# Three-Dimensional Structure of Mannosyl-3-phosphoglycerate Phosphatase from *Thermus thermophilus* HB27: A New Member of the Haloalcanoic Acid Dehalogenase Superfamily

Susana Gonçalves, Ana M. Esteves, Helena Santos, Nuno Borges, and Pedro M. Matias*

ITQB-Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal

⑤ *Supporting Information*

ABSTRACT: Mannosyl-3-phosphoglycerate phosphatase (MpgP) is a key mediator in the physiological response to thermal and osmotic stresses, catalyzing the hydrolysis of mannosyl-3-phosphoglycerate (MPG) to the final product, α-mannosylglycerate. MpgP is a metal-dependent haloalcanoic acid dehalogenase-like (HAD-like) phosphatase, preserving the catalytic motifs I−IV of the HAD core domain, and classified as a Cof-type MPGP (HAD-IIB-MPGP family; SCOP [117505]) on the basis of its C2B cap insertion module. Herein, the crystallographic structures of *Thermus thermophilus* HB27 MpgP in its apo form and in complex with substrates, substrate analogues, and inhibitors are reported. Two distinct enzyme conformations, open and closed, are catalytically relevant. Apo-

Crystallographic snapshot of the trapped PO_3-like intermediate

Assisted acid catalysis step of the $D_N A_N$ phosphoryl-transfer mechanism

MpgP is primarily found in the open state, while holo-MpgP, in complex with the reaction products, is found in the closed state. Enzyme activation entails a structural rearrangement of motifs I and IV with concomitant binding of the cocatalytic Mg^{2+} ion. The closure motion of the C2B domain is subsequently triggered by the anchoring of the phosphoryl group to the cocatalytic metal center, and by Arg167 fixing the mannosyl moiety inside the catalytic pocket. The results led to the proposal that in *T. thermophilus* HB27 MpgP the phosphoryl transfer employs a concerted $D_N S_N$ mechanism with assistance of proton transfer from the general acid Asp8, forming a short-lived PO_3^- intermediate that is attacked by a nucleophilic water molecule. These results provide new insights into a possible continuum of phosphoryl transfer mechanisms, ranging between those purely associative and dissociative, as well as a picture of the main mechanistic aspects of phosphoryl monoester transfer catalysis, common to other members of the HAD superfamily.
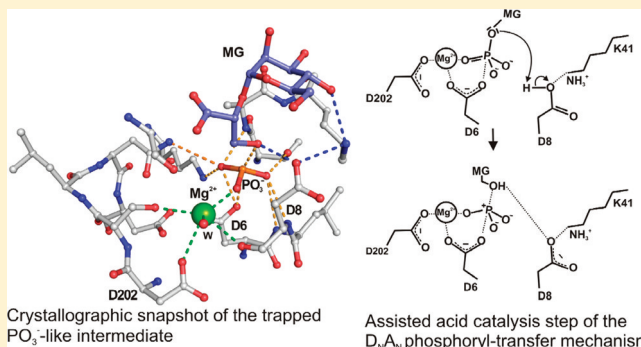
Studies of the strategies for thermo/osmo-adjustment present in many (hyper)thermophilic prokaryotes highlighted the role of organic compounds, designated compatible solutes, which are used by these organisms to control the internal water activity and protect intracellular macromolecules against a variety of stressful conditions.[1] In particular, mannosylglycerate [α-D-mannopyranosyl-(1,2)-O-D-glycerate (MG)] is a versatile compound used for cellular osmotic adjustment and thermal protection.[2−5] The structural elucidation of the enzymes taking part in the two known alternative biosynthetic pathways of MG (Figure 1) provided new insights into their rational classification and functional specialization, as well as the catalytic mechanisms at the molecular level.[6−8]

Mannosyl-3-phosphoglycerate phosphatase (MpgP, EC 3.1.3.70) is the enzyme involved in the second step of the two-step MG biosynthetic pathway, most commonly found in (hyper)thermophilic microorganisms,[9] and converts the phosphorylated intermediate mannosyl-3-phosphoglycerate (MPG) into MG as the final product. This enzyme shares sequence signature features with the mannosyl-3-phosphoglycerate phosphatase family members (MPGP; JCVI-CMR accession number TIGR01486) of the aspartate-nucleophile

haloalcanoic acid dehalogenase-like superfamily (HAD-like; SCOP version 1.75 accession number [56783]). The HAD-like superfamily includes Mg^{2+}-dependent hydrolases that catalyze the transfer of a carbon or phosphoryl group from a wide variety of substrates.[10] Members of this superfamily are topologically characterized by a three-layer α/β-sandwich having a central parallel β-sheet with strand order 54123, and by two consecutive structural motifs downstream of strand $\beta1$: a π-type helical turn known as a "squiggle" and a β-hairpin designated as a "flap". In addition, four conserved functional motifs are spatially arranged around the active site cleft in the HAD core region, shaping an electrostatic environment that stabilizes negatively charged transition-state (TS) species during the nucleophilic substitution.[10−13] Cap modules are the result of independent structural insertions at specific regions of the HAD core unit and comprise three main types: the four-helix bundle C1 cap, the $\alpha+\beta$ fold C2 cap, and the C0 cap corresponding to the single HAD core domain. These are
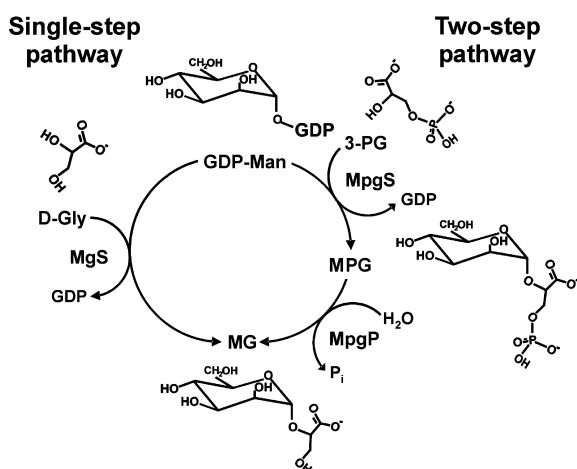
Figure 1. Two pathways for the synthesis of mannosylglycerate. In the single-step pathway (left), MgS catalyzes the direct glycosyl transfer of GDP-Man into D-Gly. In the two-step pathway (right), in the first step, MpgS catalyzes the transfer of the mannosyl moiety from GDP-Man to 3-PG, yielding MPG; in the second step, this intermediate compound is hydrolyzed by MpgP to form the final product MG. Nomenclature: MgS, mannosylglycerate synthase (EC 2.4.1); MpgS, mannosyl-3-phosphoglycerate synthase (EC 2.4.1.217); MpgP, mannosyl-3-phosphoglycerate phosphatase (EC 3.1.3.70); D-Gly, D-glycerate; 3-PG, 3-phosphoglycerate; MPG, 2-(α-D-mannosyl)-3-phosphoglycerate; GDP, guanosine diphosphate; GDP-Man, α-D-GDP-mannose; P_i, inorganic phosphate; MG, α-mannosylglycerate. Published in ref 8. Copyright 2010 American Society for Biochemistry and Molecular Biology.

further classified into classes and subclasses, which increase in complexity as a function of the diversity of substrates and reaction systems.[10]

The MPGP falls within the predicted Cof-type phosphatase class and is characterized as having a C2-type B cap module inserted within the cross-linker region between strands $\beta3$ and $\beta4$ of the basal HAD core domain (HAD-IIB-MPGP; SCOP accession number [117505]). Members of the Cof-type class are known to use sugar phosphates as substrates, and many may also be involved in the regulation of stress response mechanisms, such as the Cof-type stress response YhaX from *Bacillus subtilis*[14] and the Cof-related trehalose-6-phosphate phosphatase from *Thermoplasma acidophilum*.[15] To date, the MPGP hydrolase family is structurally represented by only the MpgP from *Pyrococcus horikoshii* in its closed conformation[7] [Protein Data Bank (PDB) entry 1wzc] and by the open conformation of the MpgP-related YedP from *Escherichia coli* of unknown function (PDB entry 1xvi; Kim et al., unpublished observations) but known to interact with inner membrane protein YedI.[16]

Remarkably, the catalytic core for the phosphoryl monoester transfer reaction is preserved among the HAD family members, regardless of the nature of the substrate. Still, the detailed understanding of the reaction coordinate landscape in these metal-dependent phosphate monoester hydrolases is currently under debate.[13,17−19] Whereas the structural data suggest an associative mechanism, comprising a two-step addition−elimination reaction (IUPAC A_N + D_N) with coupled acid/base catalysis,[12] nonenzymatic phosphoryl monoester hydrolysis supports a dissociative S_N1-type mechanism (IUPAC D_N + A_N) with coupled acid/base catalysis.[20,21] HAD core motifs I−IV effectively modulate a trigonal-bipyramidal electrostatic environment, hinting at the first model by favoring an

equatorial expansion and/or polarization at the nonbridging oxygen atoms of the phosphorane intermediate.[11−13] However, conflicting evidence that supports a dissociative mechanism arises from the observation of a trapped PO_3^- intermediate in the crystal structure of fructose 1,6-bisphosphatase,[22] and from the measured kinetic isotope effects in other monoester phosphatases, which support the formation of a PO_3^- resonance species in the first reaction step.[16,17,23] The latter studies suggested a high bond breaking order for the leaving group (β_{lg}) with a low bond order for the attacking nucleophile (β_{nuc}) and thus were explained on the basis of a concerted mechanism with higher dissociative character.[17] An integrated perspective of these two mechanisms was given by the theoretical calculations conducted with representative members of the HAD-like C1 and C2 classes.[19,24] The transfer of the PO_3^- species is concerted and should occur with a geometric associative but an electronic dissociative character, coupled with a proton transfer. Thus, the mechanism of phosphoryl transfer catalysis depends on the nature of both the electrophile and the nucleophile, as well as the enzyme's ability to stabilize the transition-state species, and a continuum of possible mechanisms between those purely associative and dissociative kinds may exist.[17,21]

Herein, we describe the crystal structure of *Thermus thermophilus* HB27 MpgP, an aspartate nucleophile HAD-like hydrolase, which belongs to the MPGP. This work completes the structural characterization of the enzymes involved in the synthesis of MG in this organism, as the structure of mannosyl-3-phosphoglycerate synthase (MpgS) was recently determined by our team.[8] Crystal soaks with substrates, substrate analogues, and inhibitors allowed us to probe and rationalize the relative positions of the Cap and HAD core regions as crystallographic snapshots of phosphoryl transfer. In MpgP, phosphoryl transfer catalysis occurs via a concerted D_NA_N mechanism, assisted by a coupled proton transfer from Asp8 to the ester oxygen. Furthermore, the functional role of the conserved salt bridge between Asp8 and Lys41 in promoting proton flux during P−O bond fission is discussed in light of a versatile catalytic HAD core that is able to synchronously modulate the hydrogen bonding network and the electrostatic potential for optimal positioning of the PO_3^- intermediate with respect to the attacking nucleophile.

The results presented may assist in resolving prior inconsistencies, as well as provide a picture of the main mechanistic aspects of phosphoryl monoester transfer catalysis, common to other members of the HAD superfamily.

■ MATERIALS AND METHODS

Expression, Purification, and Crystallization of *T. thermophilus* HB27 MpgP. The details of the expression, purification, and crystallization of apo-MpgP have been previously described.[25] Briefly, the *T. thermophilus* HB27 *mpgP* gene was cloned in a pKK223-3 vector (PL-Pharmacia; GenBank entry M77749.1) and expressed in *E. coli* BL21-(DE3), and the gene product was purified by ionic exchange chromatography. Pure and stable protein eluted as a single peak from a MonoS column (GE Healthcare) and was kept in its elution buffer [20 mM MES-NaOH (pH 6.3), 760 mM NaCl, 5 mM DTT, and 1 mM EDTA]. Prior to crystallization, the protein solution was supplemented with 2 mM fresh DTT and concentrated to 10 mg/mL. $MgCl_2$ (5 mM) and Na/KPO_4 were then added to the concentrated protein solution as additives. For crystals soaked with gadolinium(III) chloride hydrate

($GdCl_3 \cdot xH_2O$), a 2 mM lanthanide solution was used instead of $MgCl_2$ during cocrystallization. Optimized crystallization conditions were obtained by using a crystallization solution as described by Gorrec:[26] each of the carboxylic acids at 0.02 M (Na-formate, NH_4-acetate, Na_3-citrate, NaK-L-tartrate, and Na-oxamate), 0.1 M MES/imidazole (pH 6.5) (in a 1:1 molar ratio), and a 30% (v/v) precipitant mixture containing 20% (v/v) ethylene glycol and 10% (v/v) polyethylene glycol 8000 (30% EDO-PEG8K). Crystals were obtained at 293 K by the sitting drop vapor diffusion technique in 24-well crystallization plates. Each crystallization drop was prepared by mixing 1 μL of crystallization solution with 1 μL of protein solution and equilibrated against 500 μL of crystallization solution in the reservoir. Prism-shaped crystals developed within 5 days.

Crystal Soaking Procedure. The MpgP crystal used for phasing, i.e., the [Gd^{3+}] data set, was gradually transferred through six different stabilizing solutions with chemical compositions similar to that used for protein crystallization, but where the carboxylic acid (CarAc) mixture was gradually replaced with $GdCl_3 \cdot xH_2O$ (Table S1 of the Supporting Information). The removal of the CarAc mixture was conducted as an optimization strategy to increase the yield of bound ligand, by preventing the adventitious metal chelation by some of these acids. MES-NaOH buffer (pH 6.5) was used instead of MES/imidazole buffer (pH 6.5), and the final soaking solution contained 35% (v/v) precipitant mixture (EDO-PEG8K) to ensure proper cryoprotection. The final soaking with Gd^{3+} was performed at 293 K over a period of 14 h (see Table S2 of the Supporting Information), and the crystals were subsequently flash-cooled in liquid nitrogen prior to data collection.

The crystal soaks with substrates with or without inhibitors followed a procedure similar to that described above, but these ligands were added only in the final two crystal transfer solutions. As control experiments, one pseudo-Michaelis ternary complex and another mirroring the pentavalent TS species of the reaction coordinate were prepared by soaking crystals with MG and either orthophosphate or orthovanadate analogues, respectively. The composition of the soaking solutions and the final ligand concentrations, as well as the duration of the soaks, are described in Table S2 of the Supporting Information for each data set. Hereafter, each data set will be designated according to the soaking conditions used. All the soaks were conducted at 293 K. Stock solutions used for soaking were 0.5 M $GdCl_3 \cdot xH_2O$ (Sigma-Aldrich) in MES-NaOH (pH 6.5), 0.5 M $VOSO_4 \cdot 5H_2O$ (Merck) in water, 0.2 M Na_3VO_4 (Sigma-Aldrich) at pH 10.0 in water, 0.8 M $MgCl_2$ (Sigma-Aldrich) in water, 0.1 M α-D-mannosylglycerate (MG) in water, and 0.019 M α-D-mannosyl-3-phosphoglycerate (MPG) in water. The MPG substrate was produced with the *T. thermophilus* HB27 MpgS enzyme purified as previously described.[27] The experimental conditions applied for reaction and isolation of MPG were those established by Empadinhas and co-workers,[28] with slight modifications; MPG-eluted fractions from the QAE-Sephadex A25 column were detected by thin liquid chromatography (TLC) and analyzed by ^{1}H NMR. The MG substrate was obtained as previously described by Borges and co-workers.[29]

Data Collection and Processing. A highly redundant single-wavelength anomalous dispersion (SAD) data set to 2.5 Å resolution was measured in house with Cu Kα radiation (λ = 1.5418 Å) at 100 K from a crystal soaked with Gd^{3+} ([Gd^{3+}] data set), using a Bruker-AXS Proteum Pt135 CCD

detector system coupled to a Bruker-AXS Microstar-I rotating anode X-ray generator with Montel mirrors. The data were integrated with SAINT and scaled with SADABS as part of the Bruker-AXS Proteum Software Suite. Diffraction data statistics were obtained with XPREP (Bruker-AXS). A summary of the data collection and processing statistics is given in Table 1. The

Table 1. Crystallographic Data Collection, Processing, and Phase Refinement Statistics for the [Gd^{3+}] Data Set

Data Collection and Processing Statistics	
wavelength (Å)	1.5418
space group	$P2_1$
unit cell parameters	a = 39.65 Å, b = 70.77 Å, c = 92.68 Å, $β$ = 95.05°
resolution range (Å)	39.49−2.50 (2.59−2.50)
scan type	$ω$-scan
total angular range (deg)[b]	518.6
total no. of frames[b]	2593
exposure time per frame (s)	90
no. of observations	195118 (12464)[a]
no. of unique reflections	17882 (1833)[a]
$\langle I/\sigma(I) \rangle$	17.2 (4.7)[a]
R_{merge} (%)[c]	10.3 (36.5)[a]
R_{pim} (%)[d]	3.2 (13.5)[a]
completeness (%)	99.7 (98.2)[a]
multiplicity	10.9 (8.1)[a]
Wilson B (Å²)	38.0
no. of molecules in the asymmetric unit	2
V_m (Å³/Da)[e]	2.31
estimated solvent content (%)[e]	46.7
Phase Refinement Statistics	
phasing power, anomalous	1.553
anomalous R_{cullis}	0.692
SHARP figure of merit, acentric	0.467
SHARP figure of merit, centric	0.133
SHARP figure of merit, overall	0.450
Density Modification Statistics	
overall $|E^2|$ correlation[f]	0.732
figure of merit after final DM run[f]	0.825
figure of merit after NCS averaging[g]	0.786

[a]Values in parentheses are those for the highest-resolution shell. [b]In eight different crystal settings. [c]R_{merge} is the merging R factor and equals $\left[\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| \right] / \left[\sum_{hkl} \sum_i I(hkl) \right] \times 100\%$. [d]$R_{pim}$ is the precision-independent R factor and equals $\sum_{hkl} [1/(N_{hkl} - 1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl) \times 100\%$, where I is the observed intensity, $\langle I \rangle$ is the average intensity of multiple observations from symmetry-related reflections, and N_{hkl} is their multiplicity.[64] [e]According to ref 32. [f]From the SHARP optimizing density modification procedure (SOLOMON followed by the final DM run).[37] [g]From the second DM run.[38]

remaining data sets were all collected at 100 K at ESRF (Grenoble, France) ID14-4 and ID23-1 beamlines, using an ADSC Quantum Q315r detector except for the [Mg^{2+}-MG-VO_4^{3-}] data set, which was collected at DIAMOND (Didcot, U.K.) beamline I03, using a Pilatus 6M detector. Diffraction

images were processed with the XDS Program Package,[30] and the data collection statistics are summarized in Table 2. The data were further processed for use with the CCP4 Program Suite.[31] Matthews coefficient calculations[32] suggested the presence of two molecules in the asymmetric unit of all crystal structures, with values close to those calculated for the $[Gd^{3+}]$ data set ($V_m = 2.31$ Å^3/Da, and a predicted solvent content of 46.7%).

Structure Solution and Crystallographic Model Building. Using the HKL2MAP[33] graphical user interface, the $[Gd^{3+}]$ data set was scaled and analyzed with SHELXC (G. Sheldrick, personal communication), the Gd^{3+} heavy atom substructure determined with SHELXD,[34] and the phase problem solved with SHELXE.[35] The best solution from SHELXD in 100 trials gave six Gd^{3+} sites with a correlation coefficient of 46.9%, four of them with a relative occupancy that was approximately half of that of the top two. The SHELXE calculations gave a clear discrimination between the correct and inverted substructure solutions, and the phases derived from the SAD data were improved using the maximum-likelihood heavy-atom parameter refinement in autoSHARP.[36] The autoSHARP calculations found five additional sites, and a subsequent optimizing density modification procedure using SOLOMON[37] suggested a solvent content of 40.5% and two monomers in the asymmetric unit. Centroid SHARP phases were further improved by density modification with DM[38] with 2-fold noncrystallographic symmetry averaging (NCS). The initial NCS operator was obtained with PROFESSS from the top six Gd^{3+} sites. A random 5% sample of the reflection data was used for R_{free} calculations[39] during model building and refinement. Using the 2.5 Å density-modified phases from DM, 463 of the expected 518 protein residues in the asymmetric unit were built and sequenced automatically with Buccaneer/REFMAC,[40,41] leading to R and R_{free} values of 0.305 and 0.351, respectively. The model was then corrected and completed using Coot[42] prior to refinement. The main statistics from the autoSHARP/SOLOMON/DM calculations are included in Table 1.

The structures of apo-MpgP and the crystals soaked with substrates and inhibitors were determined by molecular replacement with PHASER[43] using the refined protein chain coordinates of the $[Gd^{3+}]$ crystal structure.

Crystallographic Refinement. The structures were refined using the amplitude-based Maximum-Likelihood target function with automatic weight optimization procedure as implemented through the graphics user interface of PHENIX version 1.6.1.[44] In general, the first cycle comprised a rigid body refinement followed by individual coordinate and isotropic atomic displacement parameter refinement. In the subsequent refinement stages, TLS (translation-libration-screw) rigid body refinement of atomic displacement parameters was conducted, followed by refinement of individual isotropic B factors, in addition to individual atomic coordinates. In the case of the $[Gd^{3+}]$ structure, and because of the lower data resolution, group isotropic B factors were refined for the protein chains instead of individual isotropic atomic displacement parameters (ADPs). Two groups were defined per protein residue, one for the main chain atoms and the other for the side chain atoms of each protein chain residue. For each crystal structure, four TLS groups per monomer were parametrized using the TLS Motion Determination (TLSMD) server.[45] The definition of the TLS groups comprised the segmentation of the two protein domains: the first, second, and fourth groups roughly

corresponded to the HAD-like domain, whereas the third contained at least part of the cap C2B domain (see Structure of MpgP in Results for the definitions of the HAD-like and cap C2B domains). Whereas the first and fourth TLS groups had similar boundaries for all structures, the separation between the second and third TLS groups varied. For example, while in chain B of the $[Gd^{3+}]$ crystal structure the third group closely matched the cap C2B domain, in the control $[Mg^{2+}\text{-MG-}HPO_4{}^{2-}]$ structure, part of this domain was included in the second TLS group. The inclusion of NCS restraints did not improve the refinement statistics or the quality of the model. Monomer library descriptions for new ligands were created with LIBCHECK[46] through the Sketcher interface implemented in the CCP4i Graphical User Interface.[47] Ligand and metal coordination geometry restraints were further reviewed with REFMAC5.[41] During refinement, the models were periodically inspected and corrected with Coot[42] against SigmaA-weighted $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron density maps. Inclusion of ordered solvent molecules was conducted with ARP/wARP,[48] followed by inspection in Coot.[42] Except for the $[Gd^{3+}]$ structure, in the final refinement, automatic refinement of water occupancy and constrained group occupancy refinement for nonprotein atoms were included in the refinement protocol. The stereochemical quality of the model was assessed with MolProbity.[49] The models were almost complete, varying in length after residue Arg250 except for the apo-MpgP crystal structure that presented poor electron density at residues Asp13−Leu18 in chain B; these residues were thus not included in the model. A summary of the refinement statistics, model composition, and stereochemical quality is presented in Table 3. Hereafter, each structure will be designated according to the soaking conditions used.

■ RESULTS

Structure of MpgP. MpgP crystallized as a homodimer as previously reported,[25] with each monomer related by a 2-fold noncrystallographic symmetry (NCS) axis, in both monoclinic and orthorhombic crystal structures. In the monoclinic crystal structures, the 2-fold NCS is oriented perpendicularly to the unique b-axis, and the homodimers form a right-handed helical assembly along the crystallographic c-axis. This homodimer arrangement is not present in the orthorhombic crystal structure. A secondary structure matching against all PDB, conducted with DALI[50] and SSM,[51] and using the coordinates of chain A from the apo-MpgP and chain B from the control $[Mg^{2+}\text{-MG-}HPO_4{}^{2-}]$ structures, confirmed the expected close relationship to the members of the predicted hydrolase Cof family (SCOP version 1.75 accession number [82388]) of the haloalcanoic acid dehalogenase superfamily (HAD-SF). Proteins from this family share a C2B cap module characterized by an open-face $\alpha+\beta$ unit with at an least three-stranded anti-parallel β-sheet flanked by two oppositely oriented α-helices.[52] The most similar protein structures identified with both search engines were those sharing the mannosyl-3-phosphoglycerate phosphatase protein domain (putative HAD-SF-IIB-MPGP, SCOP version 1.75 accession number [117505]), which includes the orthologous MpgP from *P. horikoshii* (PDB entries 2zos and 1wzc[7]) and the MpgP-related YedP from *E. coli* K12 (PDB entry 1xvi). All three proteins share the same fold, supporting the inclusion of *T. thermophilus* HB27 MpgP in this family and suggesting divergence from a common ancestor. Other structural relatives from the predicted hydrolase Cof family included members of the phosphoglycolate phosphatase

Table 2. Data Collection and Processing Statistics[a]

	apo-MpgP	[Mg²⁺-MPG] crystal 1	[Mg²⁺-MPG] crystal 2	[Gd³⁺-MPG] crystal 1	[Gd³⁺-MPG] crystal 2	[MPG-VO²⁺]	[Mg²⁺-MG-HPO₄²⁻] (control)	[Mg²⁺-MG-VO₄³⁻] (control)
beamline	ID14-4	ID14-4	ID14-4	ID14-4	ID14-4	ID23-1	ID14-4	I03
wavelength (Å)	0.9535	1.2520	0.9535	0.9535	0.9535	0.9792	1.2520	0.97625
space group	$P2_1$	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
cell parameters	$a = 39.52$ Å	$a = 39.93$ Å	$a = 39.37$ Å	$a = 39.79$ Å	$a = 39.87$ Å	$a = 39.97$ Å	$a = 39.90$ Å	$a = 39.95$ Å
	$b = 70.68$ Å	$b = 70.53$ Å	$b = 88.05$ Å	$b = 70.80$ Å	$b = 70.24$ Å	$b = 71.36$ Å	$b = 70.74$ Å	$b = 71.26$ Å
	$c = 95.42$ Å	$c = 92.63$ Å	$c = 146.46$ Å	$c = 97.77$ Å	$c = 91.78$ Å	$c = 92.97$ Å	$c = 92.30$ Å	$c = 91.22$ Å
	$\beta = 92.95°$	$\beta = 95.70°$		$\beta = 95.54°$	$\beta = 95.68°$	$\beta = 95.46°$	$\beta = 95.37°$	$\beta = 95.52°$
resolution (Å)	50.00–1.90 (2.01–1.90)	50.00–1.90 (2.00–1.90)	60.00–1.80 (1.91–1.80)	60.00–1.90 (2.01–1.90)	60.0–1.92 (2.03–1.92)	60.0–2.10 (2.23–2.10)	60.00–1.74 (1.79–1.74)	30.0–1.81 (1.92–1.81)
no. of observations	148324 (18433)	114899 (14881)	200217 (24572)	136161 (21687)	129380 (17654)	102785 (14554)	174136 (8286)	158322 (25502)
no. of unique reflections	40742 (5910)	39139 (5682)	46863 (6983)	39808 (6335)	37993 (5760)	30243 (4721)	50839 (3496)	46163 (7353)
completeness (%)	98.0 (88.6)	94.8 (85.7)	97.8 (91.9)	98.0 (97.3)	98.0 (92.9)	99.0 (96.5)	96.8 (91.1)	99.2 (98.8)
multiplicity	3.6 (3.1)	2.9 (2.6)	4.3 (3.5)	3.5 (3.4)	3.4 (3.1)	3.4 (3.1)	3.4 (2.4)	3.4 (3.5)
$\langle I/\sigma(I)\rangle$	18.79 (2.37)	7.19 (1.93)	19.29 (3.67)	23.79 (4.92)	20.28 (3.74)	16.78 (4.47)	11.81 (3.06)	8.99 (1.92)
R_{merge} (%)[b]	4.4 (48.1)	13.0 (73.2)	4.5 (32.4)	3.1 (26.1)	3.7 (27.5)	5.1 (28.8)	5.9 (29.1)	7.9 (54.1)
R_{meas} (%)[c]	5.2 (57.9)	15.1 (72.2)	5.1 (37.9)	3.6 (31.0)	4.4 (33.1)	5.9 (34.8)	6.5 (37.1)	9.3 (63.8)
Wilson B (Å²)	36.5	37.4	32.7	37.3	35.7	38.2	35.1	34.3

[a]Values in parentheses are those for the highest-resolution shell. [b]R_{merge} is the merging R factor and equals $[\sum_{hkl}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|]/[\sum_{hkl}\sum_i I(hkl)] \times 100\%$. [c]$R_{meas}$ is the multiplicity-independent R factor and equals $\sum_{hkl}[N_{hkl}/(N_{hkl} - 1)]^{1/2}\sum_i|I_i(hkl) - \langle I(hkl)\rangle|]/\sum_{hkl}\sum_i I_i(hkl) \times 100\%$, where I is the observed intensity, $\langle I\rangle$ is the average intensity of multiple observations from symmetry-related reflections, and N_{hkl} is their multiplicity.[64]

Table 3. Final Refinement Statistics

	[Gd³⁺]		apo-MpgP		[Mg²⁺-MPG] crystal 1		[Mg²⁺-MPG] crystal 2	
resolution limits (Å)	37.56–2.50 (2.66–2.50)		39.51–1.90 (1.97–1.90)		20.85–1.90 (1.97–1.90)		42.16–1.80 (1.87–1.80)	
R factor[a]	0.195 (0.228)		0.173 (0.242)		0.182 (0.293)		0.180 (0.215)	
no. of reflections[b]	16926 (2817)		38647 (3298)		36697 (3363)		44507 (3947)	
R_{free}[c]	0.253 (0.286)		0.228 (0.303)		0.238 (0.317)		0.230 (0.273)	
no. of reflections[b]	908 (140)		2073 (175)		1950 (193)		2342 (209)	
overall coordinate error estimate (Å)[d]	0.34		0.28		0.24		0.24	

Model Completeness and Composition

	chain A	chain B	chain A	chain B	chain A	chain B	chain A	chain B
model regions omitted	252–259	252–259	253–259	13–18, 254–259	251–259	255–259	251–259	255–259
Gd³⁺ ions	3	3	–	–	–	–	–	–
Cl⁻ ions	1	1	–	–	–	–	–	–
Mg²⁺ ions	–	–	–	–	1	1	1	1
MPG molecules	–	–	–	–	–	–	–	–
DMG molecules	–	–	–	–	1	1	1	1
[PO₄]³⁻/[PO₃]⁻ ions	–	–	1	1	1	1	1	1
[VO₃(O^δ2Asp6)]³⁻/[VO₃]⁻ ions	–	–	–	–	–	–	–	–
non-hydrogen protein atoms	1940	1944	1960	1915	1939	1953	1956	1956
disordered residues	0	2	2	1	1	1	4	2
solvent molecules	146		439		323		391	
mean B value (Å²)[e]								
protein main chain	30.5	32.1	36.7	37.0	46.7	39.8	25.6	40.9
protein side chain	31.6	34.0	43.2	44.8	56.8	51.9	33.7	48.3
ions and ligands	48.3	47.1	50.9	50.4	–	49.0	30.6	33.0
solvent	28.3		43.3		55.5		43.5	
model rmsd from ideality								
bond lengths (Å)	0.008		0.007		0.008		0.007	
bond angles (deg)	1.120		0.965		1.114		1.032	
chiral centers (Å³)	0.066		0.060		0.072		0.065	
planar groups (Å)	0.005		0.005		0.005		0.005	
model validation[f]								
% Ramachandran outliers	0.0		0.2		0.4		0	
% Ramachandran favored	95.6		98.0		95.4		97.6	
% rotamer outliers	5.8		0.3		1.8		0.3	
C^β outliers	0		0		0		0	
Clash score	19.0		15.4		16.5		13.0	

	[Gd³⁺-MPG] crystal 1	[Gd³⁺-MPG] crystal 2	[MPG-VO²⁺]	[Mg²⁺-MG-HPO₄²⁻] (control)	[Mg²⁺-MG-VO₄³⁻] (control)
resolution limits (Å)	24.96–1.90 (1.97–1.90)	28.79–1.92 (1.99–1.92)	46.28–2.10 (2.17–2.10)	38.53–1.74 (1.80–1.74)	28.77–1.81 (1.87–1.81)
R factor[a]	0.162 (0.233)	0.161 (0.220)	0.156 (0.198)	0.172 (0.225)	0.168 (0.280)

Table 3. continued

	[Gd³⁺-MPG] crystal 1		[Gd³⁺-MPG] crystal 2		[MPG-VO²⁺]		[Mg²⁺·MG-HPO₄²⁻] (control)		[Mg²⁺·MG-VO₄³⁻] (control)	
no. of reflections[b]	37877 (3648)		35831 (3271)		28710 (2700)		48242 (4569)		43834 (4274)	
R_{free}[c]	0.203 (0.276)		0.214 (0.290)		0.211 (0.276)		0.219 (0.306)		0.213 (0.334)	
no. of reflections[b]	2020 (192)		1911 (167)		1521 (160)		2592 (253)		2340 (232)	
overall coordinate error estimate (Å)[d]	0.27		0.25		0.29		0.24		0.25	
	chain A	chain B	chain A	chain B	chain A	chain B	chain A	chain B	chain A	chain B
Model Completeness and Composition										
regions omitted	253–259	253–259	251–259	252–259	253–259	253–259	251–259	255–259	252–259	252–259
Gd³⁺ ions	–	–	–	–	–	–	–	–	–	–
Cl⁻ ions	1	–	–	1	1	1	1	1	–	–
Mg²⁺ ions	1	1	1	1	1	1	1	1	1	1
MPG molecules	–	–	–	–	–	–	–	–	–	–
DMG molecules	–	1	1	1	–	–	1	1	1	1
[PO₄]³⁻/[PO₃]⁻ ions	–	1	–	–	–	–	1	1	–	–
[VO₃(Oδ2Asp6)]³⁻/[VO₃]⁻ ions	–	–	–	–	–	1	–	–	1	1
non-hydrogen protein atoms	1946	1946	1940	1946	1957	1954	1975	2020	1946	1959
disordered residues	3	1	–	–	1	1	7	7	1	4
solvent molecules	388		435		427		445		391	
mean B value (Å²)[e]										
protein main chain	40.0	40.4	37.5	30.9	31.8	40.3	40.0	32.6	35.3	28.6
protein side chain	48.5	49.1	45.0	38.6	41.0	49.0	47.2	41.5	41.9	35.9
ions and ligands	54.6	38.1	30.3	23.8	45.2	34.8	33.8	30.3	27.3	21.3
solvent	49.7		46.8		46.7		47.9		44.2	
model rmsd from ideality										
bond lengths (Å)	0.007		0.007		0.007		0.006		0.007	
bond angles (deg)	1.009		0.986		0.979		1.020		1.072	
chiral centers (Å³)	0.066		0.063		0.062		0.066		0.070	
planar groups (Å)	0.005		0.005		0.005		0.005		0.005	
model validation[f]										
% Ramachandran outliers	0		1.0		0.4		0.2		0	
% Ramachandran favored	97.6		96.2		97.4		96.7		97.2	
% rotamer outliers	0.8		2.4		0.5		1.5		0.3	
Cβ outliers	0		0		0		0		0	
Clash score	11.23		13.50		9.4		12.3		11.6	

[a]R factor = $\sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively. [b]No $\sigma(F_o)$ cutoff. [c]Cross-validation R factor computed from a randomly chosen subset of 5% of the total number of reflections, which were not used during the refinement. [d]Maximum-likelihood estimate. [e]Calculated from isotropic or equivalent isotropic B values. [f]Calculated with MolProbity.[49]
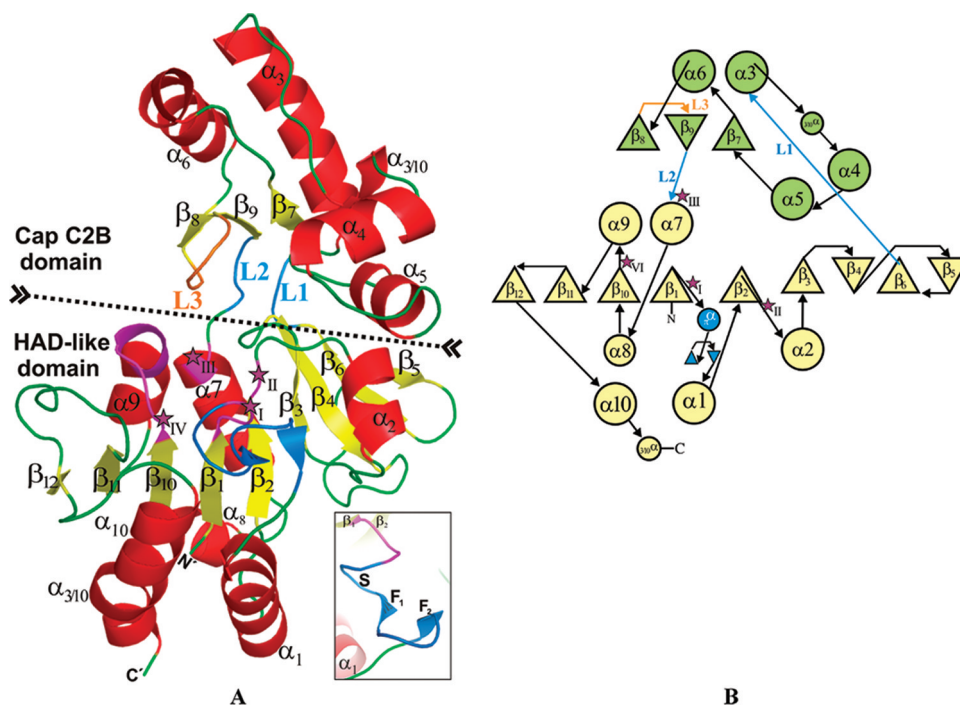
Figure 2. Three-dimensional structure of MpgP from *T. thermophilus* HB27. (A) Cartoon representation of the MpgP monomer. The dashed line indicates the interface between the HAD-like and C2B cap domains. Flexible loops L1 and L2 (cyan) connect both domains, between strand $\beta6$ and helix $\alpha7$. At the C2B cap domain, loop L3 is located at the three-stranded antiparallel β-sheet ($\beta8-\beta9\beta7$) of the Cof core unit and is involved in recognition and binding of the mannosyl moiety. The core region of the HAD-like domain consists of an α/β-sandwich with a parallel central β-sheet ($\beta11\beta10\beta1\beta2\beta3$) and two consecutive structural motifs (cyan): squiggle (S) and flap (F), immediately following strand $\beta1$. Conserved catalytic motifs I–IV are denoted with purple stars. The inset shows in detail the structural elements of the S and F motifs: a π-helical turn and a β-turn. This panel was prepared with PyMOL (http://www.pymol.org). (B) Topology diagram of the MpgP monomer. The HAD-like domain is colored yellow, and the C2B cap domain is colored green. Secondary structure catalytic elements correspond to those shown in panel A; α-helices are represented by circles and β-strands by triangles. The secondary structure elements were assigned on the basis of DSSP algorithm[65] implemented in PROCHECK[31,66] and the PROMOTIF version 3.0 output of ref 67. The topology diagram was based on TOPS.[68]

subfamily (PGPase) and the sucrose-phosphatase (SPP) subfamily as well as several predicted phosphoric monoester sugar hydrolases of unknown function. The Cof hydrolases and Cof-related members of the family are known to act with a wide spectrum of phospho-sugar compounds.

Each *T. thermophilus* MpgP monomer is composed of two domains (Figure 2): the HAD-like domain (Met1–Ser84 and Asp176–Pro259) and a cap module of the C2B subclass (Tyr89–Ala172), which is inserted between strand $\beta6$ and helix $\alpha7$ of the HAD-like domain via two flexible loops termed L1 (Leu85–Trp87) and L2 (Lys173–Ala175). The architecture of the HAD-like domain comprises the basal HAD core unit[10] (residues Met1–Gly61 and Lys177–Val219), which is composed of a three-layer α/β-sandwich with a central β-sheet formed by five parallel β-strands ($\beta11\beta10\beta1\beta2\beta3$) and is extended through the insertion of additional strands, arranged as follows: $\beta12\beta11\beta10\beta1\beta2\beta3-\beta4\beta6-\beta5$, with strands $\beta4$ and $\beta5$ antiparallel to the rest. Strand $\beta1$ is immediately followed by two consecutive highly flexible structural elements: a rare π-helical turn (residues Leu7–Leu12) and a type IV β-turn (residues Asp13–Glu17). These correspond to the squiggle and flap signature motifs, respectively, present in the HAD-SF members.[10] The C2B cap module comprises a typical $\alpha+\beta$ Cof core region (residues Tyr89–Ala103 and Glu139–Ala172) with a three-stranded antiparallel β-sheet (strands $\beta8$, $-\beta9$, and $\beta7$) surrounded by two helices ($\alpha3$ and $\alpha6$), extended by the insertion of two additional oppositely oriented helices ($\alpha4$ and $\alpha5$), which is characteristic of the MPGP. A cavity enclosing the

active site is located between the C2B cap and the HAD-like domain.

The stability of the 2-fold NCS homodimer was assessed using the Protein Interfaces, Surfaces and Assemblies (PISA) service,[53] as the measured gain in solvation energy of the buried residues upon interface formation (ΔG^i), and its biological relevance was estimated by the complexation significance score (CSS). For the apo-MpgP structure, the buried area was 7% (1540 Å^2) of the total surface area (22920 Å^2) of the homodimer, and these values were similar to those obtained for the other structures. This corresponded to estimated ΔG^i values ranging between −6 and −10 kcal/mol and to a rigid-body entropy change at dissociation ($T\Delta S^{diss}$) of 13 kcal/mol. The interface has a hydrophilic character whereby mean numbers of 16 hydrogen bonds and 18 salt bridges are formed upon dimer assembly, in addition to a disulfide bridge via two cysteine residues related by 2-fold NCS, Cys144. In all crystal structures, the estimated free energy of dissociation (ΔG^{diss}) varied between 8 and 10 kcal/mol, with the exception of those of the [Mg^{2+}-MPG] crystal 1 and crystal 2 structures, which presented much lower ΔG^{diss} values of −0.7 and 4.1 kcal/mol, respectively. A CSS value of 1.0 for the structures with higher ΔG^{diss} suggests that the dimeric assembly is stable, probably corresponding to the stable dimer form in solution previously reported.[25] In contrast, most of the proteins structurally homologous to *T. thermophilus* MpgP are predicted to be monomeric in solution. Although it is located in a structurally conserved region, at the beginning of a 3$_{10}$-helical turn that

immediately precedes helix α6 in the basal Cof unit, Cys144 is only conserved among the *T. thermophilus* species. The dimer stability is compromised in the *T. thermophilus* MpgP crystal structures where the Cys144 residues are fully reduced, as observed by the 10-fold decrease in the CSS (CSS = 0.1) and the change in ΔG^{diss} to approximately −0.7 kcal/mol in the [Mg^{2+}-MPG] crystal 1 structure. However, this reduction of the cysteine residues with concomitant dissociation of the disulfide bridge may be an artifact caused by radiation damage during synchrotron data collection.[54]

Structural Differences between Apo-MpgP and Holo-MpgP. The superposition between the HAD-like and C2B cap domains of the two monomers in the monoclinic homodimer was conducted with CCP4 LSQKAB.[31,55] The apo-MpgP and control [Mg^{2+}-MG-HPO$_4^{2-}$] structures were selected as representatives and are hereafter termed apo-MpgP and holo-MpgP, respectively. A good C$^{\alpha}$ fit was observed, as shown by the average root-mean-square (rms) coordinate deviations for apo-MpgP (0.80 Å for the HAD-like domain and 0.17 Å for the C2B cap domain) and holo-MpgP (0.73 Å for the HAD-like domain and 0.31 Å for the C2B cap domain). The largest deviations were seen for the solvent-accessible regions and especially for the flexible loops that surround the catalytic pocket. In the HAD-like domain, these comprise the squiggle and flap motifs and a 3$_{10}$-helical turn (Asp203−Asp205) preceding helix α9, while the C2B cap domain includes the β-turn that connects strands β8 and β9 (Gly165−Phe168), herein termed loop 3. Some structural deviations were also seen at the HAD-like domain boundaries involving the two helices (α1 and α10) in the top layer of the $\alpha/\beta/\alpha$-sandwich, the partially disordered strand β5 and its flanking residues (Gly74−Gly79), and the long C-terminal loop (Gly220−Pro234) that brings strand β12 and helix α10 toward the core of the central β-sheet. The higher flexibility of these regions is reflected in their B factors being higher than those of the core domains (Figure S1 of the Supporting Information).

When superposing the corresponding domains of the apo-MpgP and holo-MpgP structures, we also observed small deviations in their core regions. The apo- and holoenzyme structures clearly describe two distinct conformational states, whereby a rotation of the C2B cap domain leads from an open to a closed form of MpgP. The rotation of the C2B cap domain in apo-MpgP (chain A) to its position observed in holo-MpgP (chain B) was analyzed with DynDom in CCP4[31,56] using default parameters. This motion was rationalized as a hinge bending of ∼11° around an axis that intersects the plane of the core β-sheet between strands β3 and β4 and is oriented toward the two flexible loops, L1 and L2 (Figure 3A). The bending residues include those in the partially disordered strand β5 and the subsequent β-turn (Arg75−Arg81), and the mechanical hinge residues (Leu85, Ala86, Gly174, and Ala175), which are parts of loops L1 and L2, respectively. However, a similar analysis of the C2B cap domain rotation in apo-MpgP (chain A) to yield holo-MpgP (chain A) reveals a more complex and concerted motion involving at least three hinge axes, one of which was interpreted as a hinge motion of ∼7° around an axis equivalent to that described for holo-MpgP chain B (Figure 3B) and corresponds to a quasi-closed state of holo-MpgP chain A. These results suggest an internal flexibility of the C2B cap domain in MpgP. The crystal structures obtained from the remaining data sets fall within these three major conformational states. Chain B is predominantly found with its bound ligand and in a closed conformation, whereas chain A remains in its
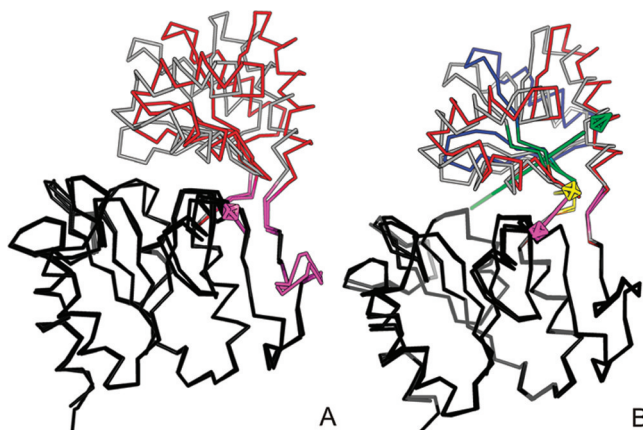
Figure 3. Hinge rotations of the C2B cap domain from the open to the closed conformations of MpgP. The protein chains, represented by their C$^{\alpha}$ trace, were superposed at their HAD-like domains and illustrate the hinge rotation of the Cap domain from the open to the closed conformation. The HAD-like domains are colored black, and the C2B cap domain of holo-MpgP used as reference for the closed conformation is colored gray. (A) Superposition of open apo-MpgP (chain A) onto closed holo-MpgP (chain B). The entire C2B cap domain (red) rotates by ∼11° about the hinge axis (magenta arrow). (B) Hinge rotations of open apo-MpgP (chain A) onto the quasi-closed conformation of holo-MpgP (chain A). Two flexible protein segments are distinguished at the C2B cap: the helical insertion motif (helices α4 and α5, residues Ala86−Glu139, red) and the Cof core region (residues Thr140−Ala172, blue). The colored arrows indicate the hinge rotation axes for the helical insertion motif (magenta, 7°; yellow, 9°) and for the Cof core region (green, 3°). The location of the bending and mechanical hinge residues is color-coded according to the color of their respective rotation axis. This figure was prepared with PyMOL (http://www.pymol.org).

open state for the shorter soaks or in a quasi-closed state for longer soaks.

These differences are most likely due to structural constraints arising from the crystallographic packing in the monoclinic crystal structures. To investigate this, solvent-accessible surface (SAS) calculations were conducted with AREAIMOL[31] for apo-MpgP to assess differences due to crystal contacts. An asymmetry in the distribution of accessible surface variations between monomers was observed, particularly in the cap domain. While no significant interactions are observed for chain B, the cap domain of chain A interacts with the C-terminal region of the HAD-like domain from two symmetry-equivalent molecules. This interaction includes strand β5 and its flanking residues, which contribute to the hinge bending motion, and probably restricts this motion, thus favoring a predominantly open conformation for holo-MpgP chain A in all the analyzed crystal structures. The analysis of these interfaces with PISA[53] showed the nature of the interactions to be predominantly electrostatic. In contrast, a single interface is observed in the orthorhombic crystal structure, between two monomers (chain A) related by a unit cell translation along the crystallographic *a*-axis, which is formed by residues from the Cof core region in one monomer and residues flanking strand β5 in the second monomer. In this structure, a nearly perfect superposition is found between one monomer and its 2-fold NCS counterpart, and both holo forms were found in their closed conformation, as shown by their hinge motion relative to apo-MpgP (11° for chain A and 12° for chain B).

■ DISCUSSION

Structural Comparison of *T. thermophilus* MpgP with Other Members of the MPGP. The different conformations found in the *T. thermophilus* MpgP crystal structures represented by apo-MpgP chain A (open) and holo-MpgP chains A (quasi-closed) and B (closed) were compared by secondary structure matching (SSM),[51] with their corresponding states in the orthologous MpgP crystal structures from *P. horikoshii* (PDB entries 2zos and 1wzc) and the MpgP-related YedP from *E. coli* (PDB entry 1xvi) (Figure S2 of the Supporting Information).

Two distinct conformations (open for chain A and closed for chain B) were seen in the crystal structure of the *P. horikoshii* selenomethionine derivative (PDB entry 2zos), which corresponded to the apo form.[7] The closed conformation in the apo form differed from that observed the in native holo-MpgP (PDB entry 1wzc), in the squiggle and flap structural motifs, and in loop 3. These differences most likely arose from the cocrystallization of the native protein with magnesium and phosphate that resulted in a different crystal packing.[7]

The superposition of the open state of *T. thermophilus* MpgP with its orthologues gave overall mean rms C^α deviations of 2.20 Å for 1xvi and 1.96 Å for 2zos (chain B), whereas for the closed state, the overall mean rms C^α deviations were 1.90 Å for 1wzc and 1.95 Å for 2zos (chain A). No structural homologue was found in the quasi-closed conformation. The superposition of the individual HAD core and C2B cap domains gave slightly better results. For the open state, the superposition of the HAD core domains of 1xvi and 2zos (chain B) with *T. thermophilus* MpgP gave mean rms C^α deviations of 1.95 and 1.66 Å, respectively, while their C2B cap domains superposed with mean rms C^α deviations of 1.87 and 1.53 Å, respectively; for the closed state, mean rms C^α deviations of 1.30 and 1.50 Å were obtained for the HAD core domain superposition of 1wzc and 2zos (chain A), respectively, with *T. thermophilus* MpgP, whereas the superposition of the C2B cap domains gave mean rms C^α deviations of 1.85 Å for 1wzc and 1.35 Å for 2zos (chain A). These relatively high values reflect the fact that although the HAD core and C2B cap domains share common folds in these structures, there are significant local differences arising from sequence insertions and deletions such as the relative orientation of α-helices and β-strands, as well as the position of loops such as the squiggle and flap flexible motifs of the HAD core region.

In the open state of *P. horikoshii* apo-MpgP, loop 3 is positioned in the plane of strands β8 and β9, while in *T. thermophilus* apo-MpgP and YedP, it follows the hinge motion of the Cap domain and is bent toward the catalytic pocket. This may result from the phosphate and sulfate ions bound to loop 3 in MpgP and YedP, respectively, and provide an explanation for the different hinge rotations observed for the *T. thermophilus* (11°) and *P. horikoshii* (19°)[7] enzymes. In the closed state, the hinge bending of the Cap domain moves loop 3 farther down into a conformation that is equivalent in both *P. horikoshii* and *T. thermophilus* MpgP.[7]

Catalytic Pocket and Its Comparison with the Other MPGP Enzymes. The catalytic pocket is located in the C2B and HAD-like interdomain region forming a binding cleft surrounded with ionizable and polar residues that are involved in substrate recognition and binding (Figure S3 of the Supporting Information). The four sequence signature motifs present in nearly all the aspartate nucleophile HAD-SF members are placed at stereospecific regions of the HAD core domain and are known to form a trigonal-bipyramidal electrostatic mold that efficiently stabilizes negatively charged pentavalent transition-state species[10,13] (Figure 2 and Figure S4 of the Supporting Information). Motif I is represented by Asp6-X-Asp8 in *T. thermophilus* MpgP, Asp8-X-Asp10 in *P. horikoshii* MpgP, and Asp13-X-Asp15 in *E. coli* YedP and is located at the C-terminal end of strand β1. Motif II consists of Thr39 (Ser40 in *P. horikoshii* MpgP and Ser47 in *E. coli* YedP) and is placed at the C-terminal end of strand β2. Motif III is placed at the N-terminus of helix α7 and contains Lys177 (Lys180 in *P. horikoshii* MpgP and Lys188 in *E. coli* YedP) oriented toward motifs I and II. Motif IV consists of the acidic signature Gly201-Asp202-Ser203-X-Asn205-Asp206 (Gly203-Asp204-Ser205-X-Asn207-Asp208 in *P. horikoshii* MpgP and Gly213-Asp214-Gly215-X-Asn217-Asp218 in *E. coli* YedP) that flanks the 3_{10}-helical turn between strand β10 and helix α9.

The recognition and binding of the MG group are accomplished by the interfacing residues arranged at the top and back of the catalytic pocket (Figure 4A). A diagram of the MG molecule along with the atom numbering scheme used in this work is displayed in Figure 4B. The unbiased sigmaA-weighted $|F_o| - |F_c|$ electron density for the ligands and water molecules in the $[\text{Mg}^{2+}\text{-MG-HPO}_4{}^{2-}]$ control structure is displayed in Figure 4C, and the corresponding ligand electron density for the other complexes discussed in this work is included in Figures S5 and S6 of the Supporting Information. At the α-face of the mannosyl moiety, a hydrogen bond network is established between HAD-like core residues Lys41 (Lys42 in *P. horikoshii* MpgP and Lys49 in *E. coli* YedP) and Glu45 (Glu46 in *P. horikoshii* MpgP and Glu53 in *E. coli* YedP) with C3(OH) and C4(OH), the hydroxyl groups of C3 and C4 in the sugar ring. At the β-face, atoms from residues in loop 3 (Arg167 N in MpgP, Arg170 N in *P. horikoshii* MpgP, and Arg177 N in *E. coli* YedP) and strand β9 (His170 $N^{\varepsilon 2}$ in MpgP, Thr173 O^γ in *P. horikoshii* MpgP, and His180 $N^{\varepsilon 2}$ in *E. coli* YedP) stabilize the carboxylate group of the glyceryl moiety. These residues line the region of the inner pocket surface that molds an electrophilic environment for mannosyl sugar recognition and binding. In particular, the side chain of Arg167 plays a key role in fixing the mannosyl moiety by making a bidentate hydrogen bond that extends from one of its terminal NH_2 groups to the C6(OH) group of the mannosyl moiety, and to the main chain carbonyl oxygen of Thr122 (Thr127 in *P. horikoshii* MpgP and Thr130 in *E. coli* YedP). This second interaction makes an important contribution to the closing of loop 3. The carbonyl oxygen of Gly165 (Gly168 in *P. horikoshii* MpgP and Gly175 in *E. coli* YedP) also accepts hydrogen bonds from the hydrolyzed ester O13, hereafter termed O13(H), and a water molecule coordinating the cocatalytic metal center. The catalytic roles of Gly165 and Arg167 will be further discussed in the following sections. Tyr110 (Tyr115 in *P. horikoshii* MpgP and Phe118 in *E. coli* YedP) is located at the interface of the two C2B cap subdomains. Tyr110 O^γ, together with Thr140 O^γ (Thr145 O^γ in *P. horikoshii* MpgP and Thr148 O^γ in *E. coli* YedP), stabilizes the hydroxyl group from the C2 epimer, C2(OH), and both are likely to assist in the stereochemical discrimination of the mannosyl group from other sugar compounds. The C3(OH) hydroxyl of the mannosyl moiety is also anchored by Arg135 $N^{\zeta 1}$ (Arg140 in *P. horikoshii* and His143 in *E. coli* YedP). The shorter side chain of His143 in YedP strengthens the notion
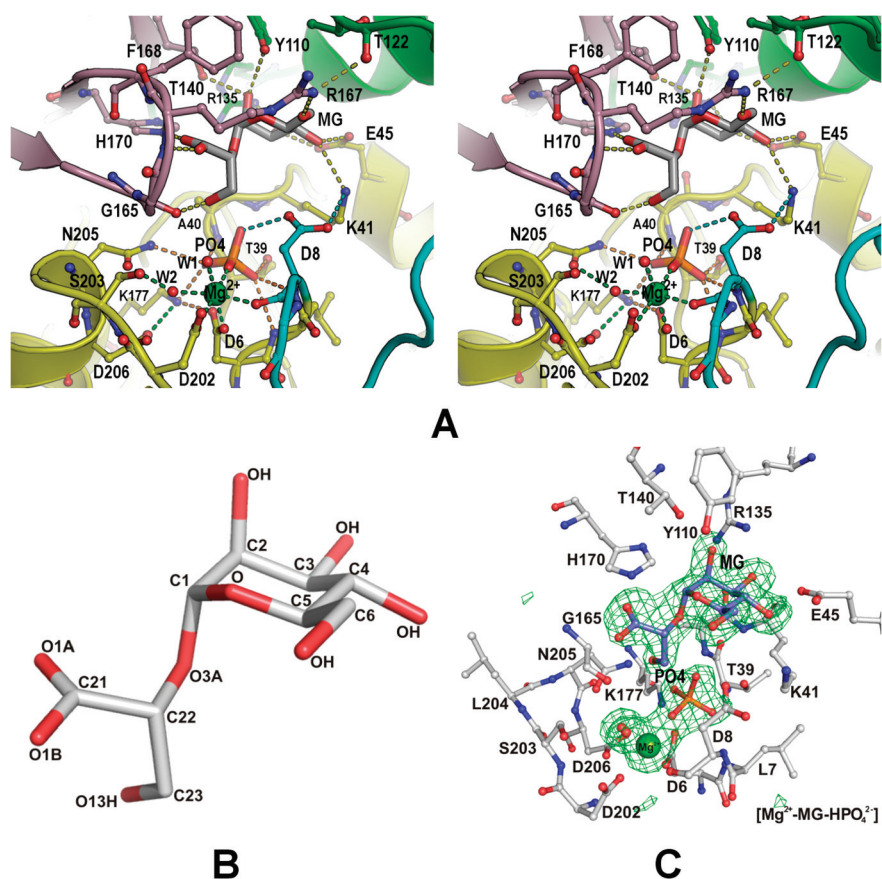
Figure 4. MpgP catalytic pocket. (A) Stereoview of the MpgP active site. The ternary complex bound in the closed conformation of the $[Mg^{2+}\text{-}MG\text{-}HPO_4^{2-}]$ control structure is represented. Residues involved in the recognition and binding of either the MPG substrate or the MG and HPO_4^{2-} products are represented as balls and sticks; the HPO_4^{2-} and MG ligands are displayed as sticks, and the water molecules and the Mg^{2+} ion are shown as spheres. Dashed lines indicate hydrogen bonds and are color-coded according to their binding pocket regions: green for the cocatalytic metal site, orange for the phosphate binding site, and yellow for the MG binding site. Atom colors are as follows: green for magnesium, orange for phosphorus, gray for MG carbon atoms, blue for nitrogen, and red for oxygen. The carbon atoms of MpgP are colored according to the structural regions: yellow for the HAD-like region, cyan for the squiggle and flap motifs, green for the helical insertion of α4 and α5 of the C2B cap module, and light pink for the Cof core motif. (B) Representation of the MG ligand. MG is oriented like it is in panel A. The carbon and oxygen atoms are labeled according to the nomenclature used in the text. (C) View of the catalytic pocket in chain B of the $[Mg^{2+}\text{-}MG\text{-}HPO_4^{2-}]$ control crystal structure, showing the unbiased sigmaA-weighted $|F_o| - |F_c|$ electron density map covering the ligands. The map was obtained with PHENIX by simulated annealing after omission of the ligands and the water molecules coordinated to the cocatalytic Mg^{2+} metal ion. The map is represented as a green mesh contoured at the 3σ level. Protein residues and the MG ligand are shown in ball-and-stick representation, and the Mg^{2+} metal ion and the water molecules are drawn as spheres. Atom colors are as follows: green for magnesium, orange for phosphorus, blue for nitrogen, red for oxygen, white for protein carbon atoms, and slate blue for MG carbon atoms. This figure was prepared with PyMOL (http://www.pymol.org).

that it is a MpgP-related enzyme, likely to use as a substrate a sugar compound modified at C3.

Interestingly, the ability of *T. thermophilus* and *P. horikoshii* MpgPs to use glucosyl-3-phosphoglycerate (GPG) as a substrate, in addition to MPG,[9] can be explained by the lack of steric hindrance toward the C2(OH) group of the glucopyranosyl ring. The unrooted phylogenetic tree based on the amino acid sequence alignment of functionally characterized MpgPs and GpgPs and homologues predicts three ill-defined clusters: one containing the MpgPs of *T. thermophilus* and *P. horikoshii* along with several homologues from (hyper)thermophilic organisms, a second comprising known GpgPs and homologues, and a third containing both MpgPs and GpgPs (Figure S7 of the Supporting Information). This convoluted distribution suggests the existence of a common ancestor with the ability to recognize both MPG and GPG as substrates. Interestingly, the topology of this phylogenetic tree contrasts clearly with that of MPG synthase/ GPG synthase in which the clusters for MpgSs and GpgSs are

far apart, in line with the strict substrate specificity of the enzymes.[9]

Coordination of the Cocatalytic Metal Ion Binding Site. In all crystal structures except $[Gd^{3+}]$, a coordinated Mg^{2+} metal ion was refined at the cocatalytic metal binding site. This happened even with $[Gd^{3+}\text{-}MPG]$ crystal 1 and 2 structures, soaked with $GdCl_3$ instead of $MgCl_2$, which was prepared with the aim of trapping a reaction coordinate mimic (Table S2 of the Supporting Information). However, no anomalous signal was detected for Gd^{3+}, and the electron density peak present at the cocatalytic metal binding site was also interpreted as a Mg^{2+} metal ion, probably present in trace amounts in the soaking solution and with a high affinity for the metal site. This may also explain the presence of a Mg^{2+} ion in the $[MPG\text{-}VO^{2+}]$ crystal structure.

The binding of the cocatalytic Mg^{2+} ion should follow a structural rearrangement of the squiggle and flap motifs, as suggested by their different conformations in chains A and B of apo-MpgP, and in chain A of the metal-bound apo form in the
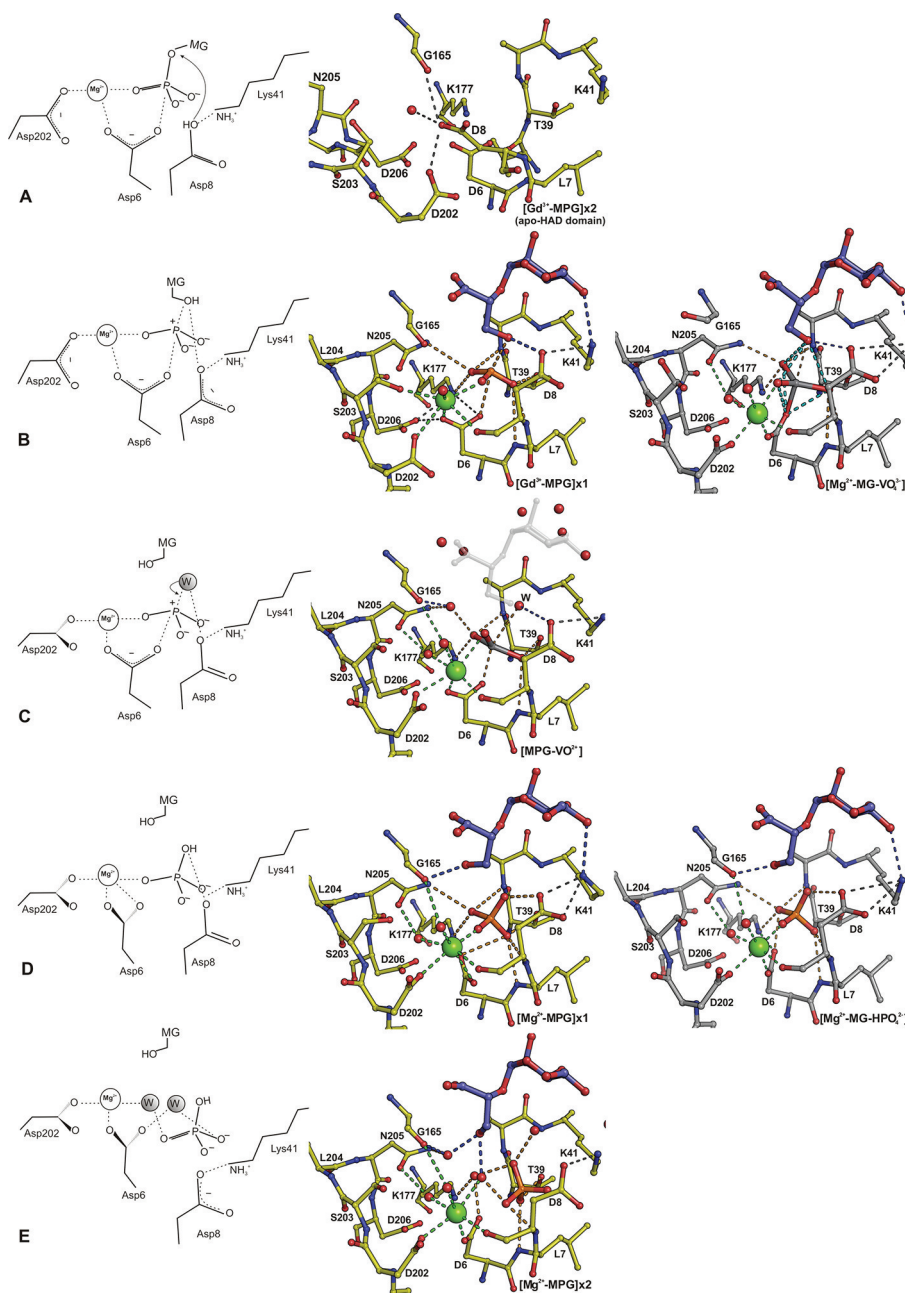
Figure 5. Catalytic events in the proposed phosphoryl transfer mechanism for MpgP. Except for the snapshot in panel A, for each event a diagram is shown to the left, with its corresponding crystallographic snapshot (B, D, and E) or mimicking PO_3^- intermediate structure (E) in the middle and the respective control crystal structure (B and D) at the right. In each diagram, only key residues involved in the synchronous migration of the cocatalytic Mg^{2+} ion with the PO_3^--transferring species and in the assisted acid catalysis are depicted. For the sake of clarity, minimal electrostatic interactions between the catalytic residues and the migrating species are shown as dashed lines, while arrows indicate the assisted proton transfer mechanism and nucleophilic attack by the catalytic water molecule. In the middle and right panels, the electrostatic interactions between the catalytic HAD core residues and the migrating Mg^{2+} and PO_3^- species are displayed as dashed lines and are color-coded according to their respective binding sites (green for the cocatalytic Mg^{2+} binding site and bright orange for the phosphoryl binding site), while those extending toward the MG binding site are colored blue. Electrostatic interactions between protein residues as well as hydrogen bonds between protein residues and water molecules are shown as gray dashed lines. (A) Active enzyme, following the structural rearrangement of the HAD core residues in concert with the binding of the cocatalytic Mg^{2+} ion and the phosphoryl group of the MPG substrate. The Asp8−Lys41 salt bridge positions Asp8 in-line with the ester oxygen for assisted acid catalysis. In the middle panel, the [Gd^{3+}-MPG] crystal 2 structure illustrates one of the possible conformations acquired by the HAD core region in its apo form, and the ability of Asp8 to become protonated. (B) After cleavage of the phosphoester bond, the short-lived PO_3^- intermediate is stabilized by the trigonal-bipyramidal electrostatic mold, as illustrated by the [Gd^{3+}-MPG] crystal 1 structure in the middle panel. In the right panel, the aspartyl-vanadate adduct of the [Mg^{2+}-MG-VO_4^{3-}] crystal structure is stabilized by the trigonal-bipyramidal electrostatic mold in an ideal geometry. (C) Nucleophilic attack at the PO_3^- center by a water molecule, represented by the mimicking VO_3^- complex and its axial water molecule, W. (B−D) Side chain rotation of residues Asp6 and Asp202, in concert with the migration of the Mg^{2+} and PO_3^- species in panel B or mimicking [Mg^{2+}-VO_3^-] complex in panel C, toward their final positions (D). (E) Release of HPO_4^{2-} followed by the collapse of the trigonal-bipyramidal electrostatic mold. Atom colors are as follows: green for magnesium, orange for phosphorus, blue for nitrogen, red for oxygen, yellow for protein carbon atoms, and slate blue for MG carbon atoms. This figure was prepared with PyMOL (http://www.pymol.org).

[Gd^{3+}-MPG] crystal 2 structure (Figure S8 of the Supporting Information). Thus, enzyme activation implies the binding of the cocatalytic Mg^{2+} ion for the subsequent phosphoryl transfer process.

In the active enzyme, the coordination of the cocatalytic metal ion changes from a loose distorted octahedral geometry, as suggested by the coordination bond length values of approximately 2.2−2.8 Å [represented by chain B in the [Gd^{3+}-MPG] crystal structure (Figure 5B, middle panel)] and here termed the labile metal ion, to a tighter and ideal octahedral geometry, with coordination bond lengths in the range of 2.0− 2.2 Å, within the range of typical values for Mg^{2+} coordination (http://mespeus.bch.ed.ac.uk/MESPEUS), exemplified by the [Mg^{2+}-MG-HPO$_4^{2-}$] and [Mg^{2+}-MG-VO$_4^{3-}$] control structures (Figure 5 B,D, right panel) and here termed the fixed metal ion. There is a positional shift of ∼0.7 Å between the two types of metal centers, accompanied by changes in their first-shell ligands. The octahedral coordinating environment comprises conserved acidic residues from motif I (Asp6 O$^{\delta2}$ and Asp8 O) and motif IV (Asp202 O$^{\delta1}$), and two water molecules, W1 and W2. W1 replaces Ser203 O$^{\gamma}$ in the labile coordinating state, and both are partially involved in holding the motif IV residues assembled at the cocatalytic metal site; W2 interacts with G165 O in loop 3 in the ideal octahedral geometry. In the Michaelis ternary complex, the octahedral geometry is completed by coordination with an oxygen atom from the transferring phosphoryl group of MPG (Figure 5B, middle panel) or by its hydrolyzed orthophosphate product (Figure 5D). A higher coordination number ($N = 7$) in the [Mg^{2+}-MPG] crystal 1 structure may be transiently reached as a consequence of the side chain rotation of Asp6 (see Discussion).

Binding and Stabilization of the Phosphoryl Species. The binding modes of the orthophosphate observed in chain B of the [Mg^{2+}-MG-HPO$_4^{2-}$] control crystal structure and in chain B of the [Mg^{2+}-MGP] crystal 1 structure are similar. Three oxygen atoms are stabilized by the amide hydrogen atoms from the motif I peptide backbone (Leu7 N and Asp8 N) and the neighboring residue Ala40 (Ala40 N) in motif II, the cocatalytic Mg^{2+} metal ion, Lys177 N$^{\zeta}$ in motif III, and Asn205 N$^{\delta2}$ in motif IV (Figure 4A). The fourth oxygen of HPO$_4^{2-}$ accepts hydrogen bonds from the hydrolyzed ester oxygen O13(H) of the MG leaving group and Asp8 O$^{\delta2}$. These residues form an electrostatic mold with a trigonal-bipyramidal-like geometry that assists in the stabilization of a developing TS species.[13] This geometry is illustrated by the aspartyl-vanadate adduct of the [Mg^{2+}-MG-VO$_4^{3-}$] control crystal structure (Figure 5B, right panel). The vanadium atom is equatorially coordinated by three oxygen atoms (restrained bond lengths of 1.9 Å) and axially by Asp6 O$^{\delta1}$ and MG O13 with V−O distances of 2.0 and 2.7 Å, respectively, in agreement with previous results.[13] In this structure, the phosphoryl analogue orthovanadate (VO$_4^{3-}$) underwent an equatorial polarization with expansion of the axial V−O bond length, prior to its nucleophilic attack by Asp6 O$^{\delta2}$ with subsequent formation of an aspartyl-vanadate adduct. However, this may not truly reflect the activation chemistry for phosphoryl compounds.[57] Indeed, and contrary to the stronger additive character required for the formation of a phosphoryl-adduct intermediate as expected by the associative mechanism of phosphoryl transfer,[57] the electron density of the transferring PO$_3^{-}$ species in the active site of the [Gd^{3+}-MPG] crystal 1 structure suggests a mixture of nonhydrolyzed R-PO$_4^{2-}$ and hydrolyzed PO$_3^{-}$ species [hereafter designated as the PO$_3^{-}$-like intermediate (Figure S6 of the

Supporting Information)], reinforcing the notion of a concerted dissociative-like mechanism for this first step, with weaker bond order formation from the attacking nucleophile Asp6 (see Phosphoryl Transfer Mechanism in MpgP for details). In vivo, the short-lived PO$_3^{-}$ intermediate is very likely stabilized by the trigonal-bipyramidal electrostatic mold described above as suggested by the [Gd^{3+}-MPG] crystal 1 structure (see also Discussion for details; Figure 5B, middle panel). The phosphorus center should become positively charged and stabilized by the hydrolyzed ester oxygen of the MG leaving group, either as an alkoxide anion C23(O13^{-}) or as a hydroxyl group O13(H), and by Asp6 O$^{\delta1}$. These two oxygen atoms are axially located in relation to the plane defined by the three pseudoequatorial oxygen atoms of the PO$_3^{-}$-like intermediate. These two atom groups are hereafter termed bridging and nonbridging oxygen atoms, respectively. In contrast with the salt bridge between Asp10 and Arg170 in the holo-MpgP structure of *P. horikoshii*,[7] a salt bridge between corresponding residues Asp8 and Arg167 is not observed in the *T. thermophilus* MpgP structures. This salt bridge was suggested to prevent the reverse phosphoryl transfer reaction,[7] yet such role is doubtful because of the absence of a true Michaelis complex bound at the active site. Instead, in the *T. thermophilus* MpgP structures, Asp8 interacts with Lys41 and assumes a conformation similar to that observed for several members of the IIB and III HAD subfamilies[13] (Figure 4A). In the [Gd^{3+}-MPG] crystal 1 structure, Asp8 O$^{\delta2}$ is oriented in line with O13(H) of the MG leaving group and makes a bifurcated hydrogen bond with O13(H) (2.5 Å) and with the amine group of Lys41 (2.8 Å) (Figure 5B, middle panel). These interactions highlight the role of Asp8 as a general acid catalyst (see Phosphoryl Transfer Mechanism in MpgP for details). On the other hand, in the [Mg^{2+}-MPG] crystal 1 structure, when the bridging oxygen atoms move away from their axial positions, a salt bridge between Asp8 and Lys41 is likely to form (Figure 5D, middle panel).

In the [Mg^{2+}-MPG] crystal 2 structure, the HPO$_4^{2-}$ ion is found closer to the pocket exit, suggesting a HPO$_4^{2-}$ release concerted with the collapse of the trigonal-bipyramidal electrostatic mold, primarily caused by the absence of the axial stabilization from Asp8 O$^{\delta2}$ (Figure 5E, middle panel).

Structural Changes in the MG Binding Pocket during MPG Binding and Catalysis. The anchoring of the phosphoryl group of the MPG substrate to the cocatalytic Mg^{2+} ion favors the hinge bending of the C2B cap module toward the MG binding region of the catalytic pocket, shaping it for the recognition and binding of the sugar moiety and resulting in the closed conformation of the holoenzyme (Figure 6). This shape is also found for chain B in the crystal structure of the [MPG-VO^{2+}] metavanadate complex, in which neither bound MPG nor MG was observed; instead, water molecules are hydrogen-bonded to the same residues that interact with the hydroxyl groups of the sugar substrate (Figure 5C, middle panel). Locally, residues in loop 3 are also key mediators in the reaction cycle while binding to the sugar substrate. In particular, Arg167 should play a central role in the global bending of the C2B cap module by stabilizing the mannosyl moiety via the C6(OH) group as previously described.

In the [Mg^{2+}-MPG] crystal 1 structure, the hydroxymethyl group C23O13(H) rotates by 108° about its σ-bond to the preceding carbon atom, forming hydrogen bonds with Gly165 O (Figure 5D, middle panel). On the other hand, in the [Mg^{2+}-MG-VO$_4^{3-}$] crystal structure, the O13(H) group of MG is still hydrogen-bonded with Asp8 O$^{\delta2}$ and axially coordinates the vanadium atom (Figure 5B, right panel). Therefore, the
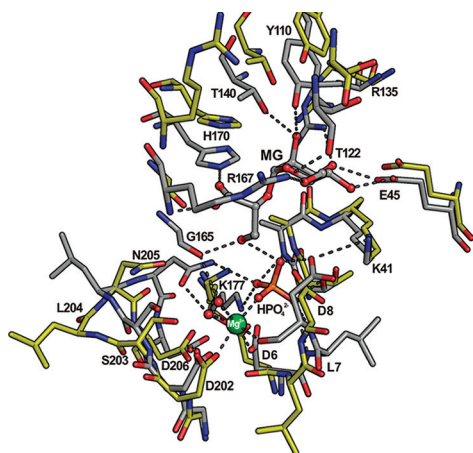
Figure 6. Enzyme activation and closure via hinge bending. Overlay of the active sites of apo-MpgP chain B and $[Mg^{2+}\text{-}MG\text{-}HPO_4^{2-}]$ control structure chain B, which correspond to the open and inactive state and the closed and active state of the enzyme, respectively. The structures are superposed at their HAD-like domains and illustrate the changes occurring in the HAD core region upon the binding of the cocatalytic metal ion Mg^{2+} and the phosphate moiety of MPG (represented by the ligand HPO_4^{2-}). Residues involved in the recognition and binding of either the MPG substrate or the MG and HPO_4^{2-} products are drawn as sticks, whereas the HPO_4^{2-} and MG ligands are shown in ball-and-stick representation. Water molecules and the Mg^{2+} ion are displayed as spheres. Hydrogen bonds established between ligands and holo-MpgP, and relevant for hinge bending, are represented as dashed lines. Atom colors are as follows: green for magnesium, orange for phosphorus, gray for MG carbon atoms, blue for nitrogen, and red for oxygen. The carbon atoms of MpgP are colored yellow for apo-MpgP and gray for holo-MpgP. This figure was prepared with PyMOL (http://www.pymol.org).

enzyme-inhibited state obtained by the formation of the aspartyl-vanadate adduct in the $[Mg^{2+}\text{-}MG\text{-}VO_4^{3-}]$ crystal structure can be said to mimic the acid-assisted catalysis that corresponds to the first step in the phosphoryl transfer catalytic mechanism. Gly165 O should play a primary role in the second step, by stabilizing the conformational change of the C23O13(H) group observed in the $[Mg^{2+}\text{-}MPG]$ crystal 1 structure and thus allowing the nucleophilic attack of the PO_3^- center by an activated water molecule. It should be noted that in the $[Mg^{2+}\text{-}MG\text{-}VO_4^{3-}]$ crystal structure the Gly165 carbonyl is flipped, pointing in the opposite direction from that found in all the other *T. thermophilus* MpgP crystal structures herein presented, but this is probably not relevant to the conclusions drawn from this structure regarding the phosphoryl transfer catalytic mechanism.

Phosphoryl Transfer Mechanism in MpgP. During the crystal soaking procedures, the crystallized MpgP displayed a catalytic rate slow enough to allow the trapping of hydrolyzed species at several stages of the phosphoryl transfer reaction. These were rationalized as crystallographic snapshots, reflecting a series of molecular and structural events that take place during the catalytic cycle. Additionally, the soaks conducted with the hydrolyzed products (MG and HPO_4^{2-}), and also in the presence of vanadium(V) oxoanion compounds (VO_3^- and VO_4^{3-}), were used as control crystal structures, allowing a detailed analysis of the binding residues and of the extended electrostatic network that is engaged in overcoming the energetic barriers that arise along the reaction coordinate. These structural results paralleled the crystallographic snapshots of MPG hydrolysis, thus supporting the insights into the

catalytic mechanism derived therefrom. The catalytic cycle of MPG hydrolysis is schematically represented in Figure 5, and the key events of the reaction coordinate are interpreted on the basis of the structural data derived from the crystallographic snapshots and their corresponding control crystal structures.

In the $[Gd^{3+}\text{-}MPG]$ crystal 1 structure, the trapped PO_3^--like species exhibits an apparently asymmetric axial coordination, i.e., 2.7 Å to hydrolyzed ester O13 and 2.9 Å to Asp6 $O^{\delta1}$ (Figure 5B, middle panel). This difference is on the same magnitude as the overall mean coordinate error estimate for this crystal structure (0.27 Å), and therefore, its significance may be questioned. However, it is likely that the residues in the active site are more rigid and thus have lower coordinate errors than those on the surface of the enzyme. This asymmetry suggests that the phosphoryl transfer reaction in MpgP should employ a concerted dissociative-like mechanism ($A_N D_N$), but with a weaker S_N2 character for the stabilization of the resonance species of the PO_3^- intermediate. In this view, the reactivity of the MG leaving group should overcome that of the attacking nucleophile Asp6, with a transition state in which the bond to the leaving group is more extensively broken than the bond to the nucleophile is formed.[17,21] Higher-resolution diffraction data would be needed to verify this hypothesis.

At this stage, the MG leaving group remains fixed in its binding site, whereas the PO_3^--like species is stabilized by the trigonal-bipyramidal electrostatic mold formed by motifs I–IV. This is in agreement with previous studies conducted with the predicted hexose phosphate phosphatase (HPP) from *Bacteroides thetaiotaomicrometer*,[13] in which the role of this mold was assessed in connection with the stabilization of the trigonal-bipyramidal TS species and the catalytic turnover. A dianionic monoester phosphate should be the reactive form of the MPG substrate, as suggested by the pH dependence of the MpgP catalytic activity, with its highest turnover rate at pH ~6.0.[17,58] This fact highlights the need for an external acid catalyst, required to enhance the rate of phosphoryl transfer,[13,59] contrary to the intramolecular proton shift postulated for monoanions of phosphate monoesters.[17,60] The $[Gd^{3+}\text{-}MPG]$ crystal 1 structure represents acid-assisted catalysis in the first step of the phosphoryl transfer mechanism (Figure 5B, middle panel). Asp8 is predicted to act as the proton donor, supported by the hydrogen bond between hydrolyzed ester O13 and Asp8 $O^{\delta2}$ (2.5 Å), together with the trapped PO_3^--like intermediate, and implying the coupled protonation of the hydroxymethyl anion C23O13$^-$ with the P–O bond fission. Asp8 can be transiently protonated prior to the binding of the cocatalytic metal ion (Figure 5A). This is illustrated by the $[Gd^{3+}\text{-}MPG]$ crystal 2 structure, in which chain A Asp8 is hydrogen-bonded to the carbonyl oxygen of Gly165 (Figure S8 of the Supporting Information). Indeed, this structure contains snapshots of the final stages of the catalytic cycle: in chain B, the HPO_4^{2-} product has been replaced by two water molecules, followed by a destabilization of the cocatalytic metal center, whereas in the quasi-closed conformation of chain A, the labile cocatalytic metal ion is released, allowing Asp8 to receive a proton from a water molecule of the solvent channel that fills the metal binding site. On the other hand, the energetic barrier for proton donation is expected to be significantly lowered, both by the relatively high pK_a predicted for the MG leaving group[17,59] and by the electrostatic effect between Asp8 and Lys41. In the latter, the observed salt bridge that has been proposed for several members of the IIB and III HAD subfamilies[61] will have the effect of decreasing the pK_a of Asp8 and increasing the pK_a

of Lys41, as would be generally the case for a coupled acid–base pair,[62] thus destabilizing the protonated state of Asp8 and favoring the donation of its proton to the ester oxygen of MPG, while this is being stabilized for the in line assisted acid catalysis (Figure 5B).

The interaction between Asp6 $O^{\delta 1}$ and the positive charge at the phosphorus atom in the PO_3^- intermediate (Figure 5B) facilitates P–O bond fission and contributes to the synchronous migration of the anchored PO_3^- species with the loosely bound cocatalytic Mg^{2+} ion, when comparing their positions with those in the $[Mg^{2+}\text{-}MPG]$ crystal 1 structure, which represents the final reaction products (Figure 5B,D). More significantly, this relatively small motion is accompanied by large changes in the orientation of the carboxyl groups of Asp6 and Asp202: both rotate by ∼90° with respect to their orientation in the $[Gd^{3+}\text{-}MPG]$ crystal 1 structure. In particular, the rotation of Asp6 seems to be directly related to the electrostatic interaction between Asp6 $O^{\delta 1,2}$ and the phosphorus atom in HPO_4^{2-}, whereas the rotation of Asp202 appears to be connected with the coordination of Mg^{2+} by Asp202 $O^{\delta 1}$ (Figure 5C). Similar orientations were observed for the carboxylate group of Asp6 in the $[Mg^{2+}\text{-}MG\text{-}VO_4^{3-}]$ and $[Mg^{2+}\text{-}MG\text{-}HPO_4^{2-}]$ control crystal structures, which replicate those of the $[Gd^{3+}\text{-}MPG]$ and $[Mg^{2+}\text{-}MPG]$ crystal 1 structures, respectively. In the $[Mg^{2+}\text{-}MG\text{-}VO_4^{3-}]$ control crystal structure, Asp6 $O^{\delta 1}$ is bound axially to the vanadium center (2.0 Å) (Figure 5B, right panel), whereas in the $[Mg^{2+}\text{-}MG\text{-}HPO_4^{2-}]$ control crystal structure, Asp6 $O^{\delta 2}$ interacts axially with a nonbridging oxygen atom of HPO_4^{2-} (2.8 Å) and has probably rotated because of electrostatic repulsion prior to product release (Figure 5D, right panel). Thus, the structural template provided by the active catalytic HAD core region has the ability to hold in proper position the two migrating electrophilic centers, while balancing the hydrogen bonding network and electrostatic potential in their vicinity as needed to overcome the energetic barriers of P–O bond fission and P–O bond formation. Indeed, the migration of the cocatalytic Mg^{2+} ion with the concomitant stereoinversion of the PO_3^- intermediate was previously proposed for crystals of fructose-1,6-bisphosphatase (FBPase), grown under an equilibrium mixture of substrates and products.[22] Also, crystallographic snapshots of *Lactococcus lactis* β-phosphoglucomutase (β-PGM) revealed a migration of the cocatalytic Mg^{2+} center by 0.7 Å synchronously with the enzyme activation by phosphorylation, and in concert with a rotation of the carboxylate groups from motifs I and IV.[63]

The second step of the phosphoryl transfer reaction comprises the nucleophilic water attack on the PO_3^- intermediate, leading to the final products represented by the ligands in the end reaction $[Mg^{2+}\text{-}MPG]$ crystal 1 structure (Figure 5C,D). The rotation of the hydroxymethyl group C23O13(H) away from Asp8 to form a hydrogen bond with Gly165 O leads to the exposure of the positive charge on the PO_3^- intermediate, which may then be readily attacked by a water molecule, followed by the release of a proton. In contrast with the associative phosphoryl transfer mechanism of the C2B phosphatase HPP[13] where a role as a general acid/base catalyst was proposed for the second Asp residue (Asp10), a similar role for Asp8 is questionable in MpgP. Its expected reduction in pK_a, which results from the salt bridge established with Lys41, favors a deprotonated state throughout the second step and conflicts with Asp8 playing a role as a general base. Instead, Asp8 should mainly play a structural role during this stage by

maintaining the trigonal-bipyramidal electrostatic mold that holds the PO_3^- intermediate. Structural evidence of the preservation of the Asp8–Lys41 salt bridge during the second step of the reaction is given by the PO_3^--mimicking intermediate $[MPG\text{-}VO^{2+}]$ crystal structure, in which an axial water molecule spanning from Asp8 to the VO_3^- center mirrors either the hydrolyzed ester O13(H) of MG or the attacking water molecule (Figure 5C), and by the apical interaction of Asp8 with the hydrolyzed HPO_4^{2-} product prior to its release (Figure 5D). This may help to explain the lower catalytic rate measured for the HPP D10A mutant in the presence of a low-energy leaving group.[13]

■ ASSOCIATED CONTENT

⑤ Supporting Information

Solutions used in the preparation of the $[Gd^{3+}]$ crystals by soaking (Table S1), summary of crystallization and crystal soaking procedures for all data sets with their respective bound ligands (Table S2), mean main-chain B factors for chains A and B of apo-MpgP and holo-MpgP (control $[Mg^{2+}\text{-}MG\text{-}HPO_4^{2-}]$) structures (Figure S1), structural comparison of *T. thermophilus* MpgP with its orthologous members of the MPGP (Figure S2), electrostatic potential at the molecular surfaces of apo-MpgP and holo-MpgP structures (Figure S3), structure-based sequence alignment of MpgP with its structural homologues (Figure S4), unbiased sigmaA-weighted $|F_o| - |F_c|$ electron density for the ligands and water molecules for the $[Gd^{3+}\text{-}MPG]$ crystal 2, $[Mg^{2+}\text{-}MPG]$ crystal 1 and crystal 2, and $[MPG\text{-}VO^{2+}]$ structures and the $[Mg^{2+}\text{-}MG\text{-}VO_4^{3-}]$ control structure (Figure S5), stereoview of the $[Gd^{3+}\text{-}MPG]$ crystal 1 structure showing the formation of the PO_3^- intermediate (Figure S6), unrooted phylogenetic trees based on the amino acid sequences of MpgPs, GpgPs, and homologues and MpgSs, GpgSs, and homologues (Figure S7), and alternative conformations of the catalytic HAD core in the apo state (Figure S8). This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

Coordinates and structure factors were deposited with the Protein Data Bank in Europe[69] as entries 3zty for the Gd^{3+}-soaked structure ($[Gd^{3+}]$), 3ztw for apo-MpgP, 3zu6 for $[Mg^{2+}\text{-}MPG]$ crystal 1, 3zup for $[Mg^{2+}\text{-}MPG]$ crystal 2, 3zw7 for $[Gd^{3+}\text{-}MPG]$ crystal 1, 3zwd for $[Gd^{3+}\text{-}MPG]$ crystal 2, 3zwk for $[MPG\text{-}VO^{2+}]$, 3zx4 for $[Mg^{2+}\text{-}MG\text{-}HPO_4^{2-}]$ (control), and 3zx5 for $[Mg^{2+}\text{-}MG\text{-}VO_4^{3-}]$ (control).

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: +351 21 4469669. Fax: +351 21 4433644. E-mail: matias@itqb.unl.pt.

Notes

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

MPGP, mannosyl-3-phosphoglycerate phosphatase family; MpgP, mannosyl-3-phosphoglycerate phosphatase; MPG, mannosyl-3-phosphoglycerate; MG, mannosylglycerate; HAD-like phosphatase, haloalcanoic acid dehalogenase-like phosphatase; PO_3^- intermediate, metaphosphate intermediate; HPP, *B. thetaiotaomicrometer* predicted hexose phosphate phosphatase; β-PGM, *L. lactis* β-phosphoglucomutase; FBPase, *Sus scrofa* (pig) fructose-1,6-bisphosphatase; TS, transition state; rmsd, root-mean-square deviation.

■ REFERENCES

(1) Santos, H., Lamosa, P., Borges, N., Gonçalves, L. G., Pais, T. M., and Rodrigues, M. V. (2011) Organic Compatible Solutes of Prokaryotes that Thrive in Hot Environments: The Importance of Ionic Compounds for Thermostabilization. In *Extremophiles Handbook* (Horikoshi, K., Antranikian, G., Bull, A. T., Robb, F. T., and Stetter, K. O., Eds.) pp 497−520, Springer, Tokyo.

(2) Borges, N., Marugg, J. D., Empadinhas, N., da Costa, M. S., and Santos, H. (2004) Specialized Roles of the Two Pathways for the Synthesis of Mannosylglycerate in Osmoadaptation and Thermoadaptation of *Rhodothermus marinus*. *J. Biol. Chem. 279*, 9892−9898.

(3) Alarico, S., Empadinhas, N., Simões, C., Silva, Z., Henne, A., Mingote, A., Santos, H., and da Costa, M. S. (2005) Distribution of genes for synthesis of trehalose and mannosylglycerate in *Thermus* spp. and direct correlation of these genes with halotolerance. *Appl. Environ. Microbiol. 71*, 2460−2466.

(4) Neves, C., da Costa, M. S., and Santos, H. (2005) Compatible solutes of the hyperthermophile *Palaeococcus ferrophilus*: Osmoadaptation and thermoadaptation in the order thermococcales. *Appl. Environ. Microbiol. 71*, 8091−8098.

(5) Alarico, S., Empadinhas, N., Mingote, A., Simões, C., Santos, M. S., and da Costa, M. S. (2007) Mannosylglycerate is essential for osmotic adjustment in *Thermus thermophilus* strains HB27 and RQ-1. *Extremophiles 11*, 833−840.

(6) Flint, J., Taylor, E., Yang, M., Bolam, D., Tailford, L., Martinez-Fleites, C., Dodson, E., Davies, B., Gilbert, H., and Davies, G. (2005) Structural dissection and high-throughput screening of mannosyglycerate synthase. *Nat. Struct. Mol. Biol. 12*, 608−614.

(7) Kawamura, T., Watanabe, N., and Tanaka, I. (2008) Structure of mannosyl-3-phosphoglycerate phosphatase from *Pyrococcus horikoshii*. *Acta Crystallogr. D64*, 1267−1276.

(8) Gonçalves, S., Borges, N., Esteves, A. M., Victor, B. L., Soares, C. M., Santos, H., and Matias, P. M. (2010) Structural analysis of *Thermus thermophilus* HB27 mannosyl-3-phosphoglycerate synthase provides evidence for a second catalytic metal ion and new insight into the retaining mechanism of glycosyltransferases. *J. Biol. Chem. 285*, 17857−17868.

(9) Empadinhas, N., and da Costa, M. S. (2011) Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environ. Microbiol. 13*, 2056−2077.

(10) Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., and Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: Understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol. 361*, 1003−1034.

(11) Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., Grigoriev, I. V., Wemmer, D. E., and Kim, S. H. (2002)

Structural characterization of the reaction pathway in phosphoserine phosphatase: Crystallographic "snapshots" of intermediate states. *J. Mol. Biol. 319*, 421−431.

(12) Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science 299*, 2067−2071.

(13) Lu, Z., Dunaway-Mariano, D., and Allen, K. N. (2008) The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state. *Proc. Natl. Acad. Sci. U.S.A. 105*, 5687−5692.

(14) Pragai, Z., and Harwood, C. R. (2002) Regulatory interactions between the Pho and σ(B)-dependent general stress regulons of *Bacillus subtilis*. *Microbiology 148*, 1593−1602.

(15) Rao, K. N., Kumaran, D., Seetharaman, J., Bonanno, J. B., Burley, S. K., and Swaminathan, S. (2006) Crystal structure of trehalose-6-phosphate phosphatase-related protein: Biochemical and biological implications. *Protein Sci. 15*, 1735−1744.

(16) Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H. C., Hirai, A., Tsuzuki, K., Nakamura, S., Altaf-Ul-Amin, M., Oshima, T., Baba, T., Yamamoto, N., Kawamura, T., Ioka-Nakamichi, T., Kitagawa, M., Tomita, M., Kanaya, S., Wada, C., and Mori, H. (2006) Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. *Genome Res. 16*, 686−691.

(17) Cleland, W. W., and Hengge, A. C. (1995) Mechanisms of phosphoryl and acyl transfer. *FASEB J. 9*, 1585−1594.

(18) Aqvist, J., Kolmodin, K., Florian, J., and Warshel, A. (1999) Mechanistic alternatives in phosphate monoester hydrolysis: What conclusions can be drawn from available experimental data? *Chem. Biol. 6*, R71−R80.

(19) Re, S., Imai, T., Jung, J., Ten-No, S., and Sugita, Y. (2011) Geometrically associative yet electronically dissociative character in the transition state of enzymatic reversible phosphorylation. *J. Comput. Chem. 32*, 260−270.

(20) Grzyska, P. K., Czyryca, P. G., Purcell, J., and Hengge, A. C. (2003) Transition state differences in hydrolysis reactions of alkyl versus aryl phosphate monoester monoanions. *J. Am. Chem. Soc. 125*, 13106−13111.

(21) Cleland, W. W., and Hengge, A. C. (2006) Enzymatic mechanisms of phosphate and sulfate transfer. *Chem. Rev. 106*, 3252−3278.

(22) Choe, J. Y., Iancu, C. V., Fromm, H. J., and Honzatko, R. B. (2003) Metaphosphate in the active site of fructose-1,6-bisphosphatase. *J. Biol. Chem. 278*, 16015−16020.

(23) Baxter, N. J., Bowler, M. W., Alizadeh, T., Cliff, M. J., Hounslow, A. M., Wu, B., Berkowitz, D. B., Williams, N. H., Blackburn, G. M., and Waltho, J. P. (2010) Atomic details of near-transition state conformers for enzyme phosphoryl transfer revealed by MgF_3^- rather than by phosphoranes. *Proc. Natl. Acad. Sci. U.S.A. 107*, 4555−4560.

(24) Webster, C. E. (2004) High-energy intermediate or stable transition state analogue: Theoretical perspective of the active site and mechanism of β-phosphoglucomutase. *J. Am. Chem. Soc. 126*, 6840−6841.

(25) Gonçalves, S., Esteves, A. M., Borges, N., Santos, H., and Matias, P. M. (2011) Crystallization and preliminary X-ray analysis of mannosyl-3-phosphoglycerate phosphatase from *Thermus thermophilus* HB27. *Acta Crystallogr. F67*, 390−396.

(26) Gorrec, F. (2009) The MORPHEUS protein crystallization screen. *J. Appl. Crystallogr. 42*, 1035−1042.

(27) Gonçalves, S., Borges, N., Santos, H., and Matias, P. M. (2009) Crystallization and preliminary X-ray analysis of mannosyl-3-phosphoglycerate synthase from *Thermus thermophilus* HB27. *Acta Crystallogr. F65*, 1014−1017.

(28) Empadinhas, N., Marugg, J. D., Borges, N., Santos, H., and da Costa, M. S. (2001) Pathway for the synthesis of mannosylglycerate in the hyperthermophilic archaeon *Pyrococcus horikoshii*. Biochemical and genetic characterization of key enzymes. *J. Biol. Chem. 276*, 43580−43588.

(29) Borges, N., Ramos, A, Raven, N. D., Sharp, R. J., and Santos, H. (2002) Comparative study of the thermostabilizing properties of

mannosylglycerate and other compatible solutes on model enzymes. *Extremophiles* 6, 209−216.

(30) Kabsch, W. (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Crystallogr.* 26, 795−800.

(31) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr.* D67, 235−242.

(32) Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* 33, 491−497.

(33) Pape, T., and Schneider, T. R. (2004) HKL2MAP: A graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* 37, 843−844.

(34) Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr.* D58, 1772−1779.

(35) Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. *Z. Kristallogr.* 217, 644−650.

(36) Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated Structure Solution With autoSHARP. In *Macromolecular Crystallography Protocols* (Doublié, S., Ed.) pp 215−230, Humana Press, Totowa, NJ.

(37) Abrahams, J. P., and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr.* D52, 30−42.

(38) Cowtan, K. (1994) 'dm': An automated procedure for phase improvement by density modification. In *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, pp 34−38, Daresbury Laboratory, Warrington, U.K.

(39) Brunger, A. T. (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472−475.

(40) Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr.* D62, 1002−1011.

(41) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* D53, 240−255.

(42) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr.* D66, 486−501.

(43) Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr.* D60, 432−438.

(44) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* D66, 213−221.

(45) Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr.* D62, 439−450.

(46) Vagin, A. A., Murshudov, G. N., and Strokopytov, B. V. (1998) BLANC: The program suite for protein crystallography. *J. Appl. Crystallogr.* 31, 98−102.

(47) Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr.* D59, 1131−1137.

(48) Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) ARP/wARP and molecular replacement. *Acta Crystallogr.* D57, 1445−1450.

(49) Chen, V. B., Arendall, W. B. III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr.* D66, 12−21.

(50) Holm, L., Kaariainen, S., Rosenstrom, P., and Schenke, I. A. (2008) Searching protein structure databases with DaliLite V.3. *Bioinformatics* 24, 2780−2781.

(51) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein alignment in three dimensions. *Acta Crystallogr.* D60, 2256−2268.

(52) Kim, Y., Yakunin, A. F., Kuznetsova, E., Xu, X., Pennycooke, M., Gu, J., Cheung, F., Proudfoot, M., Arrowsmith, C. H., Joachimiak, A., Edwards, A. M., and Christendat, D. (2004) Structure- and Function-based Characterization of a New Phosphoglycolate Phosphatase from *Thermoplasma acidophilum*. *J. Biol. Chem.* 279, 517−526.

(53) Krissinel, E., and Henrick, K. (2007) Inference of macro-molecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774−797.

(54) Ravelli, R. B., Leiros, H. K., Pan, B., Caffrey, M., and McSweeney, S. (2003) Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* 11, 217−224.

(55) Kabsch, W. (1976) Solution for Best Rotation to Relate 2 Sets of Vectors. *Acta Crystallogr.* A32, 922−923.

(56) Hayward, S., and Berendsen, H. J. (1998) Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins* 30, 144−154.

(57) Deng, H., Ray, W. J., Burgner, J. W., and Callender, R. (1993) Comparison of vibrational frequencies of critical bonds in ground-state complexes and in a vanadate-based transition-state analog complex of muscle phosphoglucomutase. Mechanistic implications. *Biochemistry* 32, 12984−12992.

(58) Empadinhas, N., Albuquerque, L., Henne, A., Santos, H., and da Costa, M. S. (2003) The bacterium *Thermus thermophilus*, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate. *Appl. Environ. Microbiol.* 69, 3272−3279.

(59) Cassano, A. G., Anderson, V. E., and Harris, M. E. (2004) Understanding the transition states of phosphodiester bond cleavage: Insights from heavy atom isotope effects. *Biopolymers* 73, 110−129.

(60) Kirby, A. J., and Varvoglis, A. G. (1967) The Reactivity of Phosphate Esters. Monoester Hydrolysis. *J. Am. Chem. Soc.* 89, 415−423.

(61) Lu, Z., Dunaway-Mariano, D., and Allen, K. N. (2005) HAD superfamily phosphotransferase substrate diversification: Structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry* 44, 8684−8696.

(62) Koumanov, A., Ruterjans, H., and Karshikoff, A. (2002) Continuum electrostatic analysis of irregular ionization and proton allocation in proteins. *Proteins* 46, 85−96.

(63) Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W., and Allen, K. N. (2005) Catalytic cycling in β-phos-phoglucomutase: A kinetic and structural analysis. *Biochemistry* 44, 9404−9416.

(64) Diederichs, K., and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* 4, 269−275.

(65) Kabsch, W., and Sander, C. (1983) Dictionary of Protein Secondary Structure: Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* 22, 2577−2637.

(66) Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* 26, 283−291.

(67) Hutchinson, E. G., and Thornton, J. M. (1996) PROMOTIF: A program to identify and analyze structural motifs in proteins. *Protein Sci.* 5, 212−220.

(68) Westhead, D. R., Slidel, T. W., Flores, T. P., and Thornton, J. M. (1999) Protein structural topology: Automated analysis and diagrammatic representation. *Protein Sci.* 8, 897−904.

(69) Velankar, S., and Kleywegt, G. J. (2011) The Protein Data Bank in Europe (PDBe): bringing structure to biology. *Acta Crystallogr.* D67, 324−330.