This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
IspH-RPS1 and IspH-UbiA: "Rosetta Stone" Proteins

Guodong Rao, Bing O'Dowd, Jikun Li, Ke Wang, and Eric Oldfield

The protein IspH, (E)-1-hydroxy-2-methyl-but-2-enyl 4-diphosphate (HMPPP) reductase, is an essential 4Fe-4S cluster-containing protein in the methylerythritol phosphate pathway for isoprenoid biosynthesis. Using a sequence similarity network we found that there are >400 IspH proteins that are about twice as large as most of the IspHs studied to date since their IspH domains are fused to either the ribosomal protein S1 (RPS1), or to a UbiA (4-hydroxybenzoate octaprenyltransferase)-like protein. Many of the IspH-RPS1 proteins are present in anaerobes found in the human gut and some, such as Clostridium botulinum, C. tetani and Fusobacterium nucleatum, are pathogens. The IspH-UbiAs are all found in sulfate-reducing anaerobes. The IspH domains in IspH-RPS1 are fused to 4 and in a few cases 6 tandem repeats in RPS1 that, in most organisms, bind to mRNA or form part of the bacterial ribosome. Mutants in which the four RPS1 domains were sequentially eliminated had similar IspH activity as wild-type protein, indicating they are not essential for IspH catalysis. Overall, the results are of interest since they represent the first isolation of a catalytically active IspH, as well as the identification of IspH-UbiA hybrids, two "Rosetta stone" proteins that are likely to be functionally related—IspH producing the isoprenoids required for a UbiA-like prenyl transferase; the IspH-RPS1 hybrids, perhaps, being involved in the stringent response or as Fe/O sensors.

Introduction

The methylerythritol phosphate pathway is responsible for the production of the terpene building blocks dimethylallyl diphosphate (DMAPP, 1, Scheme 1) and isopentenyl diphosphate (IPP, 2) in most bacteria, apicomplexan parasites, and plants. Its component enzymes are all essential for survival, so there is interest in their inhibition1-2 for the development of drugs to treat infectious diseases, such as malaria and tuberculosis, with fosmidomycin inhibition for the development of drugs to treat infectious diseases, such as malaria and tuberculosis, with fosmidomycin targeting the isopentenyl diphosphate reductase, having reached clinical trials for malaria3. In addition, targeting the pathway could be of use for herbicide development. IspH, (E)-1-hydroxy-2-methyl-but-2-enyl 4-diphosphate (HMPPP, 3) reductase, also known as LytB (but not to be confused with the N-acetylglucosaminidase LytB from Streptococcus spp.), is the last enzyme in the pathway4. IspH is a 4Fe-4S cluster-containing protein5 that catalyzes the 2H+/2e- reductive dehydroxylation of HMPPP into a ~1:5 mixture of dimethylallyl diphosphate and isopentenyl diphosphate. Its structure and mechanism of action have been studied extensively, both experimentally7-13 and computationally14-16, and several inhibitors have been developed17. The mechanism of action is remarkable and the consensus view now is that IspH utilizes a bio-organometallic mechanism in which an allyl species binds to the Fe-S cluster14-16.

Despite their now well-known role in isoprenoid biosynthesis, IspHs (known as LytB5 at the time) from E. coli and Campylobacter jejuni were first identified and investigated almost 20 years ago in studies of the so-called “stringent response”, in bacteria17,18. The stringent response is a stress response that can arise from diverse insults and is aimed at limiting growth to promote survival. An IspH mutant E. coli strain WW718 induced the stringent response and exhibited a penicillin-tolerant phenotype at restrictive temperatures (42 °C). This phenotype could be complemented by wild-type E. coli IspH or C. jejuni IspH17, but not by a C. jejuni IspH Q265H mutant and it was proposed that this wild-type IspH could interact “directly or indirectly” with RelA16. RelA is an enzyme that has been shown to bind to the bacterial ribosome19 and is responsible for biosynthesis of the alarmone, (p)ppGpp, guanosine tetraphosphate (or pentaphosphate), the (p)ppGpp, guanosine tetraphosphate (or pentaphosphate), the key regulator involved in the bacterial stringent response20. In addition, RelA is an emerging central regulator of multidrug tolerance and persistence21. If the interaction between IspH and RelA were disrupted, the stringent response would be induced.

In other work it has been noted that that some bacterial proteins appear to contain a fusion in which IspH is linked to another ribosomal protein, RPS1, the ribosomal protein small (30S ribosome) protein17,22,23. Fusion hybrids are often found when two individual proteins have some related activity and are called “Rosetta stone” proteins24,25 and RPS1-IspH is given as one25 example, although the functional relatedness of the 2 domains has not been proposed. Here, we report initial cloning, expression, purification, activity, inhibition, mutagenesis and spectroscopic results on one IspH-RPS1 protein. In addition, we report that several...
other proteins form fusion hybrids with IspH, one of which appears to be a functionally related prenyl synthase.

Materials and Methods

Sequence similarity network. The sequence similarity network for the IspH family proteins (InterPro number IPR003451) was generated by using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST, http://efi.igb.illinois.edu/efi-est/). An expectation-value of 10^{-120} was used to construct the network, which was then visualized by using Cytoscape 3.2.1. A single node represents sequences with at least 90% identity, while each edge joins sequences that share an e-value of 10^{-120} or smaller.

Cloning, protein expression, and purification. CthispH-RPS1 and its mutants were amplified from the genomic DNA of *Clostridium thermocellum* strain VPI 7372 [ATCC® 27405™] by polymerase chain reaction. The amplification product was digested with SacI-HF and Sall-HF (New England Biolabs, MA) and cloned into the pET-28a (+) vector (Novagen, WI). The plasmid with the correct insert was transformed into *E. coli* BL21-CodonPlus (DE3)-RIPL competent cells (Agilent, CA). For protein expression, six liters of LB broth supplemented with 25 mg/mL kanamycin and 17 mg/mL chloramphenicol were inoculated with a 0.5% overnight culture and grown at 37 °C to an O.D.600 of 0.6-0.8. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM and the cultures were grown at 28 °C for a further 24 hrs. The cells were then harvested and stored at -80 °C until further use. CthispH-RPS1 and all mutants were purified by using column chromatography employing a Ni-NTA Hispür™ resin (Fisher, NY), according to the manufacturer’s instructions. Briefly, cell pellets were thawed and suspended in the loading buffer (5 mM imidazole, 50 mM Tris/HCl, 150 mM NaCl, pH = 8.0) containing protease inhibitor cocktail tablet (Roche) and lyzed by sonication. The cell debris was discarded after centrifugation and the clear cell lysate was loaded onto the resin. The resin was then washed (50 mM imidazole, 50 mM Tris/HCl, 150 mM NaCl, pH = 8.0) and the desired protein eluted (300 mM imidazole, 50 mM Tris/HCl, 150 mM NaCl, pH = 8.0). Imidazole was removed by dialysis against storage buffer (50 mM Tris/HCl, 150 mM NaCl, 1 mMDTT, 5% glycerol, pH = 8.0). The molecular weight of wild-type CthispH-RPS1 was confirmed with MALDI-TOF (Bruker UltrafleXtrem, Boston, MA) mass spectrometry. Purity of the proteins was checked with SDS-PAGE. Protein concentrations were determined by using the Bradford assay. The concentrations of the [Fe,S]4+ clusters were measured by using UV-Vis spectroscopy. The extinction coefficient at 410 nm was taken to be 15000 M⁻¹ cm⁻¹.[28]

Fe-S cluster reconstitution. The Fe,S₄ cluster of CthispH-RPS1 and its mutants was reconstituted by using purified *Azotobacter vinelandii* IscS protein. AvIscS was expressed from a plasmid which was the kind gift from Professor James A. Inlay. Typically, as-purified IspH-RPS1 was concentrated to <2 mL (~50-100 µM), degassed by bubbling nitrogen through the solution, transferred into an anaerobic chamber (Coy Labs, Grass Lake, IL), then stirred gently in the chamber overnight in order to equilibrate with the inert atmosphere. To the gently stirred protein solution was added 0.5 mM Fe(NH₄)₂(SO₄)₂, 2.5 mM L-cysteine, 5 mM DTT and ~0.1 µM IscS. The reaction was incubated for several hours until Fe₄S₄ incorporation was satisfactory, taken to be an A₄10/A₂₈₀ ratio of ~0.4. The resulting protein solution was then centrifuged to remove any precipitate, desalted by using a PD10 column (Agilent, CA) and concentrated as necessary. IscS was not removed from the sample since it represents only ~0.1% of the total protein.

Enzyme kinetics and inhibition assay. The CthispH kinetics and inhibition assays were performed by using the methyl viologen method described previously.[27] The final assay solution contained 40 nM enzyme, 50 µM HMBPP, 40 µM dithionite and 2 mM methylviologen. The reactions were monitored at 606 nm. Km and kcat were fitted by using a Michaelis–Menten relation in Origin 9.1. The extinction coefficient for the methyl viologen radical cation at 606 nm was taken to be 13700 M⁻¹ cm⁻¹. [17,17]

Electron paramagnetic resonance (EPR) spectroscopy. All sample preparation procedures were carried out in an anaerobic chamber. Reconstituted and desalted CthispH-RPS1 was concentrated to ~0.3-0.4 mM and 20 equivalents of Na2S2O5 and the ligand of interest, added. Samples (and no-ligand controls) were incubated for 30 min and then transferred into an EPR sample tube. EPR experiments were performed at X-band (9 GHz) using a Varian E-122 spectrometer together with an Air Products (Allentown, PA) helium cryostat. Spectra were obtained at 10K at a 1 mW power level.

Protein fingerprint analysis. The E. coli transcription-translation factor Rho was assigned by protein fingerprint analysis in the Protein Science Facility at the University of Illinois at Urbana-Champaign. Briefly, the sample containing the unknown protein was trypsin-digested and the resulting mixture analyzed by using electrospray ionization (ESI) tandem mass spectrometry in positive ion mode. The results were compared against a protein database using the Mascot program.[29]

Synthetic protocols. All reagents were purchased from Aldrich (Milwaukee, WI). The purity of compounds investigated were confirmed by 1H, 13C and 31P NMR spectroscopy at 400 MHz or 500 MHz performed on Varian ( Palo Alto, CA) Unity spectrometers. High-resolution mass spectra were obtained at the University of Illinois Mass Spectrometry Laboratory. Celullose thin layer chromatography (TLC) plates were visualized by using iodine or a sulphosalicylic acid-ferric chloride stain. The syntheses and characterization of 4-12 were described previously.[10]

4-Hydroxybut-2-ynyl-5-thiolodiphosphate (13) Tris-[(tetra-n-butylammonium)thiodiphosphate was prepared using a literature procedure[31]. 1-Chloro-4-hydroxy-2-butyne (52 mg, 0.5 mmol) in CH3CN (0.5 mL) was treated with 1.36 g (1.5 mmol) tris-[(tetra-n-butylammonium)thiodiphosphate in CH3CN (2 mL) at 0 °C. The reaction mixture was allowed to warm to room temperature over 8 h, then solvent was removed under reduced pressure. The residue was dissolved in 1 mL of cation-exchange buffer (49:1 (v/v) 25 mM NaH4HCO3/2-propanol) and passed over 90 equiv of Dowex AG50W-X8 (100-200 mesh, ammonium form) cation-exchange resin, pre-equilibrated with two column volumes of the same buffer. The product was eluted with two column volumes of the same buffer, flash frozen, and lyophilized. The resulting powder was dissolved in 1 mL of 50 mM NH4HCO3, 2-propanol/CH3CN (1: 1 (v/v), 2 mL) added, and the mixture mixed on a vortex mixer, then centrifuged for 5 min at 2000 rpm. The supernatant was decanted. This procedure was repeated three times and the supernatants were combined. After solvent removal and lyophilization, followed by flash chromatography on cellulose column (2:1:1 (v/v/v) 2-propanol/CH3CN/50 mM NaH4HCO3), a white solid 24 mg (yield 15%) was obtained. 1H NMR (400 MHz, D2O): δ: 4.05-4.03 (m, 2H), 3.44-3.40 (m, 2H); 13C NMR (125.7 MHz, D2O): δ: 82.46, 81.03, 49.89, 18.43; 31P NMR (161.9 MHz, D2O): δ: 7.83 (d, J = 26.4 Hz); -9.50 (d, J = 26.4 Hz); High-resolution mass spectrometry (HRMS, ESI) calculated for C14H20O4P2S [M+H]⁺, 262.9544; found 262.9543.
Results and Discussion

IspH-RPS1: A Rosetta Stone Protein Present in Many Gut Bacteria, and Some Human Pathogens. We first investigated the network of IspH family proteins by using the Enzyme Similarity Tool from the Enzyme Function Initiative (EFI-EST; http://efi.igb.illinois.edu/efi-est/) to generate a sequence similarity network for IspH family proteins, and visualized the results by using Cytoscape12-24. An overview of 15,660 IspHs is shown in Fig. 1. Most can be sorted into clusters based on the phylum and class of the organism. In a network with an expectation value of $10^{-10}$, the major clusters shown in Fig. 1 are as follows: a, Gammaproteobacteria (dark cyan, upper left) and Betaproteobacteria (blue, lower right); b, Alphaproteobacteria (cyan); c, Firmicutes (purple); d, Firmicutes (with very long sequences, red); e, Bacteroidetes (yellow); f, Actinobacteria (orange); g, Cyanobacteria and plants (green); h, Apicomplexa (pink, genus Plasmodium, the malaria parasites) and i, Desulfovibionales (brown, an order of Deltaproteobacteria).

What is of particular interest about these bioinformatics results is that, unlike most bacterial IspHs—which typically have ~300 residues (e.g. *E. coli*, N=316 residues, and *Aquifex aeolicus*, N=297 residues)—the IspHs found in cluster D (and a few isolated nodes) are much larger, with N=600-850 residues. In most cases this is due to a fusion with several ribosomal protein S1-like tandem repeats, found in the ribosomal protein S1 (RPS1) in the 30S subunit of the bacterial ribosome. The basic IspH-RPS1 architectures are shown schematically in Fig. 1j. Notably, essentially all of the organisms that have the IspH-RPS1 fusion are anaerobic bacteria. Among the 447 IspH-RPS1 sequences, 157 are from anaerobes. The other 290 are not annotated, but are primarily from anaerobic species. As shown in the zoomed-in view of the IspH-RPS1 network (Supplementary information, Fig. S1), the majority of the IspH-RPS1s are from Firmicutes, the low G+C gram-positives, including many members from the genus *Clostridium*. They include pathogenic species such as *C. tetani* and *C. botulinum* (the causative agents of tetanus and botulism, respectively), as well as the industrially valuable strain, *C. acetobutylicum*. A few IspH-RPS1s are from gram-negatives and typically contain 6 S1 repeats, Fig. 1j, one example being *Fusobacterium nucleatum*. This organism plays a role in periodontal disease as well as being associated with colorectal carcinoma35, 36 where it is involved in protecting tumors from immune system (NK cell) attack37. We also found that 37% of the IspH-RPS1s are present in the human gut, examples being *Clostridium* sp. SS2-1, *Eubacterium hallii*, *Subdoligranulum variabile*, *Clostridium leptum*, *Coprococcus eutactus*, *Clostridium* sp. L2-50, *Anaerotruncus colilominis*, and *Clostridium asparagusforme*, the first two organisms being the 7th and 9th most abundant gut bacteria38.

As noted in the book chapter by Pellegrini and Graeber25, the IspH-RPS1 fusion hybrids are thought to be examples of Rosetta stone proteins, functionally-related fusion proteins, rather than having arisen from random fusion events. However, there have been no proposals as to what the non-IspH function might be, rather, it is the high occurrence of IspH-RPS1 hybrids that is suggestive of functional relatedness.

IspH Forms Other Fusion Hybrids. Inspired by the observation of large numbers of IspH-RPS1 fusion proteins, we next examined the sequences and conserved domains of all IspHs in the sequence similarity network. In addition to the IspH-RPS1 fusions and the plant/protozoan IspHs that appear to have an N-terminal plastid-localization domain, there are five other types of IspH fusion annotated (Fig. 1k; Fig. 2). The most abundant fusion is between IspH and a putative prenyltransferase, typically annotated as a UbiA (4-hydroxybenzoate octaprenyltransferase) family protein. There were 47 such sequences found, primarily in the gram-negative sulfate-reducing bacteria Desulfovibionales (cluster i in the Cytoscape network, brown, Fig. 1i) in which the UbiA-like prenyltransferase domain was fused to the C-terminus of IspH, Fig. 1k. UbiAs form a superfamily of proteins, Fig. S2, typically thought to contain ~9 trans-membrane helices, and the X-ray structure of UbiA has recently been reported39. The function of the UbiA-annotated domain in the IspH-UbiAs, categorized as PT_UbiA_5, is unknown. However, it has been reported that ubiquinones are not detected in sulfate-reducing bacteria40; that menaquinone-7 is the major respiratory quinone in these organisms; and a separate UbiA/MenA is evident in their genomes. It seems likely, nevertheless, that the UbiA domain is involved in prenyl transferase activity, not least since there are two "DXXD" repeats present, typical of prenyl syntheses and transferases, and the fact that IspH appears to be fused to a second prenyltransferase would of interest.

![Fig. 1](image_url)
since it suggests that the fusion may be functionally significant—a true Rosetta stone protein. In addition, the hybrid protein would likely be membrane-bound. Four other types of fusion were found only in single species (or several strains of the same species): CMP kinase; helicases; fructose bisphosphatase and UDP/PPN phosphorylase (Fig. 2a–d). Plus, in a recent report, IspHs from a cyanobacterium and a plant, photosynthetic organisms, were proposed to have an extra N-terminal conserved domain, from photosynthetic organisms.

Organization and Structure Prediction for IspH-RPS1 and IspH-UbiA. We next sought to begin to investigate the possible structures of the IspH-RPS1 and IspH-UbiA hybrids. RPS1s from most gram-positive bacteria (e.g. *Bacillus subtilis*) contain 4 S1-like repeats, Fig. 1j (top), while RPS1s from gram-negatives (e.g. *E. coli*) typically contain 6 S1-like repeats, Fig. 1j (bottom). In *E. coli*, the first two repeats are used for ribosome binding, and the last four repeats for mRNA binding during translation-initiation. There have been no X-ray structures of any IspH-RPS1s reported. The structures of small bacterial and a protozoal IspH have, however, been reported, and by way of example, we show that of the *E. coli* protein in Fig. 3a, together with that of the *P. falciparum* protein (N=318, the apicoplast-targeting domain being absent), in Fig. 3b. Both the *E. coli* and *P. falciparum* structures are characterized by a trefoil/clover-leaf arrangement of α/β domains with the catalytic 4Fe-4S cluster at the center of each structure, linked to the protein via 3 Cys thiols. Substrate, inhibitor or water molecules occupy the remaining coordination sites. By way of comparison, a computational (Phyre2 program) structure prediction of the *Clostridium thermocellum* IspH-RPS1 protein is shown in Fig. 3c. The trefoil-like IspH domain is well defined, plus there are four more (and less) defined RPS1 domains (which are based on the solution NMR structures of individually expressed RPS1 domains).22

A Phyre2 structure prediction of the *Desulfobacula toloulica* (a marine, aromatic compound-degrading, sulfate-reducing bacterium) IspH-UbiA is shown in Fig. 3d. The UbiA domain is essentially that seen in the X-ray structure of *Aeropyrum pernix* UbiA (the template used by the Phyre2 program) and the 3 conserved Cys (yellow spheres) in the IspH domain and the two Asp rich (DXXXD) sites in the UbiA domain (purple spheres) can be readily identified. The UbiA domain is predicted to contain 9 trans-membrane helices, and the IspH domain presumably lies on the cytoplasmatic side of the membrane, as shown in Fig. 3d, although in most predicted structures this localization was variable.

In addition to the 3 essential Cys found in all IspH family proteins, involved in coordinating the 4Fe-4S cluster, IspH-RPS1s contain significantly more Cys than do the more conventional IspHs: 7.4 versus 4.8 on average (in the IspH domain), and some strains (*Clostridium sp.* CAG: 678; *Mogibacterium sp.* CM50) contained as many as 14 Cys. The high cysteine content is probably related to the observation that most of the species are anaerobes, since it has been pointed out that cysteine depletion occurs on a proteome-wide scale on moving from anaerobic to aerobic unicellular organisms. The range of Cys content in the IspHs is, however, much larger than the proteome-wide average.

At present, there is no information as to whether there are important functional consequences of the IspH-RPS1 (or other)
fusions. Thus, as a first step toward answering that question, we next sought to express an IspH-RPS1, the most abundant IspH fusion protein, in order to form a basis for future studies of structure and function.

Cloning, Expression, Purification and Catalytic Activity of IspH-RPS1. Since we did not find any other isph or rps1 genes in the organisms that harbor isph-rps1, it is almost certain that the IspH domain makes IPP/DMAPP, since mevalonate pathway enzymes are absent in the genome. We first sought to see if isph-RPS1 could be expressed and purified in an active form, catalyzing IPP/DMAPP formation from HMBPP. We attempted expression of IspH-RPS1s from three bacteria: F. nucleatum, C. acetobutylicum, and C. thermocellum. Initial trials with F. nucleatum and C. acetobutylicum resulted in low expression yields or inactive protein, although UV-Vis spectra clearly indicated the presence of a Fe-S cluster (Fig. 4a, [M+H]^+). However, there was poor iron-sulfur cluster incorporation, typically ~10%, even when co-expressed with the IscS protein48 that catalyzes cluster formation. We thus next used an in vitro reconstitution method with an expressed and purified IscS. The UV-Vis absorption spectrum of the reconstituted CthIspH-RPS1 protein at 410 nm, Fig. 4b, is consistent with the presence of an “oxidized” [Fe₄S₄]⁺ cluster. This could be reduced by dithionite to the reduced [Fe₄S₄]⁻ cluster, as found with other IspH proteins49.

We then examined the catalytic activity of CthIspH-RPS1 in the conversion of HMBPP into DMAPP and IPP by using the methyl viologen assay described previously27. As shown in Fig. 4c, the enzyme has a kcat of 56 min⁻¹ and a Km of 15 µM, at room temperature. The kcat is very close to the value found with IspH from the thermophile Aquifex aeolicus while the Km is 2-fold larger (kcat=1.0 s⁻¹; Kₘ=7 μM)7. Substrate consumption and product formation were also confirmed by LC-MS (Fig. S3) and NMR analysis. So, an IspH-RPS1 can be expressed from a thermophile and its activity is similar to that of IspH from another, “IspH-only” thermophile.

Truncation Mutants: Effects on Activity and on 4Fe-4S Cluster Binding. In order to find out whether the presence of the RPS1 domain had any effects on the catalytic activity of CthIspH-RPS1, we made truncation mutants containing 0, 1, 2 or 3 S1-like repeats (called here D0, D1, D2, D3; cartoon in Fig. 4d). As shown in Fig. 4e, these truncation mutants had almost the same activity as the wild-type CthIspH-RPS1 which contains 4 S1-like repeats, indicating that...
the RPS1 domain is not essential for IspH catalysis, at least in this in vitro assay.

We then investigated the 9 GHz EPR spectra of wild-type and truncation mutants of CthIspH-RPS1, Fig. 4f, to see if there were any obvious differences in cluster electronic structure, due to RPS1 binding. The wild-type CthIspH-RPS1 spectrum (Fig. 4f, pink) had g values of [2.022, 1.910, 1.896, 1.826] in the g=2 region, typical of a reduced S = 1/2 Fe/S cluster. In addition, an isotropic signal at g=5.0 was assigned to an S = 7/2 species (see Fig. S4 for a simulation), similar to a previously observed [Fe,S] cluster from benzoyl-CoA reductase. Higher-spin species are not as obvious in E. coli IspH (Fig. 4f, yellow), and the g values at g=2 were also slightly lower [1.998, 1.900, 1.873, 1.791]. The mutants all exhibited essentially the same EPR spectra as the wild-type C. thermocellum protein. We also made (though not by design) a very short C. thermocellum IspH (residues 1-171; “Cth171”) in which the C-terminal “leaf” of the cloverleaf was excised. Although inactive and missing the third essential Cys, the protein still bound a reducible Fe/S cluster and had an EPR spectrum very similar to that of the other IspHs, as shown in Fig. 4g. However, unlike the other unliganded IspHs, Cth171 exhibited just the strong signal in the g=2 region, and the spectra did not change upon ligand (3, 4, 5 and 8) binding, Fig. 4g and Fig. S5. Apparently, the 3 Cys residues are not essential for cluster binding, but are for ligand binding.

IspH-RPS1 Inhibition and a Comparison with Other IspHs. We next investigated the inhibition of CthIspH-RPS1 catalytic activity with the inhibitors (4-12) developed previously7, 30, as well as a new compound (13). The interest here is that it might be possible to develop compounds that specifically inhibit the F. nucleatum cell growth, reinstating, perhaps, the cytotoxicity of NK cells against colorectal carcinomas.35-37

Structures of the compounds tested are shown in Fig. 5a and consist of various alkyne and pyridine diphosphates and phosphonates. Representative dose-response inhibition curves are shown in Fig. 5b. These and other inhibition results are summarized in Fig. 5a together with Aquifex aeolicus IspH inhibition results. As a class, the acetylenes were the most potent inhibitors. The best inhibitor for CthIspH-RPS1 was the alkynyl compound 5, which had an IC₅₀ of 3.3 μM. Compound 5 was also the most potent alkyne inhibitor for AalsIspH, where it had the lowest IC₅₀, 0.45 μM. The pyridine-containing compounds were, in general, weak inhibitors. Interestingly, a novel alkynyl thioester compound (i.e. with a P=S bond), 13, was also very active against CthIspH-RPS1, compared to the ester compound 11. This thioester feature could potentially be useful in developing more drug-like leads that are resistant to hydrolysis, although pro-drugs will likely be needed for good cell penetration.

To investigate IspH-RPS1 ligand-protein interactions in more detail, we obtained 9 GHz EPR spectra of liganded (the acetylene 4, the pyridine 8, and the thiole-diphosphate 13), reduced CthIspH-RPS1, Fig. 5c. Upon binding of the alkyne 4 or 13, the signal from the higher-spin species in the unliganded spectrum disappeared and the signal intensity of the spin S = 1/2 species increased, with g values of [2.099, 2.012, 1.999] in the case of 13, similar to the values found with the alkylne inhibitors reported previously, bound to EclIspH and AalsIspH. Pyridine compounds such as 7-10 are another class of IspH inhibitors, but these all had only weak activity against CthIspH-RPS1, resulting in a significant amount of unliganded signal in the EPR spectrum of 8, despite the ligand being present in a 20-fold excess.

IspH-RPS1 Binds to E. coli Transcription Termination Factor Rho. A surprising observation during protein expression and purification was that when heterologously expressed in E. coli, CthIspH-RPS1 always co-purified with a protein of 46 kDa, as seen in the MALDI spectrum in Fig. 4a. This protein (Mᵦ, Fig. 4a) bound strongly enough to CthIspH-RPS1 that it co-purified with a CthIspH-RPS1 construct containing both a His-Tag and a Strep-Tag, in a two-step affinity chromatographic purification. The binding partner was subsequently found to be the E. coli transcription termination factor Rho through protein MS fingerprint analysis (Fig. 5a), with a sequence coverage ratio of 69%. The Rho factor is a helicase that unwinds the DNA-RNA junction during transcription termination and has a strong binding affinity for nucleotides.24, 52. It seemed possible that the binding between CthIspH-RPS1 and EchRho might also be mediated by nucleotide binding and indeed, we found that Rho could be dissociated from CthIspH-RPS1 by washing the Ni-NTA resin with a 0.5 mM (CT₅₅) DNA oligomer, a reported Rho binder in an X-ray study33 (Fig. 5b). It is also of interest that there is an actual (i.e., annotated) IspH-helicase fusion protein, in Bacteroidales bacterium. CF, Fig. 2b.

Possible Roles of IspH-RPS1 and IspH-UbiA. The roles of the IspH fusions in IspH-RPS1 and IspH-UbiA remain to be determined. What is, however, clearly of interest is that there are hundreds of anaerobes that contain IspH fusions, and some are pathogens or carcinogens. The very observation that IspH-RPS1 proteins exist supports previous ideas derived from work with E. coli (that lacks the RPS1 fusion) that the IspH domain might bind close to RelA when IspH binds to the ribosome, Fig. 6. Combining these results and ideas with the results of MS, EM and microbiological studies,17-19, 42 we hypothesize that the 4Fe-4S cluster might act as a “switch” to probe iron or oxygen levels, affecting RelA activity and the stringent response, a moonlighting role of a 4Fe-4S cluster containing protein that would be reminiscent of the fumarate-nitrate reductase regulatory protein, FNR19, or perhaps of aconitase, another 4Fe-4S protein that
Conclusions

Overall, the results reported here are of interest since they show that active IspH-RPS1 proteins can be expressed, and inhibited, of potential interest in the context of understanding the biology underlying IspH gene fusions, as well as in anti-infective and even anti-cancer drug discovery. There are many hundreds of organisms containing IspH-RPS1, about 37% being human gut bacteria. Some of these are in pathogens such as *Clostridium tetani*, *Clostridium botulinum*, and in *Fusobacterium nucleatum*, an oral pathogen that is also being linked to colon carcinoma. We also found that there are ~47 IspH-UbiA hybrids, all in sulfate-reducing bacteria, as well as smaller numbers of other IspH hybrids. We cloned, expressed and purified active IspH-RPS1 from *Clostridium thermocellum* and found that it had similar activity to IspH from another thermophile, *Aquifex aeolicus*. The IspH-RPS1 protein contains IspH fused to 4 RPS1 repeats: removal of 1, 2, 3, or 4 of these domains had no effect on IspH catalytic activity. We investigated the inhibition of C. thermocellum IspH catalytic activity (conversion of HMBPP to DMAPP and IPP) by a series of alkyne and pyridine diphosphates and related inhibitors finding the most potent activity with a novel alkyne thiol-diphosphate. The profiles of activity against both C. thermocellum and A. aeolicus were very similar. We also found that IspH-RPS1 bound to an *E. coli* protein, Rho, a helicase, and that there was evidence for an IspH-helicase fusion hybrid, in *Bacteroidales bacterium* CF. With IspH-UbiA, we were no successful in expressing full length active protein, but the observation of such hybrids is still intriguing given that both components are likely to be involved in isoprenoid biosynthesis. The IspH-UbiA hybrid is very likely to fit the functional definition of a Rosetta stone protein in which there are two domains that act in the same metabolic pathway. IspH makes DMAPP and IPP, and UbiA-like proteins utilize a diverse range of prenyl diphosphates. With IspH-RPS1, there are many fusion hybrids reported, but the functional-relatedness of the two domains is less clear. However, since RPS1 proteins bind close to the RelA binding site on the bacterial ribosome, a role for IspH-RPS1 in the bacterial stringent response is a possibility, consistent with previous suggestions of an IspH-RelA interaction, although further work will be required in order to fully understand the mechanism of action of both hybrids.

Acknowledgements

We are grateful to Professor John Gerlt for carrying out the first Cytoscape analysis of IspH. We thank Professors James Imlay and Hong Jin for helpful discussions, and Drs. Peter M. Yau and Brian S. Imai for carrying out the MS fingerprint analysis. We also thank Dr. Fuxing Zeng for assistance with the gel filtration experiments, and Dr. Katie Whalen for assistance with the sequence similarity network analysis. This work was supported in part by the United States Public Health Service (National Institutes of Health Grant GM065307) by a Harriet A. Harlin Professorship (E.O.), and by the University of Illinois Foundation/Oldfield Research Fund.

References


This Journal is © The Royal Society of Chemistry 20xx J. Name., 2013, 00, 1-3 | 7
