Correction

BIOPHYSICS AND COMPUTATIONAL BIOLOGY


The authors note that the following statement should be added to the Acknowledgments: “W.W. and J.L. were both supported by predoctoral fellowships from the American Heart Association, Midwest Affiliate (award 10PRE4430022 to W.W. and 11PRE7500042 to J.L.).”
IspG is a 4Fe4S protein involved in isoprenoid biosynthesis. Most bacterial IspGs contain two domains: a TIM barrel (A) and a 4Fe4S domain (B), but in plants and malaria parasites, there is a large insert domain (A*) whose structure and function are unknown. We show that bacterial IspGs function in solution as (AB)₂ dimers and that mutations in either both A or both B domains block activity. Chimeras harboring an A-mutation in one chain and a B-mutation in the other have 50% of the activity seen in wild-type protein, because there is still one catalytically active AB domain. However, a plant IspG functions as an AA*B monomer. We propose, using computational modeling and electron microscopy, that the A* insert domain has a TIM barrel structure that interacts with the A domain. This structural arrangement enables the A and B domains to interact in a “cup and ball” manner during catalysis, just as in the bacterial systems. EPR/HYSCORE spectra of reaction intermediate, product, and inhibitor ligands bound to both two and three domain proteins are identical, indicating the same local electronic structure, and computational docking indicates these ligands bridge both A and B domains. Overall, the results are of broad general interest because they indicate the insert domain in three-domain IspGs is close structural homology with the TIM (triose phosphate isomerase) barrel in dihydropteroate synthase, and the C terminus of the TIM barrel is only approximately 15 Å from the 4Fe4S cluster (in AaIspG; approximately 20 Å in TtIspG) as compared with the open mouth of the TIM barrel (A) in the second molecule in the dimer in such a way that the MEcPP substrate is trapped between the two domains, interacting with both. As can be seen in Fig. 1D, a ConSurf (10) analysis reveals many highly conserved residues (dark pink) in the TIM barrel and at the A/B interface in the dimer. IspG could, however, also in principle function as a monomer (or higher multimer), although the dimer model is more satisfying because the distance between the center of the TIM barrel is only approximately 15 Å from the 4Fe4S cluster (in AaIspG; approximately 20 Å in TtIspG) as compared with an approximately 50 Å distance to the TIM barrel in the protein folding | iron-sulfur | drug discovery | metalloprotein | metalloenzymes | structure | function and inhibition of the two- and three-domain 4Fe-4S IspG proteins

Yi-Liang Liu¹*, Francisco Guerra¹, Ke Wang, Weixue Wang, Jikun Li², Cancan Huang, Wei Zhu, Kevin Houlihan, Zhi Li³, Yong Zhang, Satish K. Nair, and Eric Oldfield

*Center for Biophysics and Computational Biology; ¹Department of Chemistry; and ²Department of Biochemistry, University of Illinois, Urbana, IL 61801; and ³Department of Chemistry, Chemical Biology, and Biomedical Engineering, Stevens Institute of Technology, Hoboken, NJ 07030

Edited by Arieh Warshel, University of Southern California, Los Angeles, CA, and approved April 3, 2012 (received for review December 20, 2011)

IspG [(E)-1-hydroxy-2-methyl-but-2-enyl-4-diphosphate (HMBPP) synthase; EC1.17.7.1, also known as GcpE] is a 4Fe4S cluster-containing protein (1) involved in isoprenoid biosynthesis in the methylerythritol phosphate pathway (2–4). It converts 2-C-methyl-D-erythritol-2,4-cyclo diphosphate (MEcPP) to HMBPP (2), which is then converted by a second 4Fe4S protein, IspH [(E)-1-hydroxy-2-methyl-but-2-enyl-4-diphosphate reductase; EC 1.17.1.2, also known as LgtB], to form isopentenyl diphosphate (3) and di-methylallyl diphosphate (4) in an approximately 5:1 ratio (3). IspG is found in most bacteria as well as in malaria parasites and plants and, because it is essential for pathogen survival and is not produced by humans, is of interest as a drug target. There are, however, two different types of IspG. In most bacteria, IspGs contain two domains: an N-terminal catalytic domain (A) and a C-terminal 4Fe4S reductase domain (B), as shown in Fig. 1A (5). In plant and in malaria parasite IspGs, the AB domains are again present, but there is also a third or insert domain, A*, located between the AB domains, as shown in Fig. 1B (5, 6). This insert domain is also found in some photosynthetic green sulfur bacteria, algae, mosses, as well as in numerous intracellular pathogens such as Babesia bovis, Theileria parva, Chlamydia trachomatis, and Leptospira interrogans, but the structure and function of this domain is not known.

A direct role in catalysis, that is, one involving processing of the MEcPP substrate, seems unlikely. As shown in Fig. S1, there are few conserved residues in the A* domain between 10 plant, malarial, and bacterial 3-domain IspGs; while in the 2-domain proteins, there are large numbers of conserved residues that are essential for catalytic activity, as shown via site-directed mutagenesis (7). This observation leads to the idea that the A* domain might play a more passive, structural role, linking the two (AB) “catalytic” domains together for catalysis. If this view is correct, then it would be predicted that all A* domains—even though they lack high sequence homology— would adopt a similar structure in order to link the A and B domains together. Likewise, it would be predicted that the EPR [as well as (ENDOR), electron nuclear double resonance and (HYSCORE), hyperfine sublevel correlation spectroscopy] spectra of the 2- and 3-domain proteins containing the reactive intermediate “X” (8) (as well as bound inhibitors) would be very similar because the A* domain is not directly involved in ligand bonding.

In recent work, X-ray crystallographic structures of IspG from Aquifex aeolicus and Thermus thermophilus (called here AaIspG and TtIspG) have been reported (7, 9) and these provide important clues as to the likely nature of the 3-domain proteins. The structure of the (AB) 2-domain protein AaIspG is illustrated in Fig. 1C and consists of a dimer of IspG molecules. There are two domains in each monomer, with the N terminus (A, blue) having close structural homology with the TIM (triose phosphate isomerase) barrel in dihydropteroate synthase, and the C terminus (B, orange) having close structural homology to sulfite/nitrite reductase (7, 9). Based on these results, it was proposed that the dimer (AB)₂ is the catalytically relevant structure, with the C terminus (4Fe4S cluster, B) of one molecule in the dimer interacting with the open mouth of the TIM barrel (A) in the second molecule in the dimer in such a way that the MEcPP substrate is trapped between the two domains, interacting with both. As can be seen in Fig. 1D, a ConSurf (10) analysis reveals many highly conserved residues (dark pink) in the TIM barrel and at the A/B interface in the dimer. IspG could, however, also in principle function as a monomer (or higher multimer), although the dimer model is more satisfying because the distance between the center of the TIM barrel is only approximately 15 Å from the 4Fe4S cluster (in AaIspG; approximately 20 Å in TtIspG) as compared with an approximately 50 Å distance to the TIM barrel in the
same chain, and can be “bent” to be even closer ([9], as seen in Fig. 1 E and F). These observations then raise two key questions: First, is the catalytic dimer model in fact correct? Second, if it is, might this structure suggest the role of the insert domain (A*) in the plant, malaria and other 3-domain proteins?

If the 3-domain IspG function as dimers, then the observation that the highly conserved catalytic residues are primarily found in the A and B domains would require an architecture along the lines of that illustrated in Fig. 1G, in which there would be no obvious role for the insert (A*) domain. If, on the other hand, the three-domain catalytically- active structures are monomeric, then a structural organization similar to that shown in Fig. 1H could be envisaged, in which the insert domain would play a similar “structural” role to the TIM barrels in the bacterial IspGs in which there is a large A/A contact patch in the (AB)2 dimer interface (7, 9) that stabilizes the protein’s structure. The AA* domain would thus act as a rigid rod with the B domain moving onto the rod to effect reduction of the reaction intermediate, then off of it to permit product release: a “cup-and-ball” model.

Results and Discussion

In the following, we focus first on the structure and function of the two-domain bacterial IspGs from Aquifex aeolicus and Thermus thermophilus, PDB ID codes 3NOY and 2Y0F. In particular, we wish to know if these proteins function in solution as monomers or dimers, because a better understanding of the structure and function of these proteins might lead to a better understanding of the structure and function of the 3-domain proteins. We then use a combination of spectroscopy and quantum chemistry to investigate ligand-binding to both 2- and 3-domain proteins—to determine if the same reactive intermediates and patterns of inhibition are found—probing functional equivalence. Finally, we use these results together with computational predictions and electron microscopy (EM) to develop a model for the A* insert domain, which we propose is a second TIM barrel, leading to an AA*B structure prediction that is consistent with the observation that a plant (Arabidopsis thaliana) IspG, AtIspG, functions as a monomer. And finally, we consider briefly possible reasons for the evolution of the 3-domain structure.

IspG Quaternary Structures. The 2-domain bacterial IspG proteins exist as dimers in the crystalline solid state, but we wished to determine if they exist as catalytically active dimers in solution, as well as whether the 3-domain enzymes function as monomers or dimers; this should help clarify the structural proposals illustrated in Fig. 1G and H. To test the hypothesis that a bacterial IspG (TtIspG) functions as a dimer, we carried out a series of experiments. First, we used glutaraldehyde cross-linking combined with SDS-PAGE and mass spectrometry to see if dimer formation occurs in solution. Cross-linking results for oxidized as well as reduced TtIspG (to see if there might be any difference in protein-protein interaction due to cluster reduction) are shown in Fig. 2A and B, together with results for lysozyme as a control (Fig. 2C). These results clearly show that most IspG molecules present in solution are cross-linked after a 30-min incubation with glutaraldehyde, independent of redox state, while no dimer formation is found with lysozyme, even after 12 h incubation, consistent with strong IspG dimer formation in solution, and only weak interactions in lysozyme. Mass spectrometry confirmed these results. Second, the results of gel filtration experiments (Fig. 2D and E) confirm that TtIspG exists in solution primarily as a dimer (MW = 94 kDa from experiment; 92 kDa expected for the dimer). Next, we need to answer the questions: is catalysis effected by monomers or by dimers in the bacterial enzyme? And if a dimer, is catalysis effected by individual molecules in the dimer (AB), or by both molecules in the dimer, (AB*)?

To answer these questions, we constructed TtIspG chimeras (AB*ΔB) consisting of two catalytically inactive mutants in which there was a mutation (m) that blocked activity in either the TIM barrel (A*ΔB) or the 4Fe4S cluster (ABmΔ), the idea being that even though neither monomers nor homo-dimers would be active, chimeras or hetero-dimers would be active, because the TIM barrel domain in a 4Fe4S cluster domain mutant (ABmΔ) could still interact with the 4Fe4S cluster domain in the TIM barrel mutant (A*ΔB). But which mutants to make? An obvious candidate for mutational analysis would be one of the Cys residues coordinated to the 4Fe4S cluster: we used C297S (6). We also need a TIM barrel mutant plus, ideally, an essential 4Fe4S domain mutant that does not prevent Fe-incorporation, because this might be more stable. To find suitable mutants, we used the SCORECONS program (11). We used a JPRED3 (12) alignment based on TtIspG as input to SCORECONS, finding (Table S1) that E232, in the TIM barrel (A), was the most essential residue, while N346, in the 4Fe4S cluster domain (B), ranked third overall (Table S1). We thus made three single mutants: E232A, C297S, and N346A, none of which were active (Fig. 2F, columns b, c, d), consistent with previous results for the equivalent E232 and C297 mutants in A. aeolicus (6, 7).

Initial attempts to generate chimeras or to facilitate hetero-dimer formation using a mixture of two mutants (E232A + C297S), in the presence or absence of detergents, yielded no activity, perhaps because homo-dimer formation was too tight. We thus coexpressed both mutants using His-tagged proteins, finding 52% (E232A/N346A) and 23% (E232A/C297S) activity (Fig. 2F, e and f). This result suggested that the catalytically active forms arise from the two individual mutants—which themselves are inactive, supporting the catalytic dimer model for the two-domain, bacterial IspGs.

In order to obtain pure chimeras containing a 1:1 ratio of each mutant, we next used a Duet vector to coexpress E232A (His)6/C297S (Strep-tag) and E232A (His)6/N346A (Strep-tag) proteins. After Ni-nitrilotriacetic acid followed by Strep-Tactin affinity chromatography (13), we obtained chimeras (AB*ΔB) containing a TIM barrel mutation in one chain and a 4Fe4S cluster-domain mutation in the second chain, as illustrated in Fig. 2G. The
two chimeras both exhibited approximately 50% activity (Fig. 2F, g and h). These results confirm that the TtIspG homo-dimer is the catalytically active species in solution, and form the basis for further experiments, with the A. thaliana protein.

With wild-type AtIspG, initial results of size-exclusion chromatography indicated protein aggregation and low catalytic activity, so we used size-exclusion chromatography to isolate nonaggregated wild-type AtIspG (Fig. 2H). The fraction of the protein that eluted as a monomer (MW 74 kDa from experiment, Fig. 2E, 82 kDa expected) was found to have good activity (approximately 8 μmol min⁻¹ mg⁻¹), comparable to that obtained with Escherichia coli IspG. In addition, there was no evidence for extensive dimer formation using glutaraldehyde cross-linking, the major effect being polymer formation (Fig. 2). Taken together, these results indicate that AtIspG exists and functions in solution as a monomer, consistent with an AA*B structure, Fig. 1H, while the bacterial 2-domain IspGs function as dimers.

**Protein-Protein and Protein-Ligand Interactions.** At present, there are no crystallographic structures of the closed conformation of any bacterial IspG, so how the MEcPP substrate [and the reactive intermediate X (2)], the HMBPP product, as well as how IspG inhibitors (2) bind, is uncertain. We thus next explored whether computational docking might help answer some of these questions, with the 2-domain AaIspG. We first used the RosettaDock program (14) to dock the B domain (orange) of AaIspG to the A domain (blue), converting the open conformation (Fig. 1 C and D) to the closed form (Fig. 1 E and F), then used Glide (15) to dock the MEcPP substrate (with or without Glu-350 bound to the 4Fe-4S cluster). The only docking poses returned were with MEcPP bound to the TIM barrel with electrostatic interactions between the MEcPP diphosphate and R101, K176 and R232, as shown in Fig. S2A. Based on the SCORECONS (11) conservation score ranking (Table S1), these residues are ranked as 20, 17, and 10 in terms of their conservation in a series of 310 IspG proteins. We then used Glide followed by molecular mechanics to generate docked poses of the putative ferrooxetane intermediate X (16), as well as of the bound HMBPP product, Fig. S2 B and C. With X, binding was to R55, R101, and N103; with HMBPP, binding was to R55, R101, R128, and K176, residues that are ranked 19, 20, 11, 10, and 17 in terms of their conservation score (Table S1). We then mutated R56, R141, and K204 in TtIspG (corresponding to R55, R128, and K176 in AaIspG) to Ala, finding in each case an approximately 95% reduction in catalytic activity, Table S1. R55, R101, and R128 have been shown to be essential for catalysis in A. aeolicus IspG by Lee et al. (7) and R141 (Thermus numbering) is essential for survival in Salmonella enterica (17). So, based on the docking results, their conserved nature (Table S1) and the mutagenesis results, these residues are all likely to be involved in MEcPP, reactive intermediate or HMBPP product binding, via their diphosphate groups, while the reactive intermediate and HMBPP product also interact directly with Fe via Fe-O and/or Fe-C bonds, the HMBPP-O-Fe interaction being similar to that seen with HMBPP binding to IspH (18).

**Spectroscopy and Quantum Chemistry.** We next investigated whether we could detect the reactive intermediate X that forms from MEcPP (2) with wild-type bacterial IspGs in the mutant and chimera bacterial proteins, as well as in the 3-domain IspG from A. thaliana, using EPR, ENDOR, and HYSCORE spectroscopy. With the C297S TtIspG protein, there is no 4Fe-4S cluster and hence, no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1. With the C297S TtIspG protein, there is no signal. With the E232A mutant, the (brown) sample has approximately 95% reduction in catalytic activity, Table S1.

**Fig. 2.** Oligomerization and catalysis of IspGs. (A) Cross-linking experiment with oxidized TtIspG. (B) Cross-linking experiment with reduced TtIspG. (C) Cross-linking experiment with lysosome. (D) Gel filtration chromatography and SDS-PAGE of TtIspG and AtIspG. Blue line: monomeric AtIspG; black line: reconstituted monomeric AtIspG; red line: dimeric TtIspG; green line: β-lactoglobulin; purple line: enolase; orange line: alcohol dehydrogenase; m: protein standard; 1: AtIspG; 2: TtIspG. (E) Molecular weight estimation of TtIspG and AtIspG. Red triangle: dimeric TtIspG (approximately 94 kDa); blue square: monomeric AtIspG (approximately 74 kDa); blue square: monomeric AtIspG; black line: reconstituted monomeric AtIspG; red line: dimeric TtIspG; green line: β-lactoglobulin; purple line: enolase; orange line: alcohol dehydrogenase; m: protein standard; 1: AtIspG; 2: TtIspG. (F) Activity measurement of TtIspG mutants and chimera. a: wild-type; b: E232A-(His)6; c: N346A-(His)6; d: C297S-(His)6; e: co-expression of E232A-(His)6 and N346A-(His)6 chimera; f: co-expression of E232A-(His)6 and C297S-(His)6 chimera; g: Duet expression of E232A-(His)6 and N346A-Strep tag chimera; h: Duet expression of E232A-(His)6 and C297S-Strep tag chimera. (G) Schematic of coexpressed/Duet TtIspG chimera. (H) High resolution gel filtration chromatography of AtIspG. Blue line: monomer AtIspG; red line: aggregated AtIspG. (I) Glutaraldehyde cross-linking experiment on AtIspG.
located in a ferraoxetane ring (16). Likewise, the $^{13}$C HYSSCORE spectra of [U-13C]-MeCP bound to both TspG chimeras (Fig. 4B and C) as well as to AtIspG (Fig. 4D) are essentially identical to those seen with EcIspG (2), again indicating the same reaction intermediate forms with both 2- and 3-domain proteins. There is also only a very small (approximately 0.15 MHz) hyperfine coupling observed with $^{13}$O1 labeled X (Fig. 4E), consistent with no direct Fe-O1 bonding.

Mechanistically then, it appears that MeCPP binds to the conserved Arg/Lys residues in the A domain and that the carboxylation that forms on ring opening is located very close to the 4Fe4S cluster domain, which then leads to formation of the reaction intermediate, X (2, 16). This intermediate is characterized by a large (approximately 12 MHz) $^1$H hyperfine coupling constant for one of the two methyl protons (16) not seen with an IspH reaction intermediate, H (2, 16). To see to what extent it might be possible to predict these spectroscopic observables, we used density functional theory (DFT) (19, 20). The structural model used was [Fe$_2$S$_2$(SM)$_2$](-C(=O)-H)-CH$_2$-CH$_2$-OH-O$^-$ (Fig. 5A) with S = 1/2. All calculations were performed with the Gaussian 09 program (19) using a Wachters' basis (62111111/31311111/3111) for Fe, 6-31G(d) for other heavy atoms, 6-31G(d) for hydrogens, and the BPW91 functional.

We found good accord with experiment for the H2, methyl protons, with computed hyperfine couplings of 9.1, 2.0, and 1.1 MHz for the three nonequivalent protons. This large coupling is in good accord with experimental results [A($^1$H) approximately 12 MHz] and arises from the trans (Fe-C-C-H torsion angle $\approx 172^\circ$) proton, while the gauche protons have much smaller couplings (and geometry optimized torsion angles of $S_2$, $-67^\circ$) — similar to the observation of large $^3$J trans scalar couplings (mediated via Fermi contact interactions) in NMR spectroscopy. Overall, there is a good correlation between the twelve computed hyperfine couplings determined with DFT and those determined experimentally, as shown in Fig. 5A; Table S2, although there is a systematic error in the slope, due most likely to basis/functional deficiencies. The correlation coefficient R = 0.87 is, however, very good. The computed spin densities are shown graphically in Fig. 5B and show, as expected, that significant spin density is transmitted from the cluster to the H2-trans proton (indicated with an arrow). The results of Atoms-in-Molecules (21) quantum chemical calculations (Table S3) indicate that binding to Fe is primarily electrostatic ($\nabla^2 \rho(r) > 0$) with partial covalence (21), just as found in simpler systems such as [FeF$_6$]$^{3-}$ and [Fe(CN)$_6$]$^{3-}$, Table S3.

**Inhibitor Binding.** We next briefly consider the topic of the inhibition of the bacterial and plant IspGs, of interest in the context of drug discovery as well as in the context of the equivalence of the
successful prediction of the three-helical where the use (24) of a variety of computational tools enabled computational methods. This approach has been successful with for the three-dimensional structures of the 3-domain IspGs using and experimental hyperfine couplings for H1(2), C2, H2, H3, H4(2), O1 and O3, based on experimental results in refs. 2, 16 and Fig. 4E. The line is constrained through the origin and \( R = 0.87 \), slope = 1.55. The inset shows the model used for performing the calculation. (B) Spin density for the ferrocene model. One methyl proton (indicated by the arrow) has a very large (approximately 12 MHz) coupling and originates from a trans (Fe-C2-C2-H2′) hyperfine interaction.

2- and 3-domain protein structures. In previous work, we investigated the inhibition of AalSpH (LytB) by a series of compounds, alkynes and pyridines linked to diphosphates or bisphosphonates (22), as well as two alkyne diphosphates that inhibited IspG (2), finding in both cases that the alkynes bound close to the 4Fe4S cluster. Here, we tested 25 compounds (Fig. S3) for IspG inhibition. As with IspH, only the alkynie diphosphates proved to be good IspG inhibitors. But how might they bind?

In the case of IspH, there is a highly conserved Ser-X-Asn site to which diphosphate (and malonate) ligands bind (23), but are there any highly conserved residues in the 4Fe4S domain in IspG to which the alkynie inhibitor diphosphates might bind? Based on the SCORECONS analysis (Table S1), the only highly conserved (non-cluster-bound) residues in the 4Fe4S cluster domain are N346 and R302 (corresponding to N303 and R270 in the A. aeolicus protein). When mutated to Ala, there is only 0.75% wild-type activity for N346A (Table S1), and 3.9% for R302. These residues are, however, approximately 10 Å distant from each other, and it is not possible to dock a diphosphate group to them both. So, rather than the diphosphates binding to the 4Fe4S cluster-containing domain (B) as in IspH, it appears likely that the diphosphates interact with the Arg/Lys-rich cluster in the TIM barrel. This idea is supported by the computational docking results shown in Fig. S2D in which the diphosphate group in an alkynie inhibitor (5, IC50 = 770 nM TtIspG; Fig. S3) binds to R55, R101, R128, K176, and R232. In addition, the results of a HYSCORE experiment with \(^{[13C]}\) propargyl diphosphate (6) bound to AalIspG (Fig. S4) show an approximately 6 MHz hyperfine coupling (as with EclspG), indicating that the side chain binds close to the 4Fe4S cluster, in the B domain, just as in the two-domain bacterial IspGs.

The Insert Domain (A*) Is a TIM Barrel. Finally, we propose a model for the three-dimensional structures of the 3-domain IspGs using computational methods. This approach has been successful with other prenyl syntheses, in particular with the diterpene cyclases, where the use (24) of a variety of computational tools enabled successful prediction of the three-helical αβγ domain structures, recently confirmed by two crystallographic investigations (25, 26).

The main questions of interest are: what is the structure of the insert domain, A*? And how might it interact with the N-terminal (TIM barrel, A) and C-terminal (4Fe4S cluster, B) domains? To provide possible answers to these questions, we first used several different protein structure prediction programs (27–31) to make predictions of the structure of the insert domain A*. Remarkably, in all cases, the structure predictions for the A* insert domain converged to the known N-terminal TIM barrel fold (A) found in the two bacterial IspG structures (7, 9). A representative structure (using a ClustalW2 multisequence alignment and Modeller) (32, 33) is shown in Fig. 6A for the A. thaliana (plant) IspG insert domain; results using other computer programs are shown in Fig. S5.

Clearly, the observation that all structure predictions of the insert domains indicate a TIM barrel strongly suggests that the plant (and other 3-domain) enzymes contain two TIM barrels: the catalytic N-terminal A, and the insert A*. In addition, based on the proposals (7, 9) that bacterial IspGs function with the C terminus (B) of one molecule in the dimer interacting with the N terminus (TIM barrel, A) of the second molecule in the dimer, and that the major contacts between the two AB molecules involve hydrophobic interactions between helices 7 and 8 in the two TIM barrels, it seemed likely that similar interactions might occur in the plant and other insert-domain-containing IspGs. Therefore, we next generated a simple three-dimensional model for IspG from A. thaliana, Fig. 6B, in which the structures of all three domains were generated computationally using Modeller (33), then aligned [in Molecular Operating Environment (34)] to the AalIspG structure. How A joins to A* and A* joins to B is not yet well defined. Using a ClustalW2 alignment of ten 3-domain IspGs, we then obtained the ConSurf prediction shown in Fig. 6C in which, as expected, the most highly conserved residues (dark pink) are at the AB interface while homology in the A* domain is low (white/blue, Fig. 6C). More importantly, this computational model is in good accord with electron microscopy (EM) results for AtIspG, constructed (35, 36) by aligning and combining two tomographic single particle reconstructions of negatively stained AtGcpE particles from low dose electron tomographic tilt series (Fig. S6), as can be seen in the EM (gray) superpositions shown in Fig. 6D.

If this structural arrangement is correct, it seems likely that the plant and other 3-domain IspGs may have originated via an initial gene fusion (A + B) to form a bacterial homo-dimer, (AB)2,
followed by duplication, exon loss, and recombination, eliminating one of the 4Fe4S clusters to form the (AA*B) proteins, as shown in Fig. 6E. But why? That is—why are there 3-domain IspGs? What is well known about IspG is that it is hypersensitive to oxygen (37, 38) and that when we inspect all of the organisms that employ 3-domain IspGs there is clearly a common feature—essentially all are subject to oxidative stress. In green plants, mosses, and algae, reactive oxygen species are generated during photosynthesis and can be expected to contribute to breakdown of the 4Fe-4S clusters. In the Apicomplexan parasites P. falciparum and B. bovis, the parasites infect red blood cells where there is expected to be oxidative stress (due to O₂ as well as other reactive oxygen species, ROS), and with bacteria such as C. trachomatis and L. interrogans, these pathogens again infect mammalian host cells and are expected to be susceptible to host generated ROS.

The possibility then exists that the additional TIM barrel may provide a degree of protection from host ROS by providing, e.g., a higher protein/iron ratio.

Conclusions

The results we have described above are of broad general interest because they clearly indicate, based on gel filtration chromatography, cross-linking, chimera formation, and spectroscopy, that most-bacterial IspGs function as an unusual (AB)₂ dimer in which a TIM barrel domain (A) of one chain interacts with the 4Fe-4S cluster-containing domain (B) in a second chain to form the active site. In the case of IspGs from plants, algae, malaria, and other parasites, the AB structural motifs are similar, but rather than a homo-dimer of four domains, we propose that there is the TIM barrel insert domain (A* ) that plays a primarily structural role. EPR, ENDOR, and HYSCORE spectra indicate the same reactive intermediate forms with both 2- and 3-domain enzymes, with this intermediate binding to the 4Fe4S cluster, during catalysis. The most potent inhibitors of both two and three domain enzymes, alkyne diphosphates, also bind close to the 4Fe4S cluster, as evidenced by similar EPR/HYSCORE spectra. This result indicates the same mechanism of inhibition, proposed to be inhibitor binding to a “closed” form of the enzyme with the inhibitors binding to both the A and B domains. Overall, these results should be of help in inhibitor design, a topic of considerable interest in the development of antiinfectives that inhibit isoprenoid biosynthesis in most pathogenic bacteria, as well as in malaria and other Apicomplexan parasites.

Methods

Pulsed ENDOR/HYSCORE spectra were obtained on a Bruker Elexys E 580-10 FT-EPR X-band EPR spectrometer using a Bruker RF amplifier (150 W, 100 kHz–250 MHz, for pulsed ENDOR experiments) and an Oxford instruments CF935 cryostat. Davies pulsed ENDOR used a three-pulse sequence \( (\pi/2, \tau, \pi, \tau, \pi) \); \( \pi/2 \) and \( \pi \) with RF applied during \( T \). HYSCORE used a four-pulse sequence \( (\pi, \tau, \pi, \tau) \); \( \pi = 48 \text{ ns} \), with \( \tau \) applied during \( T \).

Acknowledgments

We are grateful to Pinghua Liu for providing his E. coli IspG plasmid, to H. Jomaa and J. Wiener for providing their T. thermophilus IspG plasmid, to Manuel Rodriguez Concepcion for providing his A. thaliana IspG plasmid, and to Prof. D. Dean for providing pDB1281 and pDB1282 plasmids. We thank J. Niles for helpful discussions and assistance with the ENDOR and pulsed EPR experiments. This work was supported in part by the United States Public Health Service (National Institutes of Health Grant GM63507).

References