THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 15, Issue of April 9, pp. 15491–15498, 2004

Printed in U.S.A.

Crystal Structure of Imidazole Glycerol-phosphate Dehydratase

DUPLICATION OF AN UNUSUAL FOLD*

Received for publication, November 20, 2003, and in revised form, January 13, 2004
Published, JBC Papers in Press, January 14, 2004, DOI 10.1074/jbc.M312733200

Sangita C. Sinha‡§, Barnali N. Chaudhuri‡¶, John W. Burgner‡, Galina Yakovleva‖, V. Jo Davisson‖, and Janet L. Smith‡**

From the ‡Department of Biological Sciences and the ‖Department of Medicinal Chemistry and Pharmacology, Purdue University, West Lafayette, Indiana 47907

Imidazole glycerol-phosphate dehydratase (IGPD) catalyzes the sixth step of histidine biosynthesis. The enzyme is of fundamental biochemical interest, because it catalyzes removal of a non-acidic hydrogen atom in the dehydration reaction. It is also a potential target for development of herbicides. IGPD is a metalloenzyme in which transition metals induce aggregation and are required for catalysis. Addition of 1 equivalent of Mn^{2+}/subunit is shown by analytical ultracentrifugation to induce the formation of 24-mers from trimeric IGPD. Two histidine-rich motifs may participate in metal binding and aggregation. The 2.3-Å crystal structure of metal-free trimeric IGPD from the fungus *Filobasidiella neoformans* reveals a novel fold containing an internal repeat, apparently the result of gene duplication. The 95-residue α/β half-domain occurs in a few other proteins, including the GHMP kinase superfamily (galacto-homoserine-mevalonate-phosphomevalonate), but duplication to form a compact domain has not been seen elsewhere. Conserved residues cluster at two types of sites in the trimer, each site containing a conserved histidine-rich motif. A model is proposed for the intact, active 24-mer in which all highly conserved residues, including the histidine-rich motifs in both the N- and C-terminal halves of the polypeptide, cluster at a common site between trimers. This site is a candidate for the active site and also for metal binding leading to aggregation of trimers. The structure provides a basis for further studies of enzyme function and mechanism and for development of more potent and specific herbicides.

Histidine is an essential dietary nutrient for animals but is synthesized *de novo* by plants and microorganisms. Thus, the biosynthetic pathway is a potential target for herbicide development. Histidine biosynthesis from the precursor phosphoribosyl pyrophosphate is a complex process involving nine enzyme-catalyzed steps. Many intermediates in the pathway are unstable, which has slowed the structural and mechanistic study of individual steps. Imidazole glycerol-phosphate dehydratase (IGPD)[1] catalyzes the sixth step in this pathway, the dehydration of imidazole glycerol phosphate (IGP) to imidazole acetol phosphate (1). Several IGPD inhibitors have been identified, including derivatives of triazole propyl phosphonic acid (2–4). Some of the inhibitors have substantial herbicidal activity (5). The IGPD-catalyzed reaction may have an unusual molecular mechanism because the leaving C-2 hydrogen of IGP is not acidic. In contrast, the leaving hydrogen in most enzyme-catalyzed dehydrations is relatively acidic because of an adjacent carbonyl or imine group (6).

IGPDs from fungi (3, 7), plants (8, 9), archaea, and some eubacteria are monofunctional. Other eubacteria encode bifunctional enzymes, in which IGPD is fused to histidinol-phosphate phosphatase, the penultimate enzyme of histidine biosynthesis (10–12). Metal ions are essential to IGPD catalysis (1), but the role of metal in promoting the reaction is poorly defined. In the absence of metals, plant and fungal IGPDs are stable inactive trimers (3, 13, 14). Mn^{2+} induces aggregation to a catalytically competent form that appears to be a 24-mer (3, 9, 13–15). Aggregation has confounded attempts to obtain a three-dimensional structure for IGPD. Preliminary studies (16) suggested that yeast IGPD crystallized as a 24-mer with molecular octahedral (432) symmetry consistent with its aggregation in solution. However, the cubic crystals diffracted too poorly for structure determination.

The *his3* gene of the fungus *Filobasidiella neoformans* (formerly named *Cryptococcus neoformans*) encodes the 202-residue IGPD polypeptide (7). The sequences of IGPD from *F. neoformans* and other sources have no detectable relationship with any protein of known structure. Hence, the fold and structural organization of IGPD cannot be inferred by homology. Amino acids essential to catalysis have not been identified, but several conserved histidine, glutamate, and aspartate residues may serve as metal ligands or play roles in catalysis. In particular, two repeats of the motif Asx-Xaa-His-His-Xaa-Xaa-Glu ((D/N)*X*HH*XX*E) are potential metal binding sites (7) especially as histidine imidazoles have been implicated in metal binding (17). Two occurrences of this motif (residues 69–75 and 165–171 in *F. neoformans* IGPD) are suggestive of an ancient gene duplication event.

IGPD is an interesting candidate for structural studies be-

* This work was supported by National Institutes of Health Grants GM-45756 (to V. J. D.) and DK-42303 (to J. L. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1RHY) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

§ Current address: Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390.

¶ Current address: UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Box 951570, UCLA, 611 Charles Young Dr., Los Angeles, CA 90095-1570.

** To whom correspondence should be addressed. Tel.: 765-494-9246; Fax: 765-496-1189; E-mail: smithj@purdue.edu.

[1] The abbreviations used are: IGPD, imidazole glycerol-phosphate dehydratase; IGP, imidazole glycerol phosphate; EPPS, 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid; HTPP, 2-hydroxy-3-(1,2,4-triazol-1-yl)propylphosphonate; IRL-1803, 3-hydroxy-3-(1,2,4-triazol-3-yl)cyclohexylphosphonate; GHMP, galacto-homoserine-mevalonate-phosphomevalonate.

TABLE I
Data collection and phasing

		Hg-IGPD						Free enzyme	
Data									
Wavelength		λ_1 1.0086 Å 12.2928 keV		λ_2 1.0078 Å 12.3026 keV		λ_3 0.9150 Å 13.5503 keV		CuK_α 1.5418 Å 8.0416 keV	
Data range (Å)		20–2.3 (2.38–2.30)[a]		20–2.3 (2.38–2.30)		20–2.3 (2.38–2.30)		30–2.6 (2.69–2.60)	
Unique reflections		17,387		17,463		17,324		12,103	
Average multiplicity		12.0		12.2		10.0		1.82	
Completeness (%)		99.1 (100.0)		99.3 (100.0)		98.7 (99.9)		95.6 (84.7)	
R_{sym} (%)[b]		6.4 (26.1)		6.9 (35.0)		6.6 (21.3)		6.9 (33.7)	
$\langle I/\sigma_I \rangle$		16.8 (3.9)		15.2 (2.9)		17.8 (4.6)		15.5 (3.6)	
MAD phasing									
d_{min} (Å)	Overall	6.71	4.76	3.88	3.36	3.00	2.74	2.54	2.37
No. of refs.	17,077	1021	1497	1847	2105	2370	2561	2742	2934
⟨FOM⟩[c]	0.38	0.74	0.68	0.59	0.54	0.42	0.27	0.17	0.10

[a] Values in parentheses pertain to the outermost shell of data.
[b] $R_{sym} = \Sigma_{h,i} |I_{h,i} - \langle I_h \rangle| / \Sigma_{h,i} I_{h,i}$.
[c] ⟨FOM⟩, figure of merit; average estimated cosine of phase error.

cause of the unusual chemical reaction and aggregation properties, the metal dependence of these properties, the lack of sequence similarity to proteins of known structure, and the potential as a herbicide target. We present here the 2.3-Å crystal structure of *F. neoformans* IGPD. The structure reveals a new fold in which an unusual structural motif is duplicated into a single compact domain. The structure also provides clues about the location of the active site, allows modeling of the active 24-mer, is a basis for elucidating the mechanism of dehydration, and aids the development of more potent and specific herbicides.

EXPERIMENTAL PROCEDURES

Purification and Demetallation of F. neoformans IGPD—F. neoformans IGPD was expressed using the *his3* expression vector pHIS3-T7 in *Escherichia coli* strain BL21(DE3)pLysS (7, 13). One-liter cultures were grown at 37 °C in 2× YT medium, 100 μg/ml ampicillin, 34 μg/ml chloramphenicol to an A_{550} of 1.0 and for 8 h after induction with 1.5 mM isopropyl-1-thio-β-D-galactopyranoside. Cells were harvested by centrifugation at 4200 × *g* for 10 min, resuspended in 40 ml of 20 mM EPPS, pH 8.1 (Buffer A), and disrupted by sonication. DNA was removed with 1% (w/v) in streptomycin sulfate. IGPD was purified in three steps by anion exchange chromatography (Q-Sepharose column in Buffer A with a 0–1 M NaCl gradient), precipitation in 25% saturated $(NH_4)_2SO_4$, and size-exclusion chromatography (Sephacryl S-400 column in Buffer A). Enzymatic activity was assayed by monitoring absorbance of the enol form of imidazole acetol phosphate (3). The yield from 1 liter of *E. coli* culture was 95 mg of purified IGPD with a specific activity of 5.8 μmol/min/μg of IGPD at 30 °C. Purified IGPD was concentrated to 2.8 mg/ml and stored at −80 °C.

F. neoformans IGPD was demetallated by a 6-h incubation at 55 °C in Buffer A containing 4 M urea, 400 μM EDTA, 400 μM EGTA, dialyzed for 1 h at room temperature and 14 h at 4 °C against Buffer A containing 3 g/liter Chelex 100, and then concentrated to ~14 mg/ml (0.5 mM) by vacuum dialysis. IGPD was reconstituted by the addition of $MnCl_2$ to a final concentration of 1 mM then the IGPD was stored at 4 °C and used for crystallization experiments within 48 h.

Analytical Ultracentrifugation—Demetallated IGPD was dialyzed against 0.2 M KCl, 50 mM ethanolamine, pH 7.5, which had been pretreated with Chelex 100. For sedimentation velocity experiments, a standard two-sector carbon-filled Epon centerpiece with quartz windows was treated with 0.1 M EDTA, pH 7.5, for 24 h prior to use, washed with Chelex-treated water to remove EDTA, filled with 150 μl of 1 mg/ml metal-free IGPD, and incubated in a Beckman analytical ultracentrifuge (model XLA) for 1 h at 20 °C. Absorbance scans of metal-free IGPD were collected at 5-min intervals during sedimentation at 42,000 rpm. The centrifuge was stopped, 1 equivalent of Mn^{2+}/IGPD monomer was added to the cell, the solution was carefully mixed, and the cell was incubated in the centrifuge for 1 h. Absorbance scans of Mn^{2+}-containing IGPD were collected at 5-min intervals during sedimentation at 22,000 rpm. The resulting scans were analyzed using the continuous distribution (c(*s*)) analysis module in the program Sedfit v8.7 (analyticalultracentrifugation.com). Partial specific volumes and solvent densities were calculated using the program SEDNTERP v1.07, which was obtained from the RASMB program depository (rasmb.bbri.org/rasmb/spin/ms_dos/sednterp-philo/).

TABLE II
Crystallographic refinement

	Hg-IGPD	Free enzyme
Data range (Å)	20.0–2.3	30.0–2.6
Data cutoff	$I/\sigma_I > 0$	$I/\sigma_I > 0$
R_{work}[a]	0.189	0.202
R_{free}[a]	0.228	0.251
Residues, monomer A	2–66, 73–187	2–65, 73–188
Residues, monomer B	2–65, 73–188	2–65, 73–188
Mercury	6	
Sulfate ions	5	4
Acetate ions	7	1
Glycerol	7	0
Water	247	148
r.m.s.d.[b] from target values		
Bond lengths	0.010 Å	0.009 Å
Bond angles	1.7°	1.59°
Δ*B* between bonded atoms	5.1 Å^2	3.0 Å^2
Average *B* values (Å^2)		
Protein	43.2	46.9
Solvent	71.9	63.1
r.m.s.d. between monomers	0.41 Å, 167 C_α	0.58 Å, 172 C_α

[a] R factor $= \Sigma_h |F_{obs}| - |F_{calc}| / \Sigma_h |F_{obs}|$.
[b] r.m.s.d., root mean square deviation.

Crystallization and Data Collection—IGPD was crystallized by hanging drop vapor diffusion from a 1:1 mixture of protein solution (14 mg/ml IGPD, 1 mM $MnCl_2$, 20 mM HEPPS, pH 8.1) and reservoir solution (0.9 M $(NH_4)_2SO_4$, 100 mM sodium acetate, pH 5.0). Crystals of cubic morphology grew to an average dimension of 0.2 mm in approximately 1 week. A mercury-derivative crystal was prepared by a 1.25-h soak in reservoir solution containing 1.0 mM ethyl mercury phosphate, followed by 0.5-h back soak in reservoir solution. IGPD was also crystallized with 1–2 mM concentration of two triazole phosphonic acid derivatives, 2-hydroxy-3-(1,2,4-triazol-1-yl)propylphosphonate (HTPP) (2) and 3-hydroxy-3-(1,2,4-triazol-3-yl)cyclohexylphosphonate (IRL-1803) (4).

Crystals were cryoprotected by successive transfers through reservoir solution with increasing glycerol to a final concentration of 25% (v/v) and flash-frozen in a nitrogen cold stream at 100 K. Multiwavelength anomalous diffraction data at the mercury L_{III} edge were recorded at BM-14 at the European Synchrotron Radiation Facility using a Mar 345 imaging plate detector. At each wavelength, data were recorded in two 30° sweeps related by inverse geometry. The same regions of reciprocal space were recorded in "high resolution" and "low resolution" passes. Data from native crystals and from IGPD crystals grown with either HTPP or IRL-1803 were recorded on an R-axis IIc imaging plate using CuK_α radiation. Data were processed using the HKL package (18). IGPD crystallized in the cubic space group $P2_13$ with unit cell parameter 105.3 Å, which is consistent with two polypeptides/

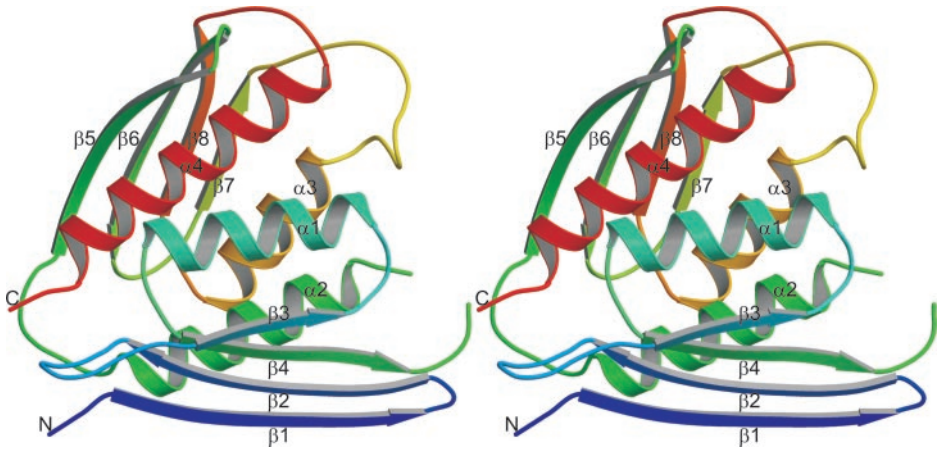

FIG. 1. **Polypeptide fold of IGPD.** The stereo ribbon diagram is color-ramped from *blue* at the N terminus to *red* at the C terminus. Secondary structures are labeled as are their terminal residues. This figure and other molecular illustrations were made using MolScript (37) and Raster3D (38).

asymmetric unit (V_m = 2.2 Å^3/Da, ~42% solvent). Data quality is summarized in Table I.

Multiwavelength Anomalous Diffraction Structure Determination and Refinement—Structure determination was complicated by translational non-crystallographic symmetry along the body diagonal of the cubic unit cell, revealed by a peak in the native Patterson map (~10% of the origin peak) at u = v = w = 0.47. Two mercury sites were located by inspection of a Patterson map calculated with coefficients $|F_\Delta|$, which were derived from the multiwavelength anomalous diffraction data using the program SOLVE (19). Four additional mercury sites were found using automated procedures in SOLVE. The final figure of merit was 0.38 for six mercury sites (Table I). Ten cycles of phase refinement via 2-fold averaging, solvent flattening, and histogram matching in the program DM (20) yielded an interpretable 2.37-Å map from $|F_{obs}|$, ϕ_{best}.

The model was built into a 2.37-Å electron density map using the program O (21). All refinement was done with maximum likelihood amplitude and phase probability targets using CNS (22) and the data at λ_3 (Table II). Atomic occupancies of protein residues in dual positions were refined once. The final refined model consists of two IGPD monomers, A and B, related by an approximate 2-fold axis (179.5° rotation) parallel to the cubic unit cell axis. Monomers A and B reside in different IGPD trimers and give rise to the native Patterson peak. In each monomer, one short internal peptide and 14–15 C-terminal residues are disordered and missing from the model. Met-1 was not present in the crystallized protein (data not shown). Three out of four cysteine side chains were modified with mercury. No potential Mn^{2+} sites were identified in electron density maps. Arg-97 and Asp-106 in each monomer lie in disallowed regions of the Ramachandran plot; however, their backbone conformations are well supported by electron density. The conformation of Arg-97 (ϕ = +70, ψ = −53) is stabilized by backbone hydrogen bonds of both the NH and C=O groups, by a bidentate salt bridge to Glu-115 and by stacking with the Tyr-98 side chain. The conformation of Asp-106 (ϕ = +47, ψ = −116) is stabilized by its position in a type II β-turn (residues 105–108) and by a hydrogen bond of the backbone C=O and side chain of His-157 in an adjacent monomer. Arg-97, Asp-106, Glu-115, and His-157 are invariant among IGPD sequences, suggesting that these residues as well as their conformations may be important for IGPD function. The model is available in the Protein Data Bank with accession code 1RHY.

Crystals of the free enzyme and those grown in presence of inhibitors HTPP or IRL-1803 were isomorphous with mercury-IGPD crystals. For each of these structures, the starting model for refinement was based on the mercury-IGPD structure. The three structures are virtually identical. The most reliable model was obtained from crystals grown without inhibitor (Table II). No density was observed for Mn^{2+}, HTPP, or IRL-1803 in these structures. The absence of bound Mn^{2+} was confirmed by the examination of anomalous difference Fourier maps. The discussion is limited to the higher resolution mercury-IGPD structure because there are no significant differences between the mercury-IGPD and free-enzyme structures (root mean square deviation is 0.38 Å for 180 C_α atoms).

Modeling of an IGPD 24-Mer—Models of IGPD 24-mers with octahedral symmetry were constructed from the crystallographic trimers. Using the trimer as a rigid body, a two-dimensional search was done through 120° rotation in 10° steps around the trimer 3-fold axis and through 140-Å translation in 5-Å steps along the trimer 3-fold. This search covers all of the rotation space, and a translation in which the ends of the range have no intertrimer contacts. At each grid position, a 24-mer was constructed from the trimer by application of 4- and 2-fold symmetry. Solutions were scored by tabulating the number of "bad" (d < 2.5 Å) and "good" (2.5 Å < d < 3.6 Å) interatomic contacts between trimers and the ratio of good to bad contacts. Full atomic coordinates were used for the search, which was done with trimers constructed from both the crystallographic A and B monomers. Most of the search space yielded impossible solutions with interpenetrating trimers. Fine searches (1° rotation step, 1-Å translation step) were done around six solutions with good:bad ratios of 3.5–16. Solutions were evaluated by visual inspection.

RESULTS AND DISCUSSION

Structure Determination—F. neoformans IGPD produced in *E. coli* had a heterogeneous metal content and yielded poorly diffracting crystals. To improve crystal quality and to label the protein with a single anomalous scatterer, IGPD was demetallated to produce the catalytically inactive 70-kDa species and then reconstituted with each of several metals. Many of the reconstituted proteins crystallized, but only crystals of IGPD reconstituted with Mn^{2+} were suitable for high resolution analysis. As Mn^{2+} does not have an x-ray absorption edge convenient for multiwavelength anomalous diffraction, crystals were treated with mercury, and the structure was determined by mercury multiwavelength anomalous diffraction. The crystallographic model agrees well with the diffraction data and with stereochemical criteria (Table II). The estimated coordinate error is 0.29 Å (23). The final 2.3-Å model includes one monomer in each of two trimers. Monomers A and B are nearly identical, except for two loops comprising 12 residues (11–15 and 25–31). The variable loop conformations are not influenced by crystal lattice contacts. The remaining 167 C_α atoms superimpose with a root mean square deviation of 0.4 Å. Likewise, the trimers of monomers A and B are nearly identical with a root mean square deviation of 0.4 Å for 501 C_α atoms.

IGPD Fold: Gene Duplication and Unique Arrangement of a Rare Structural Motif—The IGPD polypeptide forms a single domain (Fig. 1) consisting of a bundle of four α-helices (α1-α4), sandwiched between four-stranded, mixed β-sheets (β1-β2-β4-β3 and β5-β6-β8-β7). The fold possesses an internal repeat, in which the first β-sheet and two α-helices (β1-β2-β3-α1-β4-α2, residues 2–93) have an identical topology to the second β-sheet and two α-helices (β5-β6-β7-α3-β8-α4, residues 94–187). The half-domain motif includes a rare left-handed $\beta\alpha\beta$ crossover (β3-α1-β4 and β7-α3-β8). The half-domains are related by pseudo-dyad symmetry (157° rotation) such that 77 pairs of C_α atoms superimpose with a root mean square deviation of 2.1 Å (Fig. 2*A*).

The structural repeat matches an internal sequence repeat exhibited by IGPDs from all biological sources. The sequences of residues 1–93 and 94–202 of *F. neoformans* IGPD are 19% identical (Fig. 2*B*). The most conserved feature of the internal

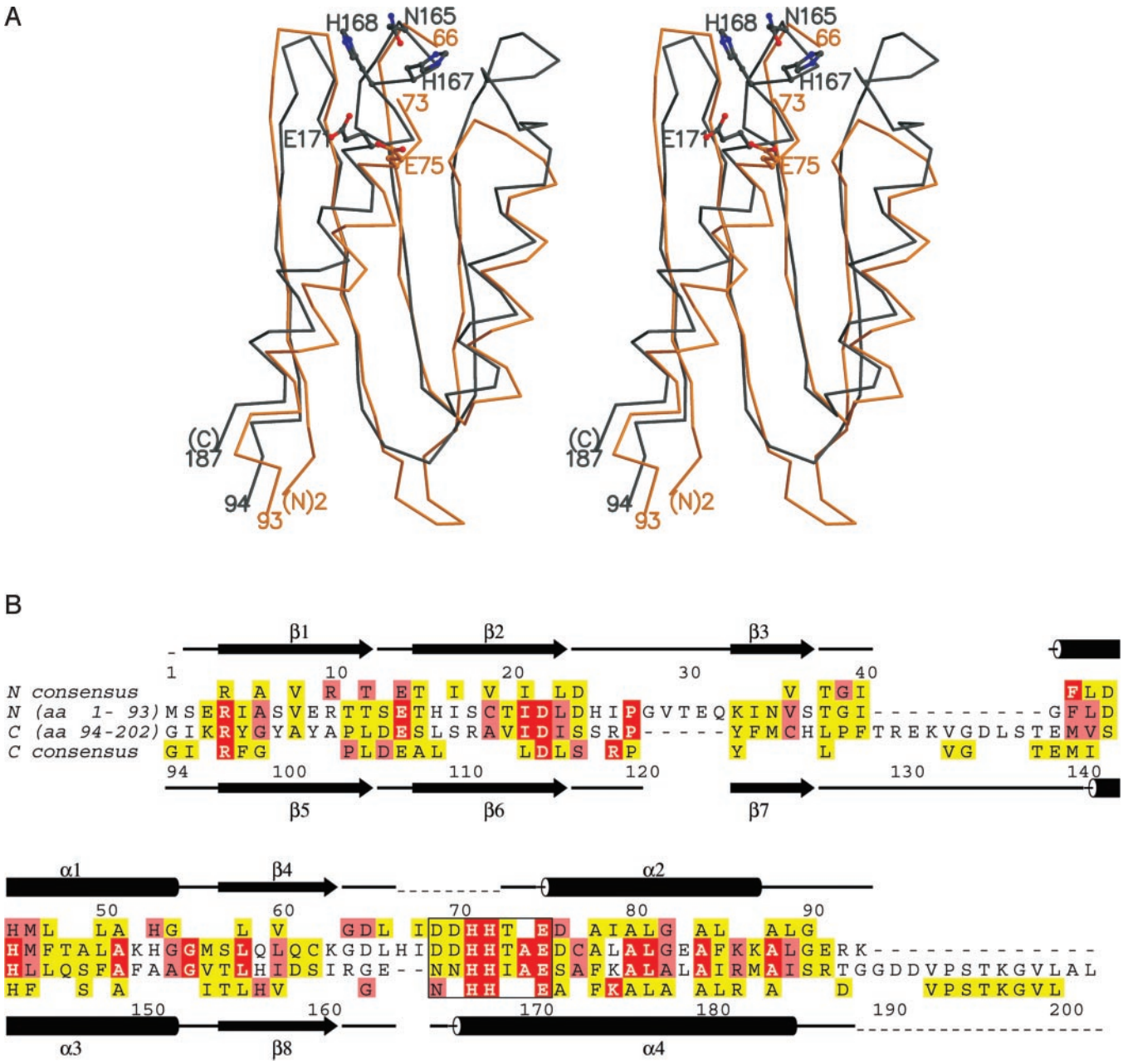

FIG. 2. **Internal repeat in IGPD.** *A*, stereo view of the superposition of C_α traces of the N and C-terminal halves of the IGPD monomer. The N-terminal half is traced in *yellow*, and the C-terminal half is in *gray*. All atoms are shown for conserved side chains in the C-terminal (D/N)*X*HH*XX*E motif, colored by atomic type (*gray*, carbon; *red*, oxygen; *blue*, nitrogen). Residues at the N- and C terminus of each half-fold are indicated. This and other superpositions were done using the program O (21). *B*, structure-based sequence alignment of the two halves of IGPD. Evidence for gene duplication and fusion is found in conservation in a multiple-sequence alignment (39) of IGPD from 56 organisms, indicated as N-terminal (*above*) and C-terminal (*below*) consensus sequences. Secondary structure elements corresponding to the *F. neoformans* IGPD structure are indicated *above* and *below* the sequence alignment. α-Helices are represented by *cylinders*, β-strands by *arrows*, loops by *black solid lines*, and disordered regions by *dashed lines*. The degree of conservation is indicated by the background color, with invariant residues in *red* (*white characters*), 95% invariant in *pink*, and sites of conservative substitution in *yellow*. The (D/N)*X*HH*XX*E motif is *outlined*. This figure was made using ALSCRIPT (40). *aa*, amino acids.

repeat is the (D/N)*X*HH*XX*E motif in loops β3-α1 (residues 69–75) and β8-α4 (residues 165–171). The repetition of sequence and structural elements suggests a gene duplication in the evolution of IGPD. Gene duplication is a recurring theme in enzymes of histidine biosynthesis. The enzymes that catalyze the two steps preceding the dehydration reaction of IGPD, phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase and imidazole glycerol-phosphate synthase, also have internal repeats and are thought to have evolved by similar gene duplication and fusion events (24). These enzymes are homologs but are unrelated to IGPD.

The structure data base contains no proteins with folds like the IGPD fold, based on a topology search (25) of the Protein Data Bank. However, several proteins have subdomains with topologies identical to the IGPD half-domain (Table III and Fig. 3). The fold was recognized first in a few nucleic acid binding proteins (26). Recently the half-domain was found in several other proteins and is a conserved structural feature of the GHMP kinase superfamily (27) (Table III). The structure-based sequence identity of the N- or C-terminal half-domain of IGPD with any of these proteins is 6–12%, indicative of a very ancient relationship, if any. Apart from the GHMP superfamily, the

TABLE III
Structures with the IGPD half-fold

r.m.s.d., root mean square deviation.

Protein	Protein Data Bank code	Ref.	r.m.s.d.	C_α atoms superimposed[a]	Sequence identity[b]	Domain locale[c]	Function[d]
			Å		%		
IGPD: C-terminal half domain			2.10	77	19	D	S, C
Multi fragment	1b63	41	3.4	69	9	I	U
Phosphomevalonate kinase[e]	1k47	42	3.5	68	12	D	C
Polynucleotide phosphorylase	1e3p	28	4.6	78	8	D	S
			4.2	79	11		
Elongation factor G	1dar	43	4.4	75	11	I	U
Mevalonate kinase[e]	1kvk	44	3.3	74	9	D	C
Homoeserine kinase[e]	1fwk	27	3.7	69	12	D	C
Mevalonate diphosphate decarboxylase	1fi4	45	3.7	71	10	D	C
16 S ribosomal protein S9	1fjg	46	3.6	65	9	I	U
RNase P protein	1a6f	47	4.2	66	6	I	U

[a] Superpositions on the IGPD N-terminal half-domain.
[b] Sequence identity based on structure superposition.
[c] D, part of a larger domain; I, independent of other secondary structures.
[d] S, subunit contacts; C, catalysis; U, unknown.
[e] Member of the GHMP kinase superfamily.

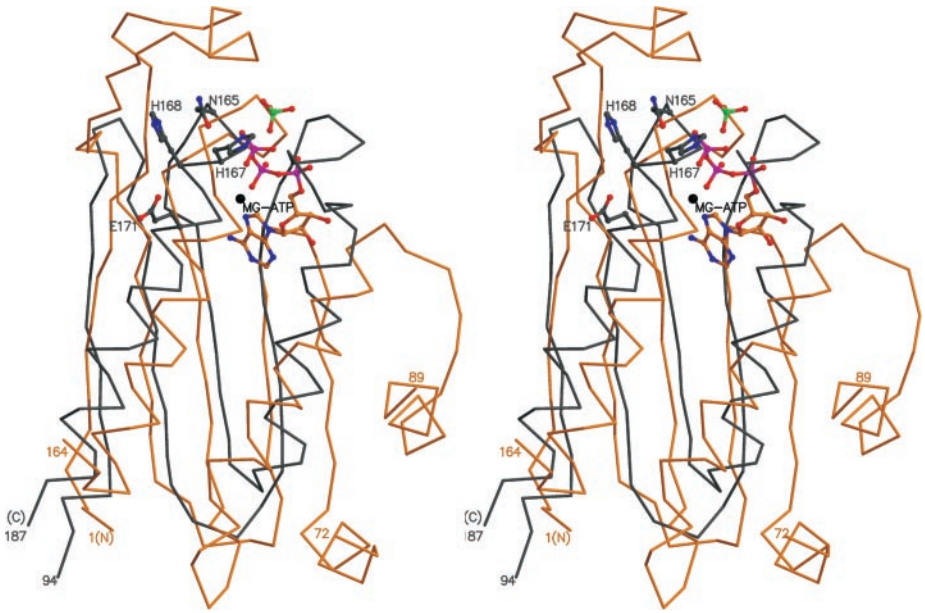

FIG. 3. **The IGPD half-domain in other proteins.** The C-terminal half-domain of IGPD (*gray*) and the N-terminal region of mevalonate kinase (*yellow*) (Protein Data Bank 1KVK) are superimposed in this stereo C_α trace. All atoms are shown for side chains of conserved residues in the C-terminal (D/N)*X*HH*XX*E motif of IGPD, along with a sulfate ion from the crystallization solution. The N-terminal domain of mevalonate kinase binds ATP such that the ATP β-phosphate and the sulfate ion in IGPD are in analogous positions at a helix N terminus. Atomic coloring (*red*, oxygen; *blue*, nitrogen; *magenta*, phosphorus; *green*, sulfur) is used for atomic details with *gray* carbon for IGPD and *yellow* carbon for mevalonate kinase. Residue numbers at the polypeptide termini are indicated.

motif is not associated with a common function.

The presence of the IGPD half-domain motif in other proteins supports the hypothesis that it constitutes an initial folding unit, which has evolved to produce the IGPD subunit by gene duplication. Duplication of this motif within a protein is not unprecedented. Analogs of the IGPD half-domain occur twice in *Streptomyces antibioticus* polynucleotide phosphorylase, presumably also through gene duplication and fusion events (28). However, the two repeats do not associate as they do in IGPD. Interestingly, both polynucleotide phosphorylase repeats are involved in subunit contacts within a trimer, but these also differ from the IGPD trimer contacts.

In other proteins, the IGPD half-domain occurs either as one element of a larger domain or as an independent domain (Table III). The arrangement of two half-folds into a single domain lacking other secondary structures is unique to IGPD.

Quaternary Structure in Solution—The quaternary structures of metal-free and Mn^{2+}-containing IGPD were determined by velocity analytical ultracentrifugation using the program Sedfit to determine the distribution of protein species with different sedimentation coefficients (29, 30). The distribution (c(*s*)) of sedimentation coefficient (*s*) for metal-free IGPD and for IGPD containing 1 equivalent of Mn^{2+}/monomer (~50 μM) are compared in Fig. 4*A*. Within measurement limits of the instrument, no protein was lost in the conversion of metal-free to Mn^{2+}-containing IGPD. In each c(*s*) scan, a single predominant peak accounts for more than 98% of the total protein. These data show that the addition of 1 Mn^{2+}/monomer induces the formation of an aggregate that has eight times the mass of the single species present without metal. Masses of 66.6 kDa for the metal-free protein and 549 kDa for Mn^{2+}-containing IGPD were calculated from sedimentation equilibrium experiments (31) run under the same conditions (data not shown). These masses correspond to the trimer observed in the crystal structure and to an octamer of trimers (24-mer). The absence of peaks of intermediate sedimentation coefficient suggests that the assembly of 24-mers is highly cooperative.

Quaternary Structure in the Crystal—IGPD crystallized as a trimer (Fig. 4*B*), which is its quaternary structure in solution in absence of metals (Fig. 4*A*). Trimeric aggregation is consistent with the lack of bound Mn^{2+} in the crystal structure, despite the presence of Mn^{2+} in crystallization solutions. We infer that the crystalline trimer corresponds to the solution trimer because of the large size and extensive hydrophobic character of the interface. About 15% (1086 $Å^2$) of the surface area of each monomer is buried in subunit interfaces. Subunit

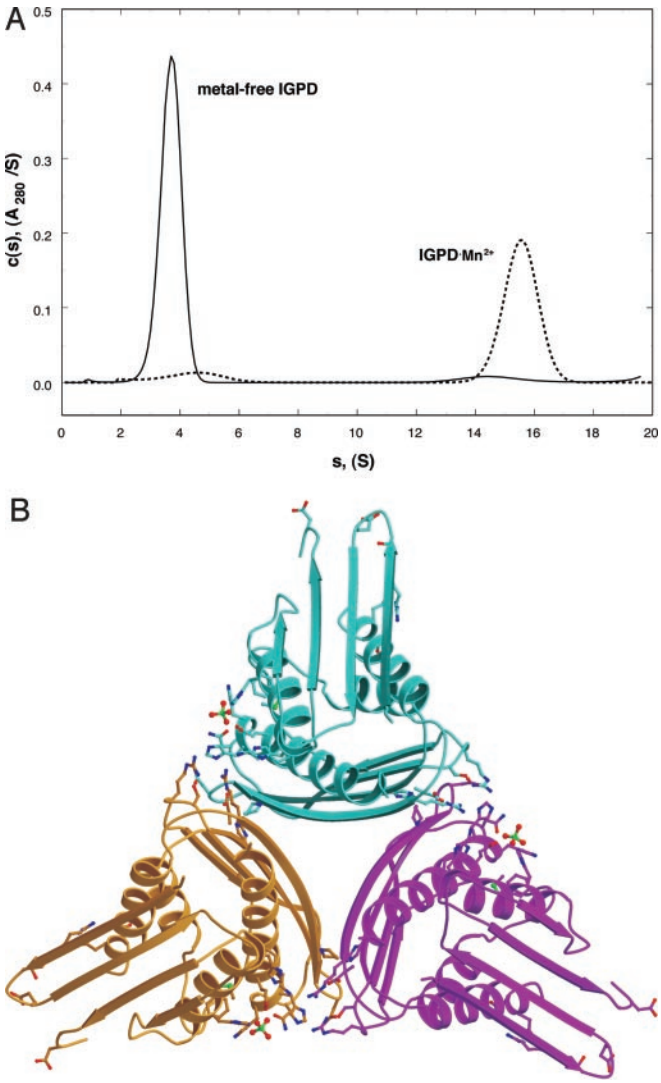

FIG. 4. **IGPD quaternary structure.** *A*, distribution of sedimentation coefficients (c(*s*)) for metal-free IGPD (*solid line*) and the Mn^{2+} complex (*dashed line*) with 1 equivalent of Mn^{2+}/subunit. The distributions are calculated from the radial position of the boundary as a function of time. The calculated values for the observed weight-average sedimentation coefficient (*s*) and frictional ratio (f/f_0) are s = 3.7 S and f/f_0 = 1.24 for demetallated IGPD and s = 15.5 S, f/f_0 = 1.18 for the Mn^{2+}-containing enzyme. *B*, IGPD trimer. The subunits are rendered in different colors in this ribbon diagram viewed along the 3-fold axis. Atomic details are shown for side chains of highly conserved residues and for the sulfate bound to the C-terminal (D/N)*X*HH*XX*E motif at the N terminus of helix α4. Conserved cluster 1 is at the outermost edge of each subunit and includes Arg-10 (β1), Thr-12 (β1), Glu-14 (β1-β2), Gly-39 (β3-α1), Gly-64 (β4-α2), Asp-65 (β4-α2), His-71 (β4-α2), His-72 (β4-α2), Glu-75 (α2), Asp-76 (α2), and Gly-82 (α2). Conserved cluster 2 is between the subunits and includes Phe-42 (α1), His-45 (α1), Met-46 (α1), His-53 (α1), Pro-104 (β5), Asp-106 (β5-β6), Gly-163 (β8-α4), Asn-165 (β8-α4), His-167 (α4), His-168 (α4), Glu-171 (α4), and Lys-175 (α4) from one subunit and Arg-97 (β5), Asp-115 (β6), Ser-117 (β6-β7), Arg-119 (β6-β7), and His-157 (β8) from an adjacent subunit. Carbon atoms are colored to match the ribbon; other atoms are described in the legend to in Fig. 3.

contacts in the trimer are made by the C-terminal half-domain such that the β-sheet (β5-β6-β8-β7) forms one wall of a central triangular core with the β-strands approximately perpendicular to the 3-fold axis and side chains forming the subunit interface. Because of extensive subunit contacts, the inner C-terminal β-sheet is more ordered than the outer N-terminal β-sheet. The hydrophobic side chains (Leu-179, Tyr-98, Tyr-100, Tyr-102, Ala-103, Pro-104, Leu-109, Val-113, Ile-161, Met-183) and the aliphatic parts of Arg-182, Lys-96, and Arg-111 form an extensive hydrophobic core between subunits. The bottom surface of the trimer is positively charged because of an Arg-4, Arg-92, Lys-93, Lys-96, Arg-97, Lys-175, Arg-182, and Arg-187 from each subunit. Intersubunit contacts at this surface include the side chains of Tyr-98 and Lys-175 and sulfate-mediated salt bridges between Arg-97 of one subunit and Lys-175 and His-53 of another. The top surface of the trimer forms a conical polar depression filled with 18 ordered water molecules.

IGPD does not aggregate beyond the trimer in the crystal structure reported here. This is correlated with the absence of bound Mn^{2+} in crystals, despite a modest 2-fold molar excess of Mn^{2+} in the crystallization solution and its requirement for the growth of diffraction-quality crystals. It is likely that carboxylate buffer (100 mM acetate) or high salt (0.9 M $(NH_4)_2SO_4$) chelated Mn^{2+} from the protein, which crystallized when the metal was gone.

Active Site—More than 50 putative IGPD sequences, having 30–75% pairwise identity, are obvious homologs of *F. neoformans* IGPD. All of these sequences contain an unusually large number of histidines, of which four are invariant. Twenty-eight residues are invariant among at least 95% of IGPD sequences (Fig. 2*B*) and are the basis for our ideas about the IGPD active site. The conserved residues map to two distinct clusters on the IGPD structure (Fig. 4*B*). These sites are obvious candidates for catalysis and metal binding.

Conserved cluster 1 consists of 11 residues from one subunit including the partly disordered N-terminal (D/N)*X*HH*XX*E motif (residues 71–75). Conserved cluster 2 forms a shallow solvent-accessible cleft at the subunit interface and contains the ordered C-terminal (D/N)*X*HH*XX*E motif (residues 165–171) and a total of 17 conserved residues from two subunits. We searched two structural data bases (32, 33) for constellations of active-site residues similar to the positions of invariant residues in clusters 1 and 2 of IGPD. No matches were found between IGPD and any other enzyme active site.

The (D/N)*X*HH*XX*E motif in each conserved cluster in the IGPD trimer is a potential site for metal binding. We predict that the disordered N-terminal (D/N)*X*HH*XX*E motif in conserved cluster 1 becomes ordered upon metal binding, which induces aggregation of trimers to form the active 24-mer. The ordered C-terminal (D/N)*X*HH*XX*E motif in conserved cluster 2 binds a sulfate ion from the crystallization solution, thus demonstrating that this site may accommodate the IGP phosphate. The sulfate binding site at the (D/N)*X*HH*XX*E motif of IGPD is structurally analogous to the binding site for ATP β-phosphate in GHMP kinases, although the remainder of ATP cannot be a model for IGP binding because of steric clashes caused by differences in the protein structures (Fig. 3).

Modeling an IGPD 24-Mer—Catalysis by *F. neoformans* IGPD and aggregation beyond the trimer state are both dependent on metal ions. It is unknown whether metal participates directly in the chemical reaction, as in many dehydratases, or whether catalysis is simply aggregation-dependent. The crystal structure reveals a 60-kDa metal-free trimer, which is known to be inactive. The ultracentrifugation experiments show that Mn^{2+} induces formation of a 24-mer of *F. neoformans* IGPD, corresponding to an octamer of trimers, without detectable intermediate forms (Fig. 4*A*). A similar situation occurs in yeast IGPD, where crystallographic and solution data indicate a 24-mer of octahedral symmetry (3, 13, 15, 16). The simplest explanation for the biochemical and structural data is that metal coordination sites between trimers are the basis for the formation of a 24-mer and that trimers cannot aggregate to form the 24-mer in absence of metal, even at the high protein concentrations of crystallization.

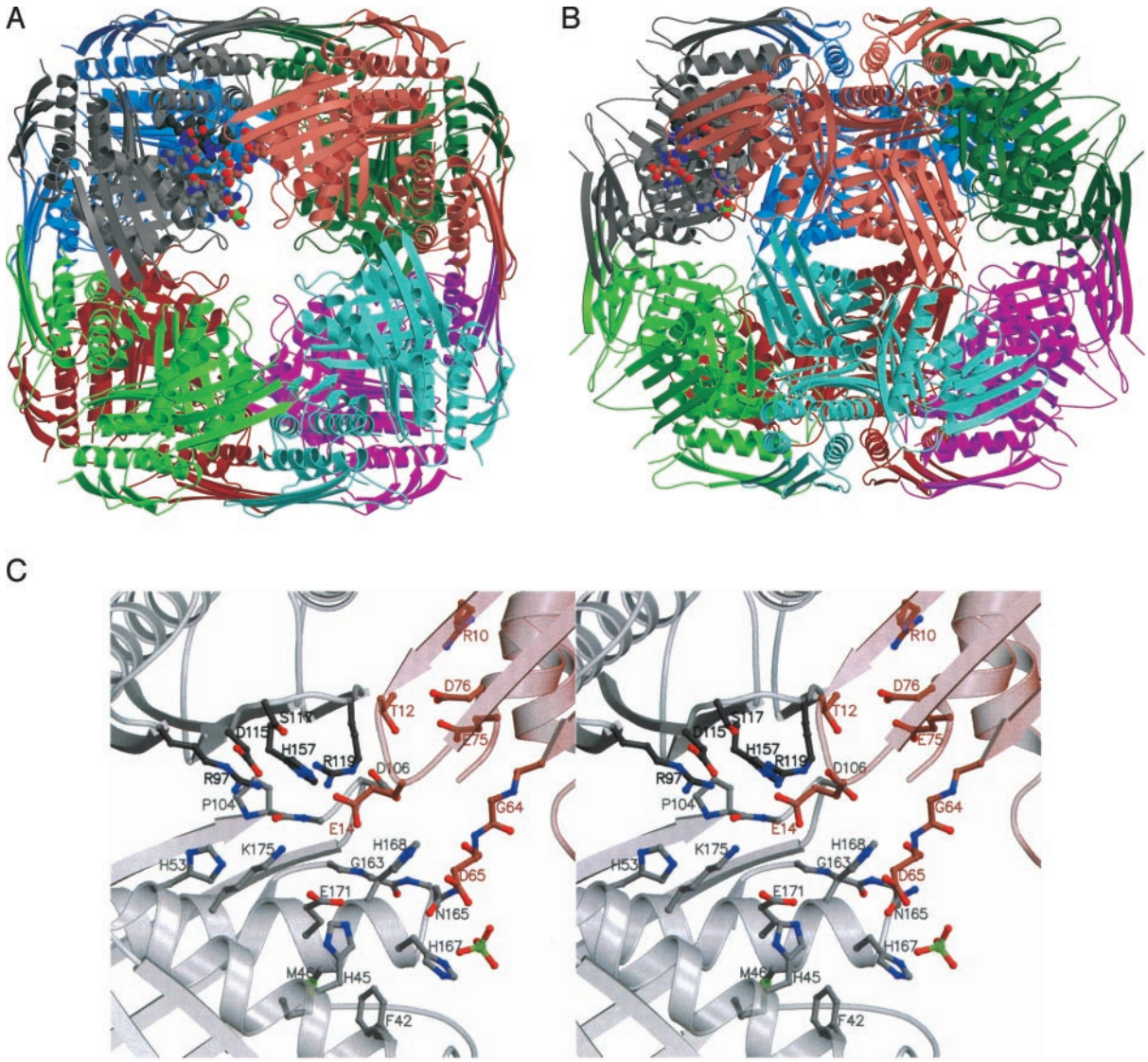

FIG. 5. **Model of IGPD 24-mer.** *A*, view along the octahedral 4-fold axis. In this ribbon diagram, each trimer is colored differently. All side chain atoms are shown for 1 of 24 equivalent conserved sites, formed by a conserved cluster 1 in the *salmon* trimer and a conserved cluster 2 in the *gray* trimer. Atoms are colored as in Fig. 2, with *gray* and *salmon* carbon. *B*, view along the octahedral 2-fold axis. *C*, the proposed active site of the IGPD 24-mer. The *salmon* trimer contributes a conserved cluster 1 and the *gray* trimer a conserved cluster 2, which consists of residues from two subunits, shown in *light* and *dark gray*. The 24-mer contains 24 of these sites. Conserved side chains are shown in atomic detail and labeled in this stereo diagram. A sulfate ion is bound to the C-terminal (D/N)*X*HH*XX*E motif of the *gray* trimer. The site includes disordered residues 66–72 from the *salmon* trimer, including conserved residues Asp-69, His-71, and His-72 from the N-terminal (D/N)*X*HH*XX*E motif. Backbone ribbon and atoms are colored as in *A*, with *dark* and *light gray* for two subunits of the *gray* trimer. To display all residues clearly, the viewing angle differs here and in *A* and *B*.

Models of potential IGPD 24-mers were constructed based on two assumptions. First, the 24-mer aggregate should contain the trimer found in the crystal structure. A constant trimer structure is predicted from the extensive hydrophobic contacts of monomers and the equivalence of monomer A trimers and monomer B trimers in the crystal structure. Second, the 24-mer should be symmetric. We know of no experimental data supporting construction of an asymmetric 24-mer. A symmetric aggregate of 24 identical subunits with a trimeric subassembly must have octahedral symmetry. There are only two degrees of freedom in constructing octahedral 24-mers from trimers, a rotation of the trimer around the octahedron 3-fold axis and a translation of the trimer along the 3-fold axis. All possible 24-mers were built in a two-dimensional rotation-translation search and were evaluated by three criteria. First, the 24-mer should have several favorable interatomic contacts between trimers and few steric clashes. Second, the 24-mer should have potential metal binding sites between trimers. These sites presumably involve either or both of the conserved (D/N)*X*HH*XX*E motifs. Third, the N terminus of each polypeptide in the 24-mer should be positioned to permit fusion to another enzyme of the histidine pathway. Aggregation beyond the trimer state occurs even in those eubacterial IGPDs that are fused (13, 34–36).

Molecular modeling resulted in only three IGPD 24-mers with reasonable contacts between trimers. One of these has N termini inaccessible for fusion and was not considered further. The two plausible 24-mers have identical radial positions for the trimer and differ by a 25° rotation of the trimer within the octahedron. The 24-mers are 120 Å in diameter along the 4- and 2-fold directions and 140 Å in diameter along the 3-fold direction. These dimensions agree well with electron micrographs of negatively stained IGPD in which the equidimensional particle is ~110 Å in diameter (data not shown).

In the first 24-mer, conserved residues occur exclusively in two types of sites, which are considered as potential sites for intertrimer metal binding and for catalysis. The first site oc-

curs at the particle 4-fold axis and is formed by residues in a conserved cluster 1 from each of the four trimers. The boundaries of the four disordered N-terminal (D/N)*X*HH*XX*E motifs are within 12 Å of one another and constitute a plausible intertrimer metal site. The second type of site, containing the C-terminal (D/N)*X*HH*XX*E motif, is not a plausible intertrimer metal site, because it is more than 20 Å from any other (D/N)*X*HH*XX*E motif. Thus, this site, which constitutes all of the conserved cluster 2, is a candidate for the active site. This 24-mer is consistent with a model in which aggregation is cryptically related to catalysis. In this model, metal must participate in catalysis or in substrate binding because the putative active site involves residues from only one trimer, and metal-free trimeric IGPD is inactive. A higher affinity metal site for aggregation and a lower affinity site for catalysis could explain why catalysis is not observed in absence of aggregation. However, this 24-mer cannot explain why two metal binding motifs and higher aggregation in IGPD have been conserved throughout evolution.

These complications do not exist for the second 24-mer in which all highly conserved residues from neighboring trimers form a single conserved site. The boundaries of the disordered N-terminal (D/N)*X*HH*XX*E motif in conserved cluster 1 are within 15 Å of the C-terminal (D/N)*X*HH*XX*E motif in conserved cluster 2 of another trimer (Fig. 5). This 24-mer is consistent with a model in which the active site includes residues from more than one trimer. Thus it explains why crystallization of IGPD in the presence of inhibitors did not yield co-crystals. Evolution of IGPD is also simpler to explain with this model because all highly conserved residues congregate in one site, and for this reason we favor the second 24-mer over the first 24-mer. The model leads to no prediction about the role of metal in catalysis. However, any metal bound to the N- and C-terminal (D/N)*X*HH*XX*E motifs, presumably responsible for formation of the 24-mer, would be near or in the active site (Fig. 5*C*).

Conclusion—The crystal structure of IGPD reveals a new fold in which a half-domain of unusual topology is duplicated to produce a compact domain. Such an arrangement has not been observed previously for the half-domain, which has parallel β-structure with a rare left-handed crossover connection and also occurs in the GHMP kinase family and several other proteins. Each half-domain includes a (D/N)*X*HH*XX*E motif that is highly conserved in IGPD sequences. The active form of IGPD is a 24-mer, which forms in the presence of transition metals such as Mn^{2+}. IGPD crystallized in a metal-free, inactive, trimeric form, despite inclusion of Mn^{2+} in crystallization solutions. Symmetry constraints allowed modeling of an octahedral 24-mer from the observed trimer. One modeled 24-mer stands out as a prospective structure for intact IGPD. In this model, all highly conserved residues congregate in one site, although they derive from four half-domains in three subunits contained in two trimers. This site is proposed as the active site and as a metal binding site. The evolutionary path from a metal-binding half-domain to a gene duplication event to an intersubunit active site in a large symmetric aggregate is interesting to contemplate.

REFERENCES

1. Ames, B. N. (1957) *J. Biol. Chem.* **228,** 131–143
2. Hawkes, T. R., Cox, J. M., Barnes, N. J., Beautement, K., Edwards, L. S., Kipps, M. R., Langford, M. P., Lewis, T., Ridley, S. M., and Thomas, P. G. (1993) *Brighton Crop Protection Conference-Weeds*, Vol. 2, pp. 739–744, Brighton, England
3. Hawkes, T. R., Thomas, P. G., Edwards, L. S., Rayner, S. J., Wilkinson, K. W., and Rice, D. W. (1995) *Biochem. J.* **306,** 385–397
4. Mori, I., Iwasaki, G., Kimura, Y., Matsunaga, S., Ogawa, A., Nakano, T., Buser, H.-P., Hatano, M., Tada, S., and Hayakawa, K. (1995) *J. Am. Chem. Soc.* **117,** 4411–4412
5. Mori, I., Fonne-Pfister, R., Matsunaga, S., Tada, S., Kimura, Y., Iwasaki, G., Mano, J., Hatano, M., Nakano, T., Koizumi, S., Scheidegger, A., Hayakawa, K., and Ohta, D. (1995) *Plant Physiol.* **107,** 719–723
6. Gerlt, J. A., and Gassman, P. G. (1992) *J. Am. Chem. Soc.* **114,** 5928–5934
7. Parker, A. R., Moore, T. D., Edman, J. C., Schwab, J. M., and Davisson, V. J. (1994) *Gene* (*Amst.*) **145,** 135–138
8. Tada, S., Volrath, S., Guyer, D., Scheidegger, A., Ryals, J., Ohta, D., and Ward, E. (1994) *Plant Physiol.* **105,** 579–583
9. Mano, J., Hatano, M., Koizumi, S., Tada, S., Hashimoto, M., and Scheidegger, A. (1993) *Plant Physiol.* **103,** 733–739
10. Loper, J. C. (1961) *Proc. Natl. Acad. Sci. U. S. A.* **47,** 1440–1449
11. Chiariotti, L., Nappo, A. G., Carlomagno, M. S., and Bruni, C. B. (1986) *Mol. Gen. Genet.* **202,** 42–47
12. Houston, L. L. (1973) *J. Biol. Chem.* **248,** 4144–4149
13. Parker, A. (1995) *Medicinal Chemistry and Molecular Pharmacognosy*, p. 218, Purdue University, West Lafayette, IN
14. Tada, S., Hatano, M., Nakayama, Y., Volrath, S., Guyer, D., Ward, E., and Ohta, D. (1995) *Plant Physiol.* **109,** 153–159
15. Glaser, R. D., and Houston, L. L. (1974) *Biochemistry* **13,** 5145–5152
16. Wilkinson, K. W., Baker, P. J., Rice, D. W., Rodgers, H. F., and Stillman, T. J. (1995) *Acta Crystallogr. Sect. D* **51,** 845–847
17. Petersen, J., Hawkes, T. R., and Lowe, D. J. (2000) *J. Inorg. Biochem.* **80,** 161–168
18. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–325
19. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D* **55,** 849–861
20. Cowtan, K. D. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31,** 34–38
21. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
22. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D* **54,** 905–921
23. Read, R. J. (1986) *Acta Crystallogr. Sect. A* **42,** 140–149
24. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000) *Science* **289,** 1546–1550
25. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
26. Murzin, A. G. (1995) *Struct. Biol.* **2,** 25–26
27. Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L., and Zhang, H. (2000) *Structure* (*Lond.*) **8,** 1247–1257
28. Symmons, M. F., Jones, G. H., and Luisi, B. F. (2000) *Structure* (*Lond.*) **8,** 1215–1226
29. Schuck, P. (2000) *Biophys. J.* **78,** 1606–1619
30. Schuck, P., Perugini, M. S., Gonzales, N. R., Howlett, G. J., and Schubert, D. (2002) *Biophys. J.* **82,** 1096–1111
31. Lebowitz, J., Lewis, M. S., and Schuck, P. (2002) *Protein Sci.* **11,** 2067–2079
32. Wallace, A. C., Borkakoti, N., and Thorton, J. M. (1997) *Protein Sci.* **6,** 2308–2323
33. Kleywegt, G. J. (1999) *J. Mol. Biol.* **285,** 1887–1897
34. Brady, D. R., and Houston, L. L. (1973) *J. Biol. Chem.* **248,** 2588–2592
35. Vasington, F. D., and LeBeau, P. (1967) *Biochem. Biophys. Res. Commun.* **26,** 153–161
36. Staples, M. A., and Houston, L. L. (1979) *J. Biol. Chem.* **254,** 1395–1401
37. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
38. Merritt, E., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
39. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
40. Barton, G. J. (1993) *Protein Eng.* **6,** 37–40
41. Ban, C., Junop, M., and Yang, W. (1999) *Cell* **97,** 85–97
42. Romanowski. M. J., Bonanno, J. B., and Burley, S. K. (2002) *Proteins Struct. Funct. Genet.* **47,** 568–571
43. Aevarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., Al-Karadaghi, S., Svensson, L. A., and Liljas, A. (1994) *EMBO J.* **13,** 3669–3677
44. Fu, Z., Wang, M., Potter, D., Miziorko, H. M., and Kim, J. P. (2002) *J. Biol. Chem.* **277,** 18134–18142
45. Bonanno, J. B., Carme, E., Eswar, N., Pieper, U., Romanowski. M. J., Valentin, I., Gerchman, S. E., Kycia, H., Studier, F. W., Sali, A., and Burley, S. K. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 12896–12901
46. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature* **407,** 327–340
47. Stams, T., Niranjanakumari, S., Fierke, C. A., and Christianson, D. W. (1998) *Science* **280,** 752–755