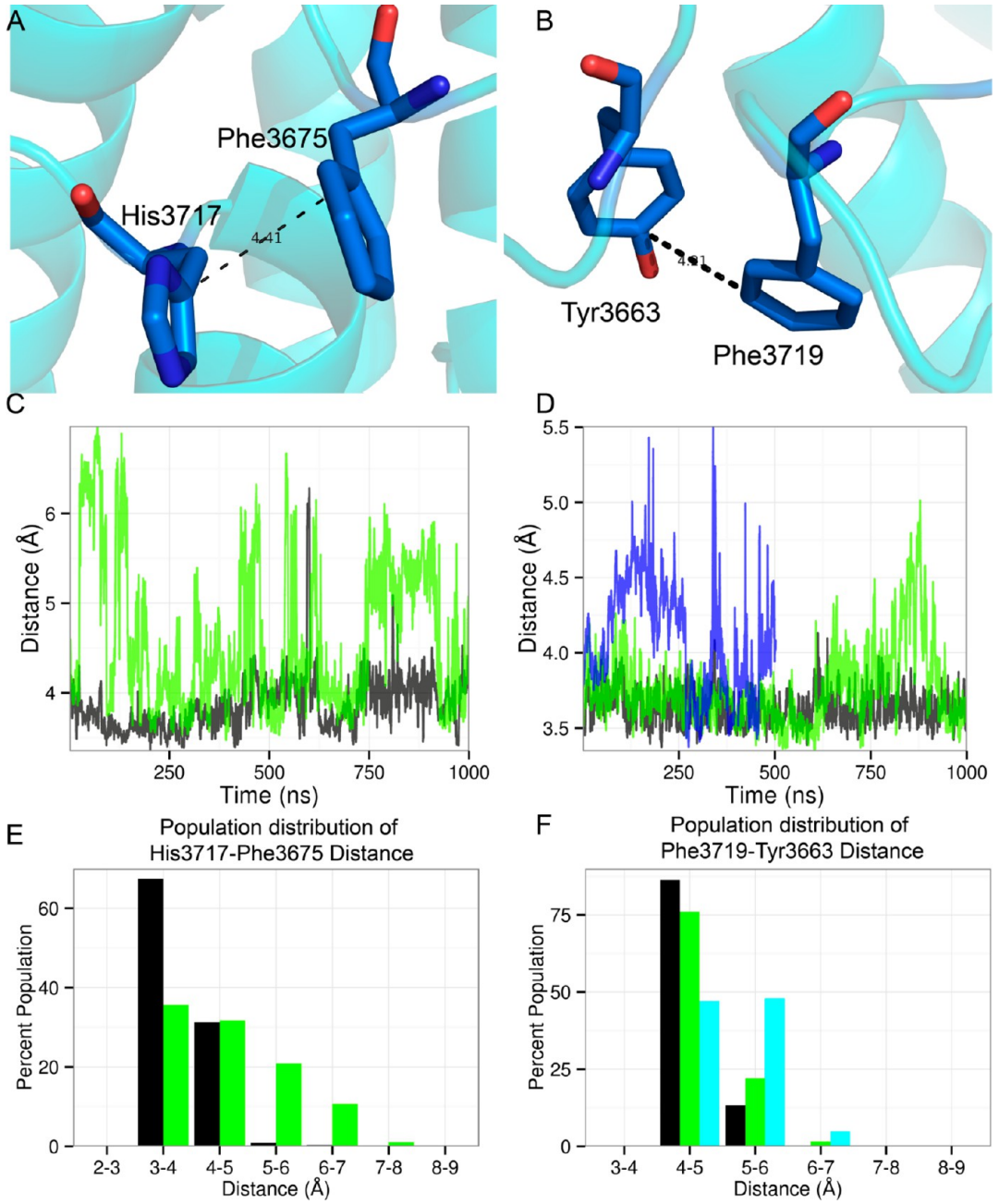


Figure 8. Interaction of Fic motif with other residues. (A) $\pi-\pi$ aromatic interactions of catalytic H3717 and F3675. (B) $\pi-\pi$ aromatic interaction of active site F3719 and Y3663. The curves in A and B were running average filtered (1 ns time window). (C) Variation of H3717−F3675 and F3719−Y3663 (D) over time length of simulations for the three systems. The shortest distance between rings was calculated for this analysis. (E) Population distribution of the interaction distance between H3717 and F3675 represented as a bar graph. The line represents density curve fitted to the bar graph. (F) Population distribution of interaction distance F3719 and Y3663 represented as a bar graph. Black bars represent holo IbpAFic2, where it is in complex with ATP and Mg^{2+} ion coordinated by α and β phosphates. Green bars represent apo IbpAFic2 complex, and blue bars represent IbpAFic2 in complex with Cdc42.

F3719 in the Fic motif is obvious, the sequence stretch harboring Tyr3663 and Phe3675 is less conserved in these sequence-based alignments. In fact, a histidine can be seen in MSA at the site corresponding to Tyr3663 (Figure 9A). Given the high sequence divergence in Fido family of proteins, we analyzed the conservation profile of these interacting aromatic pairs in genome-wide search by using structure-based profile−profile alignment. Interestingly, a significantly higher degree of conservation of His, Pro, Phe, or Tyr was observed in the structure based profile−profile alignments at sites corresponding to Phe3675 and Tyr3663. Therefore, profile−profile alignments revealed that the residues involved in the aromatic−cyclic interactions are conserved across genomes (Figure 9B,C). The Venn diagram for His3717−Phe3675 interaction shows that in 4280 proteins the aromatic−cyclic interactions are conserved (Figure 9B). Of the 6240 proteins where catalytic histidine (His3717) is conserved, 1889 (~30%) proteins do not have the aromatic or cyclic ring containing amino acid conserved at the position corresponding to Phe3675. Of these 1889, 1321 proteins are Doc proteins,

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

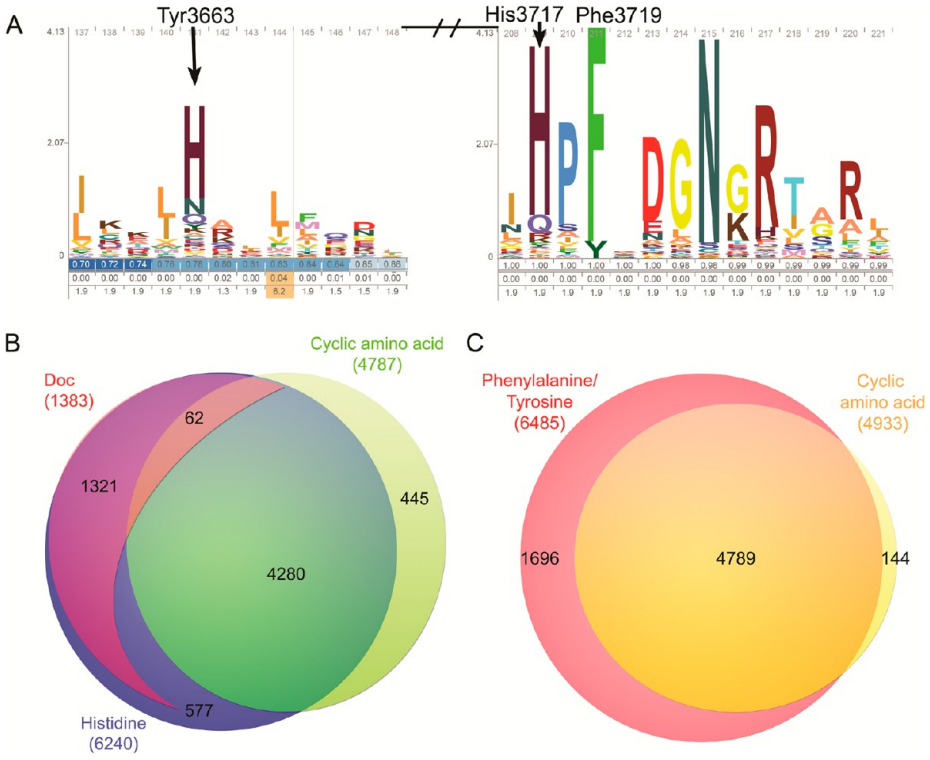


Figure 9. Genome-wide conservation of cyclic amino acid interactions. (A) HMM logo representing concurrent conservation of histidine, a cyclic amino acid, corresponding to position Y3663 and phenylalanine or tyrosine at position F3719. (B) Venn diagrams representing genome-wide conservation of cyclic amino acid interactions corresponding to H3717 and F3675. The deep blue circle represents proteins with conserved catalytic histidine, and lime green represents proteins with conserved cyclic amino acid at position F3675. Overlapped region indicates proteins where both catalytic histidine and cyclic amino acid at the F3675 position is conserved. Cases where interactions corresponding to H3717−F3675 are not conserved are mostly Doc proteins (violet colored segment), which has a missing switch I binding loop. (C) Venn diagram representing genome-wide conservation of ring interaction F3719 and Y3663. Pink circle represents conservation of active site phenylalanine (IbpAFic2 position: 3719). The yellow circle represents conservation of a cyclic residue at position Y3663. The conservation numbers used in the Venn diagram are based on structure-based profile−profile alignments.

lacking the β hairpin harboring Phe3675. Hence, approximately 90% of Fic proteins having a catalytic histidine conserved either have the aromatic−cyclic interaction conserved or are Doc proteins. Figure 9C shows the conservation of second aromatic-cyclic interaction. Of 6485 proteins which have either Phe or Tyr at the corresponding position of Phe3719, 74% have the aromatic−cyclic interaction conserved. Therefore, a genome-wide search reveals that the two crucial aromatic−cyclic interactions identified by our MD simulation studies on IbpAFic2 domain are likely to be conserved in most Fic proteins.

Structure-Based Analysis Reveals a New Fic Subfamily Lacking Conserved Inhibitory Glutamate. Our MD simulations on apo and holo forms of the IbpAFic2 domain showed K3612, a residue outside the Fic motif, to interact with ATP (Figure 10A blue stick). In apo crystal structure 3N3U, Lys3612 (Figure 10A green stick) is away from the catalytic pocket. However, in the holo structure upon binding of ATP, Lys3612 moves into the catalytic pocket and interacts with γ phosphate of ATP. The plot of shortest distance between γ phosphate of ATP and NZ atoms of Lys3613 over the 1 μs trajectory of IbpA_H3 revealed that at several time points they are positioned at an interacting distance (Figure 10B). This suggests that Lys3613 might be important for facilitating ATP binding. However, a structural superimposition of IbpAFic2 (3N3U) on other Fic proteins like HpFic showed that the position corresponding to Lys3612 of IbpAFic2 is occupied by a glutamate which is conserved in a large group of Fic domains (Figure 10A pink stick). Earlier experimental studies have shown this residue to be the conserved inhibitory glutamate of Fic domains. It has been shown that a glutamate residue on a helix in N-terminal or C-terminal of Fic domains or from an interacting protein encoded by a neighboring gene on the genome can regulate the AMPylation activity of Fic domains by obstructing the ATP binding site.[6,16] On the basis of the presence of this conserved inhibitory glutamate on a helix in N-terminus, C-terminus, or on another interacting protein Fic domains have been classified into three groups. Therefore, it was intriguing to note that in the IbpAFic2 domain the inhibitory glutamate has mutated to a positively charged Lys residue that aids in the binding of ATP. Thus, IbpA could belong to a novel class of Fic domains whose AMPylation activity is not regulated by inhibitory glutamate, but rather enhanced by the presence of activating Lys. To analyze if Fic domains containing positively charged Lys/Arg are also present in other organisms, we carried out genome-wide search using structure based profile−profile alignments as described in the Methods section. Interestingly, 151 Fic proteins showed a conservation of positively charged Lys/Arg residue at a position corresponding to inhibitory Glu (Figure 10C−E), and detailed analysis revealed that most of these 151 Fic domains show significant sequence divergence from IbpAFic2. Even though the number is smaller compared to the total number of Fic proteins, our analysis has identified a novel class of Fic

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

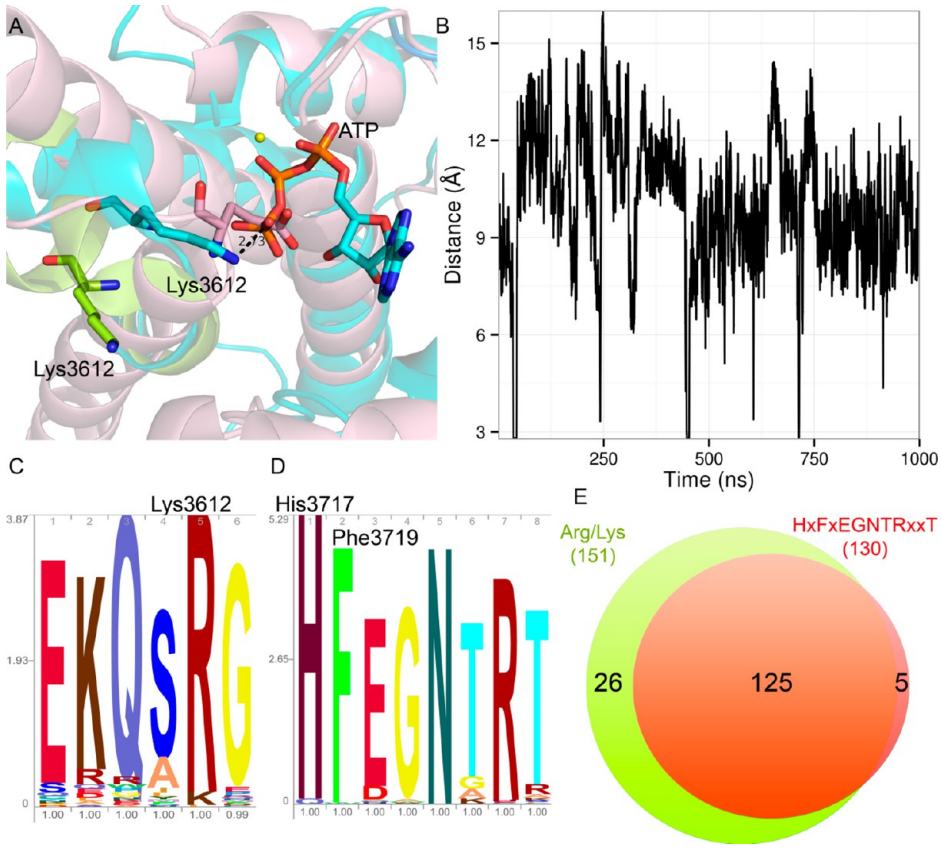


Figure 10. Variations in the regulatory helix of Fic domain. (A) Cartoon representation of inhibitory glutamate blocking ATP-binding site in HpFic (pink). Lys3612 residue is occupying the same position in IbpAFic2 (green). Snapshot of lysine interacting with γ phosphate of ATP during MD simulations. (B) Distance variation between the side chain of Lys3612 and γ phosphate of ATP. The curve was running average filtered (1 ns time window). (C) HMM logo of modified inhibitory motif in Fic proteins with arginine or lysine in place of inhibitory glutamate. (D) HMM logo of the conserved Fic catalytic motif in proteins where the inhibitory motif is modified. (E) Venn diagram representing frequency of occurrence of arginine or lysine in place of inhibitory glutamate (green), a slightly modified Fic motif with threonine instead of second glycine (red) and frequency of their co-occurrence (overlap).

domains. Though these domains were majorly present in organisms from streptococcal genus (taxonomical class - firmicutes), they were also present in classes such as proteobacterial, verrucomicrobia, and eukaryotic classes (Supplementary Figure S8 and Supplementary File 2). We were intrigued to investigate whether the loss of the inhibitory residue had other functional implications on Fic domains. Interestingly, the loss of the inhibitory motif correlated with a slightly modified Fic motif. As can be seen from Figure 10E, 125 out of these 151 Fic proteins carrying Arg/Lys in place of the inhibitory glutamate had a modified Fic motif of HxFxEGNTRxxT (Figure 10D) in place of the usual Fic motif HxFxEGNGRxxR. Mutation of the second arginine in the regular Fic motif to threonine in this new Fic family might result in loss of an important ATP binding interaction (Figure 2C). The positively charged Lys/Arg in place of inhibitory Glu provides an additional ATP binding residue to compensate this loss of ATP binding interaction because of Arg to Thr mutation in the Fic motif. Therefore, this novel Fic family identified by our bioinformatics analysis has evolved by correlated mutations involving inhibitory Glu and ATP binding Arg. It is tempting to speculate that AMPylation activity of this class of Fic domains lacking inhibitory glutamate can probably be regulated by phosphorylation of threonine in the modified Fic motif. In order to check the solvent accessibility of this threonine, one of the protein sequences from this novel subfamily of Fic domain (UNIPROT accession – H7H3Z4) was modeled using Swiss-Model,[55] and the solvent accessible surface area of the threonine was computed using NACCESS.[56] The accessibility results showed that though the threonine was not on the surface of Fic domain, it was more accessible than the core residues of the protein. Earlier studies have observed that partially exposed threonine residues in proteins can also be phosphorylated.[57] Therefore, it is possible that the threonine in the modified Fic motif can be phosphorylated despite being partially exposed to solvent.

DISCUSSION

Recent experimental studies indicate that Fic domain containing proteins catalyze AMPylation reaction, while proteins belonging to Doc, a subfamily of Fic, have been shown to catalyze phosphorylation[43,45] reaction. On the other hand AnkX, another Fic domain containing protein, catalyzes phosphocholine transfer.[58] AvrB, belonging to a family of proteins structurally homologous to Fic, has been implicated in phosphorylation of its target Rin4.[19] In addition to the diversity in the enzymatic reactions (Supplementary Figure S2) the Fido domains catalyze, Fic proteins which carry out AMPylation have also been shown to recognize a wide range of substrates (Supplementary Table S1).[10,59] In order to decipher the structural basis of substrate recognition by Fic domain, we have carried out microsecond scale explicit solvent atomistic

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

simulations on apo, holo, and substrate/product bound IbpA Fic domain (IbpAFic2). The results from the simulation studies on this representative Fic domain have been extrapolated to other Fic domains by using a structure-based sequence analysis approach.

Even though experimental studies had established the crucial role of Mg^{2+} ion in catalysis of the AMPylation reaction by Fic domains, two different coordination modes of Mg^{2+} ion have been proposed in the literature based on biochemical studies by Lounge et al.[17] on the VopS and Mg^{2+} bound crystal structure of the VbhA Fic domain.[16] Our explicit solvent molecular dynamics studies revealed that for maintaining the contacts between ATP and catalytically important active site residues, the Mg^{2+} ion needs to be coordinated by α and β phosphates (as seen in the VbhA crystal structure). Comparison of the microsecond MD simulations on apo and holo IbpAFic2 and 500 ns simulations on the substrate/product bound IbpAFic2−Cdc42 complex revealed a novel allosteric mechanism for substrate recognition by the Fic domain. ATP binding at the catalytic site near the switch I binding region induced conformational changes at the distal switch II binding region which interacts with the substrate protein (Figure 4). This allosteric communication between switch I and switch II binding regions of IbpAFic2 were found to be mediated by the interactions between Asn3667 (present on the nucleotide binding loop) and Ser3534 or Glu3526 present on the switch II binding region. The ligand-induced conformational flexibility can be verified by using either dynamic NMR studies or isothermal titration calorimetry which might indicate the increased entropy upon ligand binding. Apart from identifying residues mediating this novel allosteric communication involved in substrate recognition, MD simulations in combination with phylogenetic analysis helped in deciphering the functional role of certain crucial conserved aromatic residues outside the Fic motif. MD simulations revealed the novel role of ATP and Mg^{2+} binding in inducing the formation of a π-stacked interaction between His3717 and Phe3675, which facilitates proper orientation of catalytic His for inline nucleophilic attack during catalysis (Figure 8). Thus, our computational studies elucidated the structural basis for the role of Phe3675, which has been shown to be important for catalysis not only in IbpA but also in NmFic and HpFic.[18] Our simulations also revealed that aromatic ring interaction between the conserved pair Phe3719−Tyr3663 might be important for stabilizing the conserved structural fold of Fic proteins. Mutagenesis studies of these aromatic residues not only in IbpA but also in other AMPylating Fic domains would help in verifying the importance of these interactions for stabilizing the Fic fold. Also, NMR studies can shed light on the stability of these aromatic interactions.

To identify the specificity determining residues (SDR) of IbpAFic2 domain, we analyzed the interface contacts between IbpAFic2 and Cdc42, which persisted over 60% of the 500 ns simulation time. It was found that a tetrapeptide stretch (NLTK corresponding to 3667−3770 of IbpAFic2) present in the β hairpin loop connecting α2 and α3 helices of Fic domain was involved in specific interactions with the residues flanking the site of AMPylation on the substrate protein. Thus, it could constitute the SDR of Fic domains. Structure-based sequence analysis revealed a distinct conservation pattern for this stretch in different Fic subfamilies and based on the conservation patterns Fic/Doc domains could be classified into seven groups (Figure 6D). We hypothesized that Fic domains recognize their substrates for AMPylation using distinct sets of conserved recognition motifs present in different subfamilies, and hence Fic domains belonging to other subfamilies can potentially AMPylate substrates other than GTPases. It is encouraging to note that a very recent report by Ham et al.[44] which was published after completion of our work has shown that substrate of human and Drosophila Fic domain (belonging to SF6 subfamily as per our classification) is the HSP70 domain of chaperon BiP, a non-GTPase Fic substrate. Thus, our computational analysis can also help in identifying AMPylation domains that can potentially recognize novel substrates.

Another functionally important interaction revealed through MD simulation studies was that of Lys3612 with γ phosphate of ATP. Interestingly, this lysine residue is present at the same position as that of inhibitory glutamate on the regulatory helix in many Fic proteins.[6] Phylogenetic analysis revealed mutation of the inhibitory Glu to Lys/Arg in a subfamily of Fic domains. The inhibitory glutamate on regulatory helix has been proposed to be involved in autoinhibition of Fic domains by obstructing ATP binding. The observed interaction between Lys3612 and ATP in our MD simulations suggests that the Lys/Arg present on the regulatory helix in place of inhibitory glutamate could be involved in auto-enhancement of AMPylation activity. Interestingly in this subfamily the inhibitory Glu to Lys/Arg mutation is also correlated with mutation of the conserved Arg of the Fic motif HxFxEGNGRxxR to threonine in HxFxEGNTRxxT. In our earlier phylogenetic analysis, this subfamily formed a separate clade, and we had hypothesized that it may form an isofunctional subfamily.[60] It is also possible that AMPylation activity of this new subfamily of Fic domains lacking inhibitory glutamate can be regulated by phosphorylation of second Thr in the modified Fic motif. Therefore, this new Fic family present mostly in firmicutes could be an interesting target for experimental studies. The importance of Lys3612 in ATP binding and catalysis can be ascertained by *in vitro* mutagenesis experiments where the Lys residue of IbpA and corresponding Lys/Arg residues of other Fic domains can be mutated to either alanine or glycine. The AMPylation capability of these mutants can then be tested by using [α-^{32}P] ATP as ligand and Rho-GTPases as a substrate in an *in vitro* autoradiography assay. Our hypothesis that the Fic family can be subdivided into seven isofunctional subfamilies can also be tested by choosing proteins from various subfamilies and performing high-throughput NAPPA assays as described by Yu et al.[61] The results can then be confirmed using autoradiography-based AMPylation assays or kinetic studies.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00351.

Supplementary Table S1 and Supplementary Figures S1−S9 (PDF)

Spreadsheet containing a list of Fic domain containing proteins with modified regulatory motif and their source organism (XLS)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: deb@nii.res.in.

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

Funding

This work is supported by grants to National Institute of Immunology, New Delhi, from Department of Biotechnology (DBT), Government of India. D.M. also acknowledges financial support from DBT, India, under BTIS project (BT/BI/03/009/2002) and Bioinformatics R&D grant (BT/PR13526/BID/07/311/2010). S.K. acknowledges the support from DBT, India, in the form of BINC fellowship.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Itzen, A., Blankenfeldt, W., and Goody, R. S. (2011) Adenylylation: renaissance of a forgotten post-translational modification. *Trends Biochem. Sci. 36*, 221−228.

(2) Anderson, W. B., and Stadtman, E. R. (1970) Glutamine synthetase deadenylation: a phosphorolytic reaction yielding ADP as nucleotide product. *Biochem. Biophys. Res. Commun. 41*, 704−709.

(3) Worby, C. A., Mattoo, S., Kruger, R. P., Corbeil, L. B., Koller, A., Mendez, J. C., Zekarias, B., Lazar, C., and Dixon, J. E. (2009) The fic domain: regulation of cell signaling by adenylylation. *Mol. Cell 34*, 93−103.

(4) Yarbrough, M. L., Li, Y., Kinch, L. N., Grishin, N. V., Ball, H. L., and Orth, K. (2009) AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. *Science 323*, 269−272.

(5) Zekarias, B., Mattoo, S., Worby, C., Lehmann, J., Rosenbusch, R. F., and Corbeil, L. B. (2010) Histophilus somni IbpA DR2/Fic in virulence and immunoprotection at the natural host alveolar epithelial barrier. *Infection and immunity 78*, 1850−1858.

(6) Engel, P., Goepfert, A., Stanger, F. V., Harms, A., Schmidt, A., Schirmer, T., and Dehio, C. (2012) Adenylylation control by intra- or intermolecular active-site obstruction in Fic proteins. *Nature 482*, 107−110.

(7) Palanivelu, D. V., Goepfert, A., Meury, M., Guye, P., Dehio, C., and Schirmer, T. (2011) Fic domain-catalyzed adenylylation: insight provided by the structural analysis of the type IV secretion system effector BepA. *Protein Sci. 20*, 492−499.

(8) Rahman, M., Ham, H., Liu, X., Sugiura, Y., Orth, K., and Kramer, H. (2012) Visual neurotransmission in Drosophila requires expression of Fic in glial capitate projections. *Nat. Neurosci. 15*, 871−875.

(9) Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., and Punta, M. (2014) Pfam: the protein families database. *Nucleic Acids Res. 42*, D222−230.

(10) Garcia-Pino, A., Zenkin, N., and Loris, R. (2014) The many faces of Fic: structural and functional aspects of Fic enzymes. *Trends Biochem. Sci. 39*, 121−129.

(11) Kinch, L. N., Yarbrough, M. L., Orth, K., and Grishin, N. V. (2009) Fido, a novel AMPylation domain common to fic, doc, and AvrB. *PLoS One 4*, e5818.

(12) Desveaux, D., Singer, A. U., Wu, A. J., McNulty, B. C., Musselwhite, L., Nimchuk, Z., Sondek, J., and Dangl, J. L. (2007) Type III effector activation via nucleotide binding, phosphorylation, and host target interaction. *PLoS Pathog. 3*, e48.

(13) Garcia-Pino, A., Christensen-Dalsgaard, M., Wyns, L., Yarmolinsky, M., Magnuson, R. D., Gerdes, K., and Loris, R. (2008) Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. *J. Biol. Chem. 283*, 30821−30827.

(14) Campanacci, V., Mukherjee, S., Roy, C. R., and Cherfils, J. (2013) Structure of the Legionella effector AnkX reveals the mechanism of phosphocholine transfer by the FIC domain. *EMBO J. 32*, 1469−1477.

(15) Das, D., Krishna, S. S., McMullan, D., Miller, M. D., Xu, Q., Abdubek, P., Acosta, C., Astakhova, T., Axelrod, H. L., Burra, P., Carlton, D., Chiu, H. J., Clayton, T., Deller, M. C., Duan, L., Elias, Y., Elsliger, M. A., Ernst, D., Feuerhelm, J., Grzechnik, A., Grzechnik, S. K., Hale, J., Han, G. W., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W., Kozbial, P., Kumar, A., Marciano, D., Morse, A. T., Murphy, K. D., Nigoghossian, E., Okach, L., Oommachen, S., Paulsen, J., Reyes, R., Rife, C. L., Sefcovic, N., Tien, H., Trame, C. B., Trout, C. V., van den Bedem, H., Weekes, D., White, A., Hodgson, K. O., Wooley, J., Deacon, A. M., Godzik, A., Lesley, S. A., and Wilson, I. A. (2009) Crystal structure of the Fic (Filamentation induced by cAMP) family protein SO4266 (gil24375750) from Shewanella oneidensis MR-1 at 1.6 A resolution. *Proteins: Struct., Funct., Genet. 75*, 264−271.

(16) Goepfert, A., Stanger, F. V., Dehio, C., and Schirmer, T. (2013) Conserved inhibitory mechanism and competent ATP binding mode for adenylyltransferases with Fic fold. *PLoS One 8*, e64901.

(17) Luong, P., Kinch, L. N., Brautigam, C. A., Grishin, N. V., Tomchick, D. R., and Orth, K. (2010) Kinetic and structural insights into the mechanism of AMPylation by VopS Fic domain. *J. Biol. Chem. 285*, 20155−20163.

(18) Xiao, J., Worby, C. A., Mattoo, S., Sankaran, B., and Dixon, J. E. (2010) Structural basis of Fic-mediated adenylylation. *Nat. Struct. Mol. Biol. 17*, 1004−1010.

(19) Innes, R. W. (2011) Activation of plant nod-like receptors: how indirect can it be? *Cell Host Microbe 9*, 87−89.

(20) Liu, J., Elmore, J. M., Lin, Z. J., and Coaker, G. (2011) A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe 9*, 137−146.

(21) Khater, S., and Mohanty, D. (2015) novPTMenzy: a database for enzymes involved in novel post-translational modifications, *Database: the Journal of Biological Databases and Curation*, bav039.

(22) Morris, G. M., Huey, R., and Olson, A. J. (2008) Using AutoDock for ligand-receptor docking, *Current Protocols in Bioinformatics*, Baxevanis, A. D. et al., Chapter 8, Unit 8 14.

(23) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926−935.

(24) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hseih, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; and Kollman, P. A. (2010) AMBER11, University of California, San Francisco, CA.

(25) Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem. 24*, 1999−2012.

(26) Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003) Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem. 24*, 1016−1025.

(27) Allnér, O., Nilsson, L., and Villa, A. (2012) Magnesium Ion-Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput. 8* (4), 1493−1502.

(28) Yadav, G., Gokhale, R. S., and Mohanty, D. (2003) Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol. 328*, 335−363.

(29) Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., and Hagler, A. T. (1988) Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins: Struct., Funct., Genet. 4*, 31−47.

(30) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys. 23*, 327−341.

(31) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089−10092.

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224

(32) Schroeder, E. K., Basso, L. A., Santos, D. S., and de Souza, O. N. (2005) Molecular dynamics simulation studies of the wild-type, I21V, and I16T mutants of isoniazid-resistant Mycobacterium tuberculosis enoyl reductase (InhA) in complex with NADH: toward the understanding of NADH-InhA different affinities. *Biophys. J. 89*, 876−884.
(33) Zeileis, A., Grothendieck, G., Ryan, J. A., and Andrews, F. (2015) zoo: S3 Infrastructure for Regular and Irregular Time Series, 1.7−12 ed.
(34) Mongan, J. (2004) Interactive essential dynamics. *J. Comput.-Aided Mol. Des. 18*, 433−436.
(35) Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*, Springer, New York.
(36) Feig, M., Karanicolas, J., and Brooks, C. L., 3rd (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell. 22*, 377−395.
(37) Tina, K. G., Bhadra, R., and Srinivasan, N. (2007) PIC: Protein Interactions Calculator. *Nucleic Acids Res. 35*, W473−476.
(38) Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res. 33*, W244−248.
(39) Remmert, M., Biegert, A., Hauser, A., and Soding, J. (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods 9*, 173−175.
(40) Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol. 292*, 195−202.
(41) Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics 21*, 951−960.
(42) Wheeler, T. J., Clements, J., and Finn, R. D. (2014) Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinf. 15*, 7.
(43) Cruz, J. W., Rothenbacher, F. P., Maehigashi, T., Lane, W. S., Dunham, C. M., and Woychik, N. A. (2014) Doc toxin is a kinase that inactivates elongation factor Tu. *J. Biol. Chem. 289*, 7788−7798.
(44) Ham, H., Woolery, A. R., Tracy, C., Stenesen, D., Kramer, H., and Orth, K. (2014) Unfolded protein response-regulated dFic reversibly AMPylates BiP during endoplasmic reticulum homeostasis. *J. Biol. Chem. 289*, 36059−36069.
(45) Castro-Roa, D., Garcia-Pino, A., De Gieter, S., van Nuland, N. A., Loris, R., and Zenkin, N. (2013) The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nat. Chem. Biol. 9*, 811−817.
(46) Arbing, M. A., Handelman, S. K., Kuzin, A. P., Verdon, G., Wang, C., Su, M., Rothenbacher, F. P., Abashidze, M., Liu, M., Hurley, J. M., Xiao, R., Acton, T., Inouye, M., Montelione, G. T., Woychik, N. A., and Hunt, J. F. (2010) Crystal structures of Phd-Doc, HigA, and YeeU establish multiple evolutionary links between microbial growth-regulating toxin-antitoxin systems. *Structure 18*, 996−1010.
(47) Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res. 38*, W545−549.
(48) Gazit, E., and Sauer, R. T. (1999) The Doc toxin and Phd antidote proteins of the bacteriophage P1 plasmid addiction system form a heterotrimeric complex. *J. Biol. Chem. 274*, 16813−16818.
(49) McKinley, J. E., and Magnuson, R. D. (2005) Characterization of the Phd repressor-antitoxin boundary. *Journal of bacteriology 187*, 765−770.
(50) Aravinda, S., Shamala, N., Das, C., Sriranjini, A., Karle, I. L., and Balaram, P. (2003) Aromatic-aromatic interactions in crystal structures of helical peptide scaffolds containing projecting phenylalanine residues. *J. Am. Chem. Soc. 125*, 5308−5315.
(51) Bhattacharyya, R., and Chakrabarti, P. (2003) Stereospecific interactions of proline residues in protein structures and complexes. *J. Mol. Biol. 331*, 925−940.
(52) Zondlo, N. J. (2013) Aromatic-proline interactions: electronically tunable CH/pi interactions. *Acc. Chem. Res. 46*, 1039−1049.
(53) Mirny, L. A., and Shakhnovich, E. I. (1999) Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol. 291*, 177−196.
(54) Ong, L. E., and Innes, R. W. (2006) AvrB mutants lose both virulence and avirulence activities on soybean and Arabidopsis. *Mol. Microbiol. 60*, 951−962.
(55) Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2008) Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc. 4*, 1−13.
(56) Hubbard, S., and Thornton, J. (1992) Naccess V2.1.1 - Atomic Solvent Accessible Area Calculations.
(57) Kumar, N., Damle, N., and Mohanty, D. (2015) Getting Phosphorylated: Is it Necessary to be Solvent Accessible? *Proc. Indian Natl. Sci. Acad., Part A 81*, 493−507.
(58) Mukherjee, S., Liu, X., Arasaki, K., McDonough, J., Galan, J. E., and Roy, C. R. (2011) Modulation of Rab GTPase function by a protein phosphocholine transferase. *Nature 477*, 103−106.
(59) Cruz, J. W., and Woychik, N. A. (2014) Teaching Fido New ModiFICation Tricks. *PLoS Pathog. 10*, e1004349.
(60) Khater, S., and Mohanty, D. (2015) In silico identification of AMPylating enzymes and study of their divergent evolution. *Sci. Rep. 5*, 10804.
(61) Yu, X., and LaBaer, J. (2015) High-throughput identification of proteins with AMPylation using self-assembled human protein (NAPPA) microarrays. *Nat. Protoc. 10*, 756−767.

DOI: 10.1021/acs.biochem.5b00351
Biochemistry 2015, 54, 5209−5224