Structure of the YibK Methyltransferase From *Haemophilus influenzae* (HI0766): A Cofactor Bound at a Site Formed by a Knot

Kap Lim,[1] Hong Zhang,[1] Aleksandra Tempczyk,[1] Wojciech Krajewski,[1] Nicklas Bonander,[1,2] John Toedt,[1] Andrew Howard,[3,4] Edward Eisenstein,[1,2] and Osnat Herzberg[1]*

[1]*Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland*
[2]*The National Institute of Standards and Technology, Gaithersburg, Maryland*
[3]*Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois*
[4]*Biological, Chemical, and Physical Science Department, Illinois Institute of Technology, Chicago, Illinois*

ABSTRACT **The crystal structures of YibK from *Haemophilus influenzae* (HI0766) have been determined with and without bound cofactor product S-adenosylhomocysteine (AdoHcy) at 1.7 and 2.0 Å resolution, respectively. The molecule adopts an α/β fold, with a topology that differs from that of the classical methyltransferases. Most notably, HI0766 contains a striking knot that forms the binding crevice for the cofactor. The knot formation is correlated with an alternative arrangement of the secondary structure units compared with the classical methyltransferases. Two loop regions undergo conformational changes upon AdoHcy binding. In contrast to the extended conformation of the cofactor seen in the classical methyltransferase structures, AdoHcy binds to HI0766 in a bent conformation. HI0766 and its close sequence relatives are all shorter versions of the more remotely related rRNA/tRNA methyltransferases of the spoU sequence family. We propose that the spoU sequence family contains the same core domain for cofactor binding as HI0766 but has an additional domain for substrate binding. The substrate-binding domain is absent in HI0766 sequence family and may be provided by another *Haemophilus influenzae* partner protein, which is yet to be identified. Proteins 2003;51:56–67.** © 2003 Wiley-Liss, Inc.

INTRODUCTION

Methyltransferases that require S-adenosylmethionine (AdoMet) as a cofactor perform widely varied biological roles, including methylation of small molecules, nucleic acids, and proteins.[1] The cofactor product after methyl group transfer to the substrate is S-adenosylhomocysteine (AdoHcy), a potent inhibitor of most methyltransferases. The relative levels of cellular AdoMet and AdoHcy regulate the activity of methyltransferases.[2,3] AdoMet is synthesized from L-methionine and ATP by methionine adenosyltransferase.[4,5] After transmethylation by a methyltransferase, AdoHcy is converted to adenosine and homocysteine by AdoHcy hydrolase for further use of these degradation products. AdoMet is unstable in vitro both in solution and in the crystalline form and degrades into smaller molecules.[6,7] Yet, examples of both AdoMet and AdoHcy bound to crystal structures are available in the Protein Data Bank (PDB).[8]

The most common structural motif of AdoMet-dependent methyltransferases is an α/β fold containing a seven-stranded β-sheet, where all but one β-strand are parallel.[9] In these "classical" methyltransferases, the cofactor binds in an extended conformation at the C-terminus of β1-strand where a conserved glutamic acid or aspartic acid residue is positioned to interact with the ribose ring of the cofactor. The close structural similarity of this enzyme family is not manifested in their amino acid sequences, and only a weak pattern of conservatively replaced residues corresponding to the cofactor binding site is observed. Two variations on the classical fold have also been reported: In isoaspartyl methyltransferase, β-strands 6 and 7 reverse positions,[10] and in arginine methyltransferase, β-strands 6 and 7 are absent.[11,12]

AdoMet is also used by other methyltransferases that exhibit no sequence homology to the classical methyltransferases and have different folds. For example, the cobalamin biosynthesis enzyme is a two-domain protein, each with an α/β fold, in which the substrate and cofactor AdoHcy bind between the two domains.[13] Other sequence families of AdoMet-dependent methyltransferases of unknown structures may or may not adopt an already known methyltransferase fold. For example, the spoU-related sequence family is expected to constitute a new class of methyltransferases different from the classical methyltransferases.[14] SpoU, a protein associated with starvation response in *Escherichia coli*, was first suggested to be an rRNA methyltransferase based on sequence similarity with two known rRNA methyltransferases.[15] Later, the

Grant sponsor: National Institutes of Health; Grant number: PO1 GM57890.

H. Zhang's present address is Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas.

J. Toedt's present address is Department of Physical Science, Eastern Connecticut State University, Willimantic, Connecticut.

*Correspondence to: Osnat Herzberg, Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: osnat@carb.nist.gov

Received 31 May 2002; Accepted 7 October 2002

spoU protein was shown to be a tRNA guanosine(Gm)18 2′-O-methyltransferase.[16]

YibK of *Haemophilus influenaze* (HI0766)[17] is homologous to the spoU family of AdoMet-dependent tRNA/rRNA methyltransferases. At the time of writing, sequence analysis using PSI-BLAST[18] shows some 50 close relatives of HI0766 in curated sequence databases. HI0766 and its close sequence relatives have a shorter polypeptide chain (~160 amino acids) than the spoU enzyme family (230 amino acids). The HI0766 family and the homologous domain of the spoU family share 20–30% sequence identity, indicating that their structure is similar. The crystal structure of HI0766 is reported here in the free and AdoHcy-bound states, revealing a novel fold and an unusual cofactor binding mode, which are likely to be the hallmark of the spoU family as well.

MATERIALS AND METHODS

Protein Production

The gene for *Haemophilus influenzae* Rd KW20 open reading frame HI0766 was amplified from chromosomal DNA, treated with *Nde*I and *Hind*III, and cloned into a similarly digested pET17b plasmid for expression of the native polypeptide from the T7 promoter in *E. coli* strain BL21 DE3. Cells were grown at 37°C in LB medium containing 100 μg>/ml ampicillin until they reached mid-log phase, at which time expression was induced by the addition of 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) with continued incubation at 37°C for 4 h. Cells were lysed by using a French press at 10,000 psi and the soluble fraction recovered by high-speed centrifugation. The crude cell extract was dialyzed against 20 mM Tris-HCl, pH 8.4, and purified by using perfusion chromatography using a BioCAD 700E chromatography workstation (PerSeptive Biosystems) with a series of ion exchange chromatography steps. The extract was first applied to a Poros HQ 50 anion exchange column and eluted with a linear gradient of 0–1 M NaCl over 30 column volumes. The fractions enriched for HI0766 as assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were pooled, dialyzed against buffer containing a blend of 20 mM MES, 20 mM HEPES, and 20 mM sodium acetate, to yield a pH of 6.0, and applied to a Poros HS 20 cation exchange column. The column was eluted with a linear gradient of 0–1 M NaCl over 20 column volumes, and the fractions containing purified protein were pooled and dialyzed against 50 mM Tris-HCl, pH 7.5, containing 0.1 mM DTT and 0.1 mM EDTA for characterization and crystallization. The molecular weight of HI0766 was verified by MALDI mass spectrometry (Voyager, PerSeptive Biosystems) (MW = 18,400.8 observed and 18,401.5 calculated). Ultracentrifugation evaluation indicated that HI0766 exists predominantly as a dimer.

Crystallization and Data Collection

Crystals were obtained by the vapor diffusion method in hanging drops at room temperature. A 12 mg/mL protein solution in 50 mM Tris-HCl at pH 7.5, 0.1 mM DTT, and 0.1 mM EDTA, was mixed with an equal volume of mother liquor containing 20% polyethylene glycol monomethyl ether (PMME) 2000, 0.1 M sodium acetate at pH 4.6, 0.2 M ammonium acetate, and 3% ethylene glycol, and equilibrated against the mother liquor reservoir. The reproducibility of crystals was poor, and the protein tended to precipitate. Reproducible crystals were obtained at a later stage of structure determination in 1.4 M sodium malonate and 0.1 M sodium acetate at pH 4.6. In both conditions, crystals appeared after 1 week and grew to dimensions of $\sim 0.15 \times 0.15 \times 0.15$ mm^3. Cocrystallization of HI0766 with AdoMet (iodine salt; Sigma Chemicals, St. Louis, MO) was unsuccessful. Hence, the HI0766 structure with bound cofactor was obtained by soaking a crystal with 20 mM AdoMet solution for 3 days.

For diffraction data collection, crystals were flash-cooled at 100 K directly from the mother liquor without additional cryogenic solution. Data were collected by using both a home laboratory X-ray source and the IMCA-CAT 17-ID beamline at the Advanced Photon Source (APS; Argonne National Laboratory, Argonne, IL). The home X-ray source was a Siemens rotating anode equipped with MAR345 image plate. The IMCA-CAT beamline was equipped with a MAR CCD detector. Data processing was performed by using the HKL program suite.[19] Table I provides the statistics of data processing.

Structure Determination

Two crystal structures of HI0766 were determined, with and without bound AdoHcy. In both cases, the crystals belonged to space group $P4_12_12$ with virtually identical lattice parameters (Table II). The asymmetric unit contains one molecule and 35% solvent. Crystals grown in PMME 2000 and sodium malonate resulted in the same lattice parameters and showed similar diffraction qualities. The apo-form structure was first determined by the multiple isomorphous replacement (MIR) method using Hg, Pt, and Xe heavy atom derivative data (Table I). The Xe derivative was prepared by placing a crystal in a pressure chamber (Hampton Research, Laguna Niguel, CA) filled with Xe gas at 400 psi for 30 min. The computer program SOLVE[20] was used to determine the heavy atom sites. SOLVE identified two sites for Hg, two for Pt, and one for Xe.

Phase determination was conducted by using the CCP4 suite of programs.[21] The program MLPHARE[22] was used to refine the positions and occupancies of the heavy atoms and to calculate phases at 2.3 Å resolution. The phases were further improved at 2.3 Å by solvent-flattening using the program DM.[23] The resulting electron density map was interpretable, and the protein model was built on a Silicon Graphics Octane workstation using the interactive computer graphics program "O."[24] The two Hg derivatives had the highest phasing power, and two Hg positions presumed to bind to cysteine residues were used as markers in tracing of the polypeptide chain.

Refinement of the apo-form of HI0766 structure was conducted by using CNS,[25] with the data between 20.0 and 2.0 Å for which $F \geq 2\sigma(F)$. A simulated annealing molecular dynamics cycle at 4000 K was followed by alternating

TABLE I. Data Collection and Phasing Statistics

Space group	$p4_12_12$					
Cell dimension (Å)	$a = b = 40.4, c = 165.8$		No. of molecules in the asymmetric unit			1
Crystals	apo-HI0766	HI0766+AdoHcy	Hg acetate	$(Ethyl\ Hg)_3PO_4$	K_2PtCl_4	Xe
Concentration	—	—	1 mM	1 mM	2 mM	400 psi
Length of soak (hours)	—	—	24	24	48	0.5
Wavelength (Å)[a]	1.00	1.00	1.54	1.00	1.07	1.54
Resolution (Å)	2.0	1.7	2.5	2.3	2.7	2.5
No. of observed reflections	113,016	281,315	10,291	110,949	8,511	35,907
No. of unique reflections	9,754	15,394	5,955	6,559	2,599	5,296
Completeness (%)[b]	98.1 (99.8)	93.7 (69.2)	80.2 (79.8)	98.6 (99.2)	82.3 (83.8)	99.1 (99.1)
R_{merge} (%)[c]	5.8 (13.4)	4.0 (22.3)	8.5 (26.7)	9.4 (15.5)	5.8 (37.3)	8.9 (35.4)
$<I/\sigma(I)>$	14.0	18.5	10.0	16.0	5.8	10.8
Phasing statistics[d]						
Phasing power	—	—	1.88	1.77	0.80	0.63
Dispersive R	—	—	0.67	0.71	0.91	0.92
Anomalous R	—	—	0.94	0.79	0.96	0.99
Figure of merit	0.60					

[a]Data at 1.54 Å wavelength were collected on the home X-ray facility, and data at other wavelengths were collected at IMCA-CAT beamline.
[b]The values in parentheses are for the highest resolution shell (e.g., 2.09–2.00 Å for the apo-HI0766 data and 1.78–1.70 Å for the HI0766+AdoHcy data).
[c]$R_{merge} = \sum_{hkl}[(\sum_j|I_j - \langle I\rangle|)/\sum_j|I_j|]$.
[d]Phasing power = $\sum_j|F_H(calc)|/\sum_j|E|$, where $F_H(calc)$ is calculated by scattering amplitude of the heavy atom structure and E is the lack of closure error, the difference between observed and calculated derivative data. Dispersive $R = \sum_{hkl}|F_P + F_H(calc)| - F_{PH}|/\sum_{hkl}|F_{PH} - F_P|$, where F_P corresponds to the native data set, and F_{PH} correspond to derivative data. Anomalous $R = \sum_{hkl}|\Delta F^{\pm}_{obs} - \Delta F^{\pm}_{calc}|/\sum_{hkl}|\Delta F^{\pm}_{obs}|$, where $\Delta F^{\pm}$ is the structure factor difference between Friedel pairs.

cycles of positional and individual temperature factor refinement. The models were inspected and modified by using the "O" program. In the final stages of the refinement, water molecules were added in accordance with the F_o-F_c difference Fourier electron density map (where F_o and F_c are the observed and calculated structure factors, respectively), using peaks with density $\geq 3\sigma$ as the acceptance criteria.

The apo-form structure was used as a starting model for the structure refinement of the cofactor-bound enzyme. Rigid body minimization was followed by positional and individual temperature factor refinement with the data in the 20.0–1.7 Å range. The difference Fourier map with the coefficients $F_{o(bound)}$-$F_{o(unbound)}$ (where $F_{o(bound)}$ is the structure factor of HI0766 with cofactor and $F_{o(unbound)}$ the structure factor of apo-form of HI0766) revealed the presence of the cofactor lacking a methyl group attached to sulfur, and hence it was modeled as AdoHcy (Fig. 1). The energy and geometry parameters of the cofactor were obtained from the HIC-UP database (http://xray.bmc.uu.se/hicup).[26] All AdoCys atoms were included in the refinement. The cofactor exhibited unusual ribose geometry, and hence the target dihedral angle parameters were modified to allow a change of conformation. In addition to AdoHcy, an iodine ion that presumably originated from the iodine salt of AdoMet was identified. Difference anomalous Patterson and difference Fourier maps confirmed that the site is occupied by iodide.

Structure analysis was conducted by using a set of computer programs: PROCHECK for analysis of geometry,[27] XROT for superposition of structures (Kap Lim, unpublished), MS for solvent-accessible surface area calcu-

TABLE II. Refinement Statistics

	Apo-HI0766	HI0766+AdoHcy
Cell dimension (Å)	$a = b = 40.4$, $c = 165.8$	$a = b = 40.8$, $c = 165.8$
Resolution (Å)	20.0–2.0	20.0–1.7
Unique reflections $F \geq 2\sigma(F)$	9,326	14,742
Completeness (%)[a]	93.2 (89.8)	90.3 (56.3)
No. of protein atoms	1261	1261
No. of H_2O molecules	124	153
No. of other atoms	—	27
R_{cryst} (%)[b]	19.5 (23.4)	19.6 (21.5)
R_{free} (%)[c]	26.3 (32.3)	25.4 (27.3)
RMSD from ideal geometry		
Bond length (Å)	0.018	0.017
Bond angle (Å)	1.9	1.8
Average B factor (Å^2)		
Protein	29	28
H_2O	37	38
AdoHcy	—	32
Iodide	—	32
Ramachandran plot (%)		
Most favored	91.2	87.7
Allowed	7.4	11.6
Generously allowed	1.5	0.7
Disallowed	0.0	0.0

[a]The values in parentheses are for the highest resolution shell.
[b]$R_{cryst} = \sum_{hkl}|\,|F_o| - |F_c|\,| / \sum_{hkl}|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.
[c]R_{free} is computed from reflections that were randomly selected and omitted from the refinement, 1036 for the apo-HI0766 data and 1593 for the HI0766+AdoHcy data.

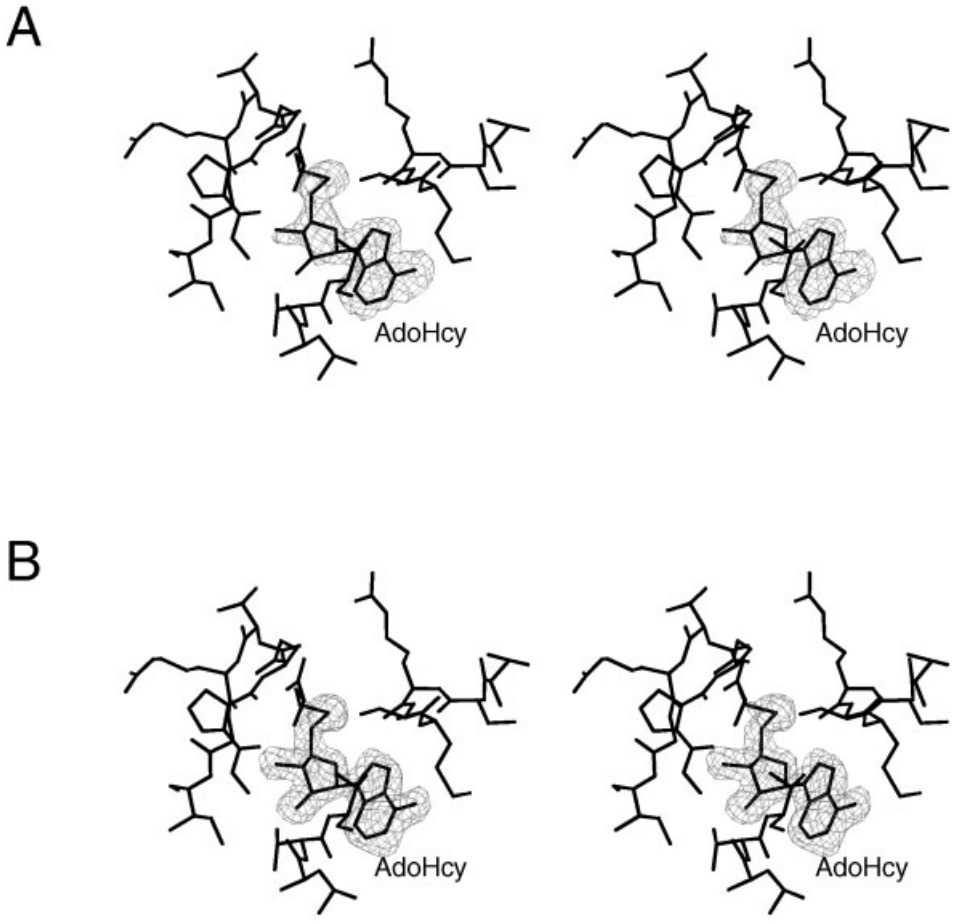

Fig. 1. Stereoscopic view of the electron density maps of HI0766 associated with the cofactor, S-adenosylhomocysteine (AdoHcy). The bound AdoHcy model and surrounding amino acid residues are shown. **A:** Difference Fourier map with the coefficients ($F_{o(bound)} - F_{o(unbound)}$), where $F_{o(bound)}$ is the structure factor of HI0766 with AdoHcy and $F_{o(unbound)}$ is structure factor of HI0766 without AdoHcy. The phases were calculated before the cofactor was added to the model. The map is contoured at 2σ level. **B:** Difference map with the coefficients ($2F_o - F_c$) and calculated phases, sigma(a)-weighted, at the end of the refinement. The map is contoured at 1σ level. For clarity, only the cofactor density is shown.

lations,[28] and Molscript[29] and Raster3D[30,31] for depiction of structure.

RESULTS AND DISCUSSION

Overall Structure

The HI0766 crystal contains one molecule in the asymmetric unit, and a tightly packed dimer is formed through a crystallographic twofold symmetry axis along the diagonal in the *ab* crystal plane (Fig. 2). The dimer formation is consistent with the ultracentrifuge result. The monomer dimensions are $\sim 30 \times 35 \times 40$ Å^3, and the dimer's longest dimension is 68 Å. The model includes 156 of the 160 amino acid residues. The electron density map indicates that the four C-terminal residues are disordered. The cofactor-free model includes 124 water molecules, and that with bound AdoHcy includes 153 molecules as well as an iodine ion. Table II summarizes the refinement statistics.

HI0766 is a single-domain protein that adopts an α/β fold with all-parallel β-sheet [Fig. 3(A)]. Currently, structure comparison using DALI[32] yields only one structural relative deposited recently in the PDB, that of MT0001 from *Methanobacterium thermoautotropicum*, a protein of unknown function (PDB entry code 1k3r; Zarembinski et al., unpublished). The Z-score obtained by DALI is 11.5, and 133 of the 262 residues align with HI0766 with a root-mean-square deviation (RMSD) in C^α atom positions of 2.8 Å, and 14% sequence identity of the aligned residues. The β1/α1/ β2/α2/β3 topology of HI0766 is the same as that of classical AdoMet-dependent methyltransferases [Fig. 3(B)] (i.e., the first β-strand is located in the center of the β-sheet, and the winding of the polypeptide chain is right-handed). The polypeptide chain then crosses over three β-strands to the other side of the β-sheet, and this is where the fold of HI0766 deviates from that of the classical methyltransferases and forms a knot.

A knotted polypeptide chain is formed by skipping three β-strands [Figs. 2(B) and 4(A)] instead of two β-strands as in the classical methyltransferases [compare Fig. 3(A) and (B)]. That is, the sheet spacing between β3 and β4 in the classical fold is occupied by β2 and β1 [Fig. 3(B)], whereas

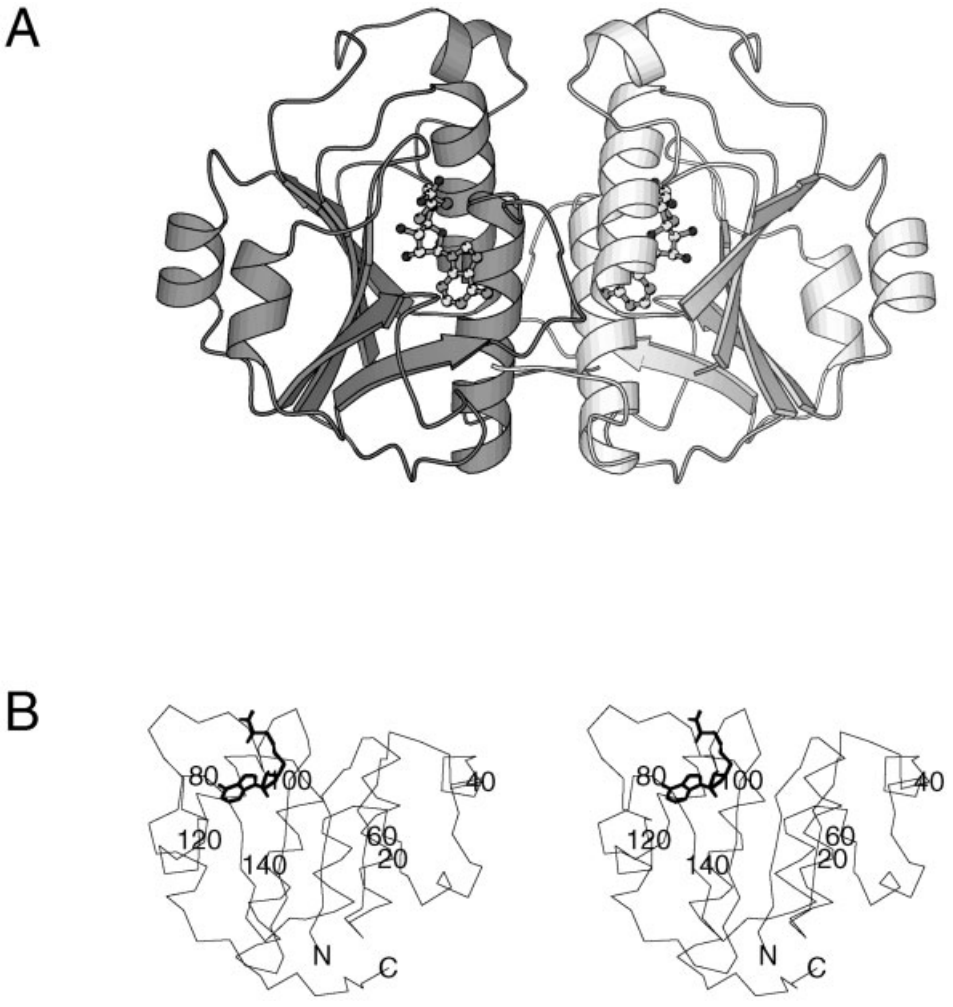

Fig. 2. Crystal structure of HI0766. **A:** Ribbon diagram of a dimer with the two monomers shown in dark and light shadings. The cofactor AdoHcy is shown as a ball-and-stick model. **B:** Stereoscopic view of the monomer C^{α} atom trace. Every 20th residue is numbered.

in HI0766 the positions of β4 and β5 are switched such that three β-strands (β2, β1, and β5) bridge the space between β3 and β4 [Fig. 3(A)]. Consequently, the polypeptide chain winds left-handed from β4 to β5. The right-handed winding of chain is then recovered by forming a knot that enables positioning of β6 on the other side of β4. The knot peptide segments 83–86 (loop between β4 and β5) and 120–122 (β6) run perpendicularly and contact one another. Hydrophobic and polar interactions between these segments and other surrounding residues support the integrity of the knot. Following β6, the C-terminus polypeptide chain forms an extended loop followed by an α-helix that interacts with the second molecule of the dimer.

The HI0766 fold lacks an antiparallel, seventh β-strand characteristic of the classical methyltransferase fold. The superposition of HI0766 and the core domain of a classical methyltransferase of catechol O-methyltransferase (PDB entry code 1vid)[33] shows that [Fig. 3(C)]: (i) the β1/α1/β2/α2/β3/α3 segment overlaps quite well; (ii) although the β4 and β5 positions have been switched, they occupy similar space, along with the respective α-helices that follows either from β5 of HI0766 or β4 of classical fold; (iii) β6 of HI0766 traverses the β6/β7 positions of the classical methyltransferase; and (iv) the C-terminal helix, α5, of HI0766 overlaps with the N-terminal helix of the catechol O-methyltransferase.

HI0766 associates into dimers both in solution and in the crystal structure. The two monomers are closely packed, with the N-terminus of two symmetry-related α-helices, α5, facing one another at ~40°. The α1 helix also participates in dimer formation. Each molecule contributes 900 Å^2 to the dimer interface, or ~12% of the total surface of a monomer. The interface residues are invariant within the HI0766 family members (Fig. 5). In particular, an extended network of hydrogen bonds is formed by symmetry-related asparagine, serine, and glutamic acid residues at the N-terminus of α5 (Fig. 6), and the ion pair between Glu143 and Arg146 at the C-terminus of α5 (not shown) projects toward the counterpart ion pair in the other monomer. At the center of the dimer interface, the

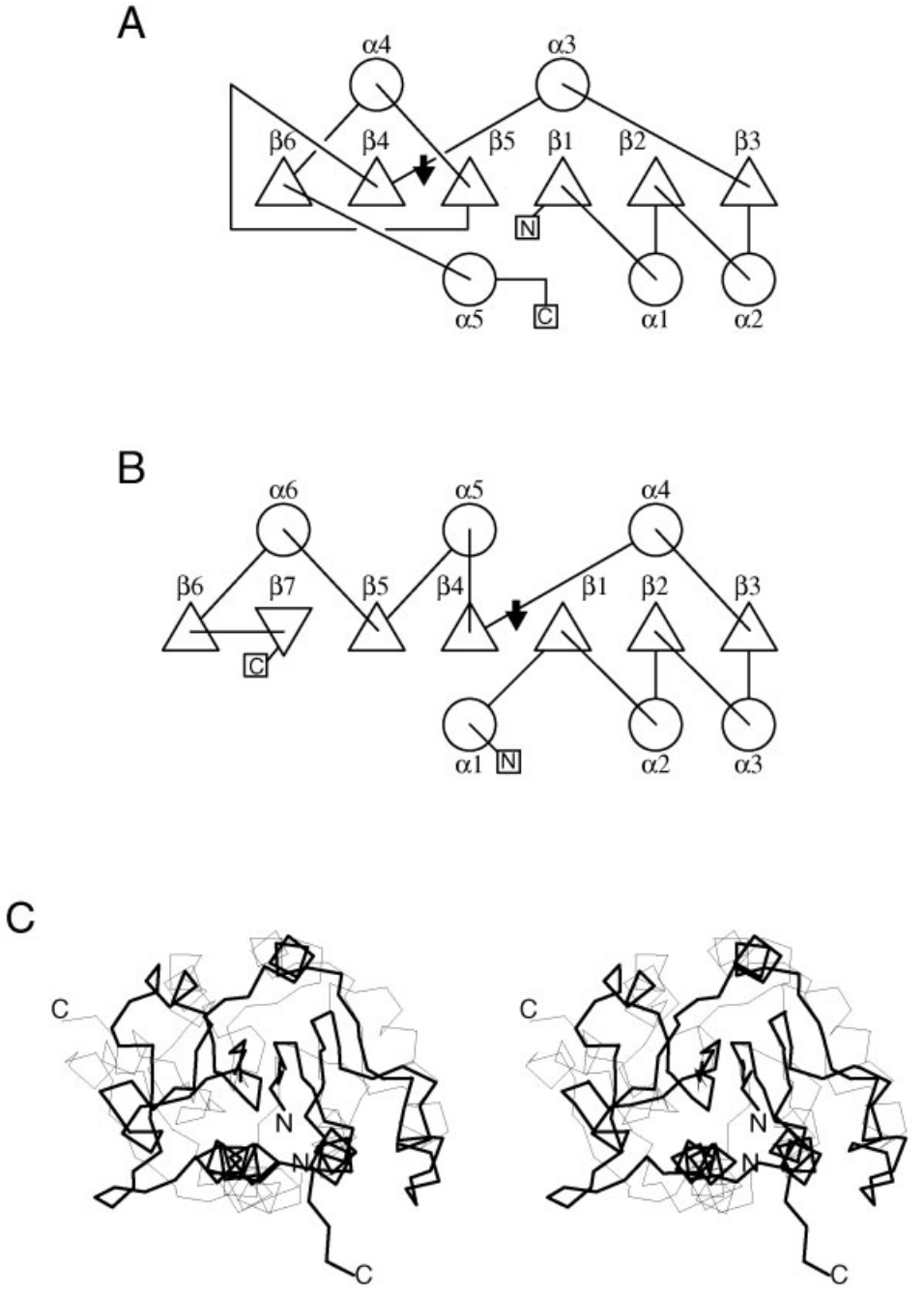

Fig. 3. Topological diagrams of the secondary structure elements of (**A**) HI0766 and (**B**) The core domain of a classical methyltransferase (catechol O-methyltransferase, PDB code 1vid).[33] An arrow in (A) and (B) indicates the topological location of the cofactor-binding site. Triangles represent β-strands and circles represent α-helices. **C:** Superposition of the first three β-strands of HI0766 (thick line tracing) and catechol O-methyltransferase (thin line tracing). Not shown is the additional N-terminal domain of catechol O-methyltransferase.

invariant residues Leu21, Val139, and Tyr142 comprise the hydrophobic component of the interface. Notably, the structure of MT0001 (1k3r) shows that it also forms a dimer in a similar manner to HI0766. This observation supports the notion that all members of the HI0766/spoU superfamily are likely to function as dimers, with dimer interfaces similar to that of HI0766.

The iodine ion seen in the AdoHcy-bound structure is located in the dimer interface close to the N-terminus of α5 and 9 Å away from the AdoHcy. It interacts with a buried water molecule, Asn135 and Asn24 of each monomer (Fig. 6). No positively charged residues are present to counter the anion charge; however, this is not unprecedented.[34,35]

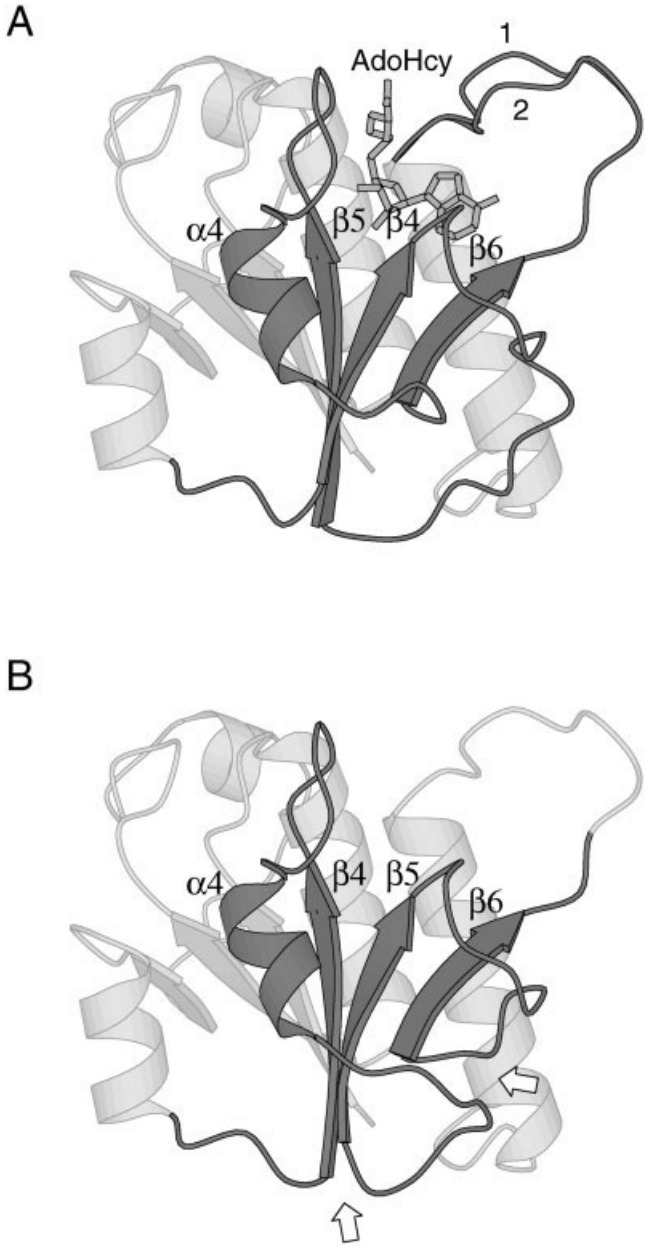

Fig. 4. The knot of HI0766. **A:** The knot region is highlighted in black and the rest of the molecule is shaded gray. Also shown is the alternate conformation of the loop 127–131 at the unbound (1) and cofactor-bound (2) states. **B:** A model that resolves the knot by shuffling polypeptide connections. Open arrows indicate where the chain breaks are made.

Cofactor Interactions and Protein Conformational Change

The overall structure of HI0766 remains the same on cofactor binding. Excluding five loop residues that undergo conformational transitions (discussed below), the RMSD in C^{α} atom positions between the bound and unbound structures is 0.3 Å. The AdoHcy cofactor binds in a topologically different location in HI0766 compared with the classical AdoMet-dependent methyltransferases (Fig. 3). In HI0766, the cofactor binds in a pocket formed by the two loops of the knot, residues 80–85 and 102–105, and the 123–133 loop connecting β6 and α5 [Figs. 2(B) and 4(A)]. There is no access from the binding pocket to the dimer interface, and the distance between the two AdoHcy molecules of the dimer is ~30 Å. The hydrogen bond network at the dimer interface extends to Glu102 and Asn132, residues involved in the AdoHcy binding (Fig. 6). This network is suggestive of possible allosteric modulation of cofactor binding.

The electron density for the adenine and ribose rings of AdoHcy is well defined, and for the homocysteine group, clear electron density is present only up to the sulfur atom (Fig. 1). The remaining atoms are exposed to the solvent and are associated with electron density lower than 0.7σ. The observed orientation of AdoHcy implies that the donor methyl group of the intact AdoMet cofactor is exposed to solvent and thus accessible to the substrate. The adenine ring of AdoHcy is surrounded by primarily hydrophobic side-chains: Leu78, Thr80, Ile122, Met124, and Met131. Binding of AdoHcy buries 338 $Å^2$ of the HI0766 surface, and many of the buried residues are conserved (Fig. 5). The interactions of the adenine and ribose moieties are largely with main-chain amide and carbonyl groups (Fig. 7). For the ribose moiety, the O2′ atom is hydrogen-bonded to the carbonyl oxygen of Leu78 and the amide nitrogen of Gly100. O3′ is hydrogen-bonded to the amide nitrogen of Gly100 and a water molecule. The O4′ atom is hydrogen-bonded to O^{γ} of Ser136. The N1 atom of the adenine moiety is hydrogen-bonded to the amide nitrogen of Ile122, and N6 is hydrogen-bonded to the carbonyl oxygen of Ile122. In addition, the N7 atom is hydrogen-bonded to the amide nitrogen of Met131.

Two regions of HI0766 undergo conformational changes on AdoHcy binding. Residues 127–131 on the β6/α5 loop move to interact with AdoHcy [Fig. 4(A)], displacing the C^{α} atoms of Arg129 and Ser130 by 3.4 Å and 2.4 Å, respectively. Residual electron density, which corresponds to the conformation of the loop in the unbound state is also present in the map, indicating that the cofactor may not be bound to all the molecules of the crystal. In unbound state, the loop is stabilized in part by the interaction between the carbonyl oxygen of Ser130 and the N^{η} atom of the invariant Arg20 of the second molecule. When the loop moves toward AdoHcy, the Ser130 carbonyl interacts with the N^{δ} atom of Asn24.

The second conformational change is subtler than that of the 127–131 segment. The peptide bond between Arg104 and Gly105 flips on cofactor binding (Fig. 8). This change was verified by omit-map-simulated annealing refinement and is reliable given the data resolution. The carbonyl oxygen of Arg104 is oriented toward the carbonyl group of Pro101 in apo-HI0766, an unusual interaction because it is electrostatically unfavorable. The peptide flip relieves electrostatic strain, and the associated free energy change may be used to enhance cofactor-binding affinity. Gly105 is invariant in the HI0766 sequence family (Fig. 5). A glycine residue at this position may be required to enable conformational flexibility, because the torsion angles at the

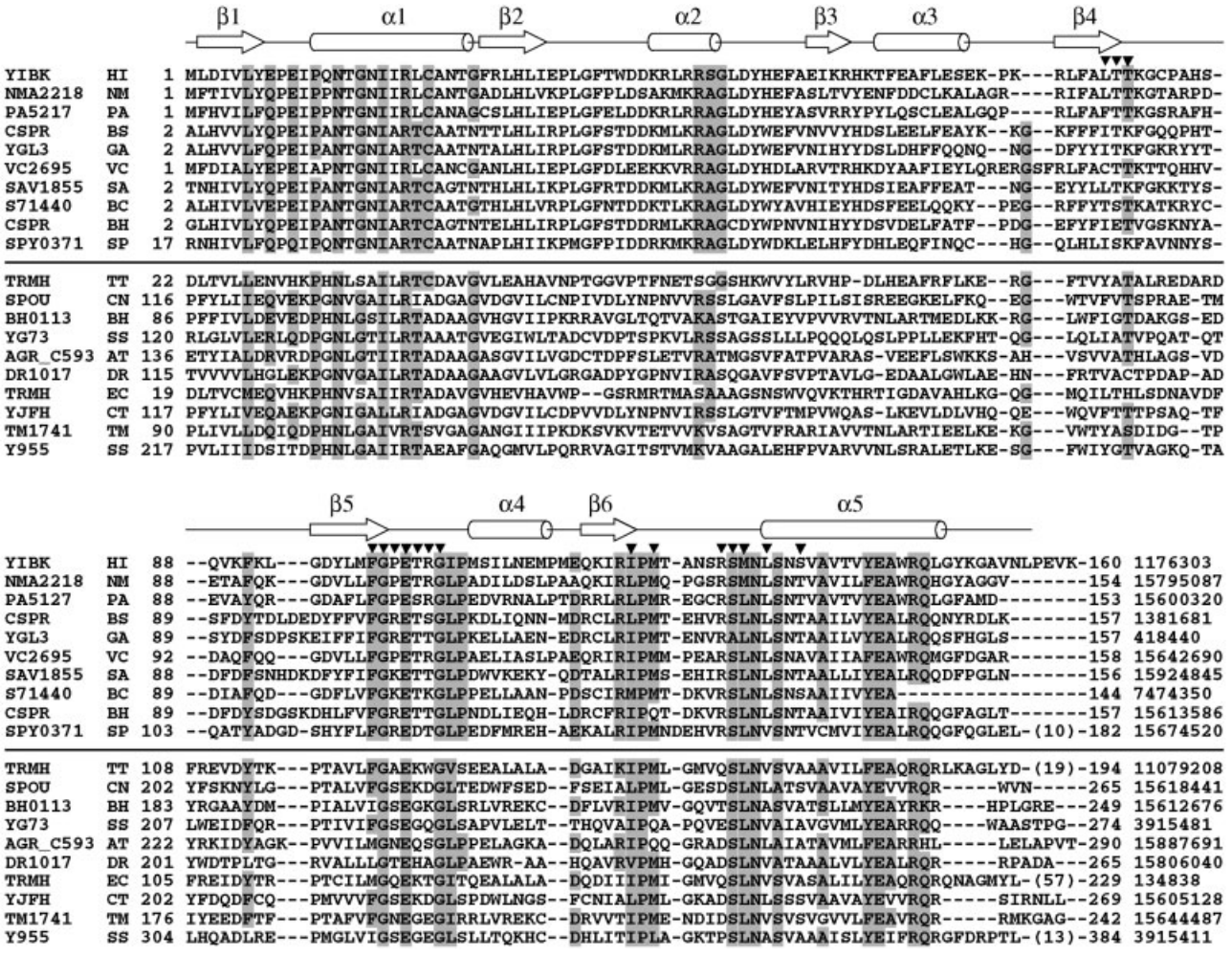

Fig. 5. Multiple alignment of HI0766 with sequences obtained from the first and second iteration cycles of a PSI-BLAST search.[18] The results from the first and second cycles are separated by a line. The alignment was conducted by using ClustalW,[44] applying 80% identity cutoff to the nonredundant protein database. Only the first 10 sequences from the first iteration cycle and the first 10 sequences from the second cycle are shown. The first column lists the protein (gene) name and the second column abbreviates species name: HI, *Haemophilus influenzae*; NM, *Neisseria meningitidis*; PA, *Pseudomonas aeruginosa*; BS, *Bacillus subtilis*; GA, *Geobacillus stearothermophilus*; VC, *Vibrio cholerae*; SA, *Staphylococcus aureus*; BC, *Bacillus circulans*; BH, *Bacillus halodurans*; SP, *Streptococcus pyogenes*; TT, *Thermus thermophilus*; CN, *Chlamydophila pneumoniae*; SS, *Synechocystis* sp.; AT, *Agrobacterium tumefaciens*; DR, *Deinococcus radiodurans*; EC, *Escherichia coli*; CT, *Chlamydia trachomatis*; and TM, *Thermotoga maritima*. The numbers in the third column indicate the positions of the first residue of the aligned region in the respective protein sequence. Numbers in parentheses indicate how many residues in an insertion are not shown. The number at the end of each sequence indicates the last residue number. The NCBI gene identification (gi) is listed at the end of the alignment in the second panel. The secondary structural elements of HI0766 (YibK_*HAEIN*) are shown as arrows (β-strands) and cylinders (α-helices). Below them is the sequence numbering for HI0766. Most similar residues from all sequences are shaded by the Blosum62 score of 1.5 (from the Belvu program, written by Erik Sonnhammer). Residues buried by the AdoHcy binding in HI0766 are indicated by dark triangles above the first sequence line.

cofactor-bound state lead to steric strain for C^{β}-containing amino acids ($\phi,\psi = 73°,168°$).

Because the structure of the HI0766-AdoHcy complex was obtained by soaking the crystal in AdoHcy, the possibility of crystal-packing constraints affecting the physiologically relevant binding mode must be considered. Inspection of the crystal packing reveals that (i) the loops that form the binding pocket are exposed to solvent and are not involved in crystal packing and (ii) the β6/α5 loop exhibits flexibility within the packing constraints and changes its conformation on binding to improve shape and electrostatic complementarity between enzyme and cofactor.

AdoHcy Conformation

The following discussion refers only to atoms and torsion angles of the AdoHcy that are well defined in the electron density map.

AdoHcy binds to HI0766 in a different conformation compared with those observed in other AdoMet-dependent methyltransferase structures. The ribose ring of AdoHcy adopts a C3′-endo conformation (as in A-DNA), whereas in most of the other methyltransferases, the C2′-endo conformation is observed (Fig. 9). In addition, AdoHcy assumes a bent structure with the homocysteine moiety positioned perpendicular to the adenine ring. This is in contrast to the

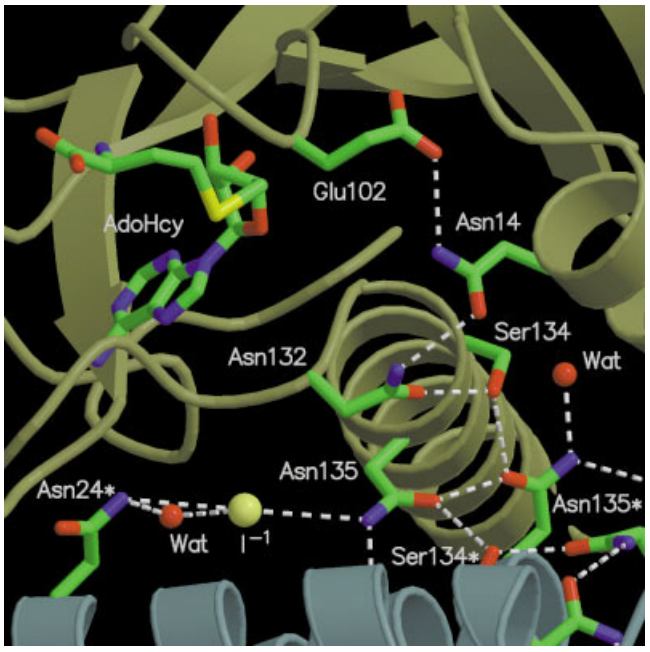

Fig. 6. Dimer interface of HI0766. Shown are the network of hydrogen bonds between the two monomers leading to the amino acid residues near the cofactor (AdoHcy) binding site. The residues labeled with asterisks are from the second monomer. Also shown are buried water molecules and an iodine ion.

extended conformation observed in other methyltransferases.

The C2′-endo and C3′-endo interconversion of ribose rings is rapid, and the former is slightly preferred in purine ribonucleosides.[36] The pseudorotation angle, computed from the five torsion angles of the ribose ring, is in the range of 0–36° for the C3′-endo form and 144–180° for the C2′-endo form.[36] For AdoHcy in HI0766, this angle is 17°. An additional torsion angle, O4′-C1′-N9-C4, defines the orientation of the adenine ring relative to the ribose ring.[37] This angle is −177° for AdoHcy bound to HI0766, whereas the value is approximately −120° in the classical methyltransferase structures. Finally, the O4′-C4′-C5′-S^{δ} torsion angle, defining the orientation of the methionine or homocysteine moiety relative to the ribose and adenine rings, is −63° for AdoHcy in HI0766, whereas this value is close to 180° in classical methyltransferases. This is the essential torsion angle that results in the bent conformation of AdoHcy in HI0766 and extended conformation in the others.

Although rare, the combination of ribose conformation and O4′-C1′-N9-C4 torsion angle exhibited in AdoHcy bound to HI0766 is not unprecedented. Two other structures in the PDB contain cofactors with the same ribose pseudorotation and O4′-C1′-N9-C4 torsion angle, cobalprecorrin-4 methyltransferase (CPMase) with bound AdoHcy, determined at 2.4 Å resolution (PDB code 1cbf),[13] and the C-terminal domain of methionine synthase (MetS) with bound AdoMet, determined at 1.8 Å resolution (PDB code 1msk).[38] The ribose pseudorotation angles in these structures are 10° and 5°, respectively, and the O4′-C1′-N9-C4 torsion angles are −174° and −167°. The folds of these proteins are different from that of the classical methyltransferases. In CMPase, the AdoHcy cofactor is bound between two five-stranded α/β domains. In MetS, the AdoMet cofactor is bound near the center of the inner surface of the protein with no obvious divisions into subdomains. The ribose and adenine groups of AdoHcy observed in the CPMase and HI0766 structures superimpose very well, with an RMSD of 0.1 Å. For MetS and HI0766, the value is 0.3 Å. The differences are manifested in the O4′-C4′-C5′-S^{δ} torsion angle, which is 80° for the cofactors in CPMase and MetS and −63° in HI0766. Furthermore, the cofactor-protein interactions do not share common features. In MetS, each side of the adenine ring of AdoMet stacks against a tyrosine residue, whereas in HI0766 and CPMase, different combinations of aliphatic side-chains and main-chain atoms surround the adenine ring.

The Knot

Knots are extremely rare in known protein structures. When they are observed, they are primarily located at the C-termini of polypeptide chains, and only a few occur in the middle of the chain.[39] Deeply knotted structures, such as HI0766, pose a folding problem. Mechanisms that give rise to a knot may include domain swapping and exchanging of secondary structure elements between duplicated domains,[39] or the movement of the N-terminal secondary structural elements into a knot during protein synthesis.[40]

There are several ways to resolve the tangled chain of HI0766, but only one maintains the core packing of the molecule and the cofactor binding-site architecture. Cutting and shuffling the structure at residues 73 (N-terminus of β4) and 94 (N-terminus of β5), and similarly, at residues 89 (on the β4/β5 loop) and 117 (on the α4/β6 loop) switches the order of β4 and β5, eliminates the knot, and recovers the right-handed winding of the polypeptide chain [Fig. 4(B)]. The rearrangement does not require major repacking of the core and does not impact the cofactor-binding site or dimer association. It also provides for a long loop that can traverse the β-sheet and connect the new β5 with β6.

The resulting fold is a classical all-parallel doubly-wound α/β structure lacking an antiparallel seventh β-stand, the hallmark of the classical methyltransferase fold. In addition, the cofactor-binding site remains in a different topological and spatial location from that of the classical methyltransferases. Transition to the classical fold requires that, in addition to the shuffling described above, major rearrangements of secondary structure units must take place. (i) The β6/α5 loop (residues 122–133) should displace β6 and insert between the new β5 and the shifted β6 to form the antiparallel β7. This is reminiscent of the serpin proteins, which act by inserting a cleaved loop into the β-sheet of the inhibited protease.[41] (ii) Either the direction of α5 should be changed to run antiparallel to α1, resulting in extensive repacking of the core, or a polypeptide chain permutation should occur to join the current N- and C-termini, and cut the chain at the N-terminus of α5.

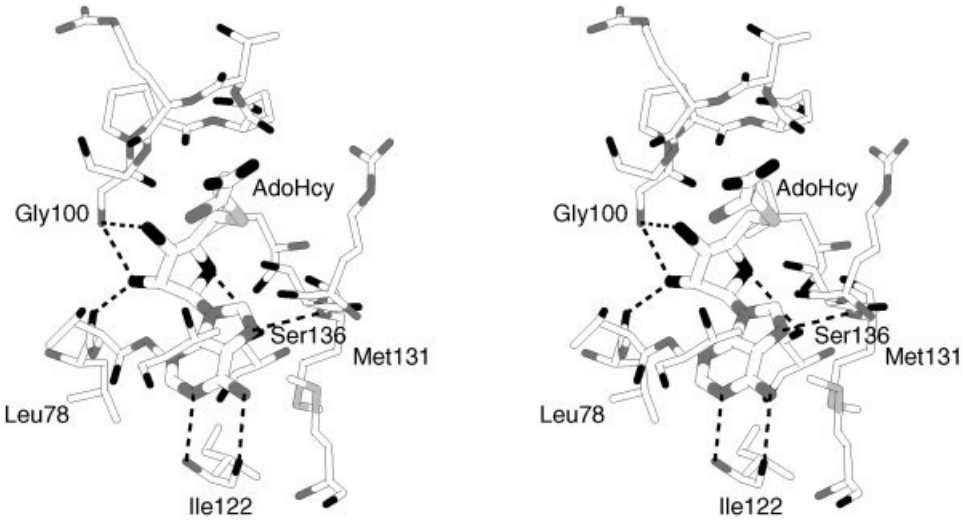

Fig. 7. Stereoscopic drawing of the interaction of AdoHcy with amino acid residues of HI0766. The labeled residues form hydrogen bonds with AdoHcy.

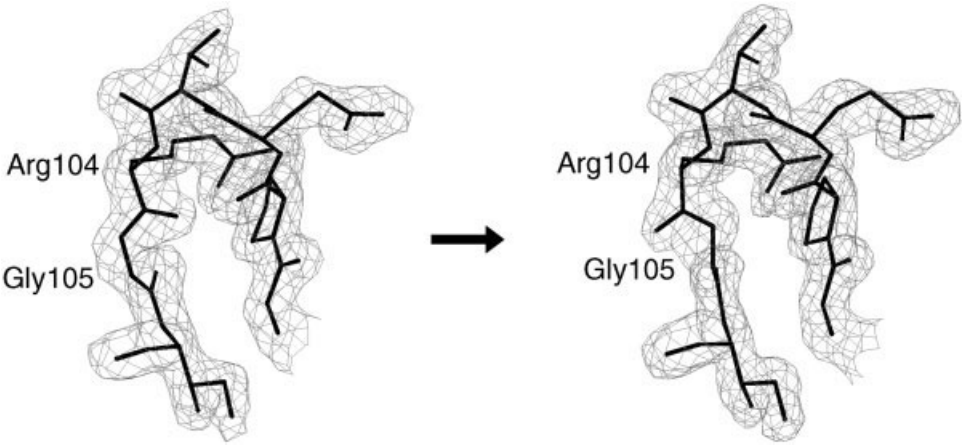

Fig. 8. Peptide flip of Arg104-Gly105. The two $2F_o - F_c$ maps contoured at 1σ show the peptide bond between Arg104 and Gly105 in the 100–106 loop before cofactor binding (**left**), and after cofactor binding (**right**).

Then α5 would form an N-terminal helix analogous to that seen in the classical methyltransferases.

If such a transition occurred in the course of evolution, elaborate DNA recombination events that result in chain shuffling and permutation must have taken place, together with changes leading to core repacking. Moreover, the transition requires that the permutation occur first, followed by the β6 shift and chain shuffling. These events are much too complex to propose an evolutionary relationship between the knotted and unknotted folds.

Function of HI0766

The close sequence relatives of HI0766 obtained after one PSI-BLAST iteration share numerous amino acid residues (Fig. 5). They are similar in length (~160 residues), in contrast to the sequences obtained after the second round of PSI-BLAST, which are 230 residues long or more. The sequence identity among the close relatives of HI0766 is >45%, whereas the longer spoU family members obtained after the second round of PSI-BLAST exhibit <35% sequence identity to HI0766. Yet the two groups share distinct sequence fingerprint regions that correspond to the cofactor-binding site and to the dimer interface.

The spoU gene from *E. coli* has been shown to code for a tRNA Gm18 2′-O-methyltransferase.[16] Other family members are annotated as rRNA or tRNA methyltransferases. These proteins consist of the cofactor-binding domain homologous to HI0766 and a second domain, likely to define the substrate specificity of the enzyme. Family members whose function was determined experimentally, rather than solely by sequence homology, indicate that the specificity domain may be formed by the C-terminal chain in tRNA methyltransferase,[16] and by the N-terminal chain in rRNA methyltransferases.[42] Based on the structural similarities, in particular the presence of the highly un-

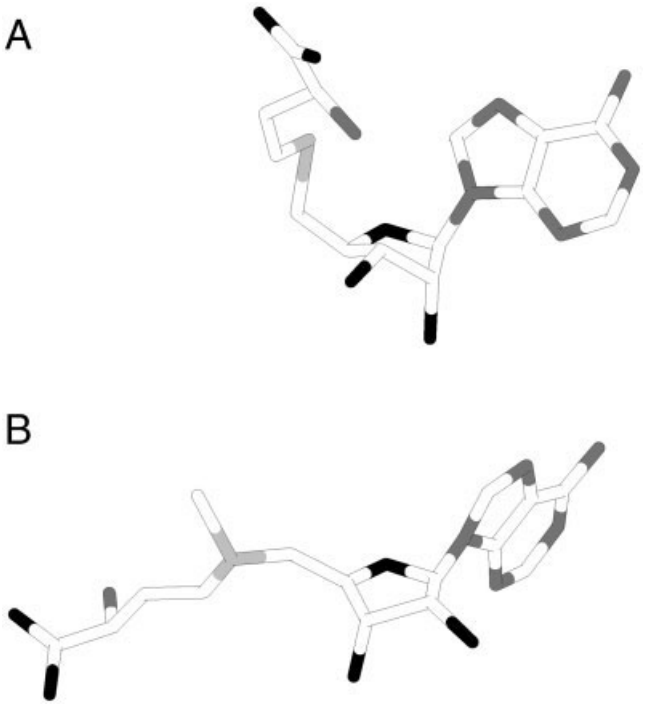

Fig. 9. Comparison of the cofactor conformations. **A:** AdoHcy from HI0766. **B:** AdoMet from catechol O-methyltransferase.[33] The two molecules were separated after superimposition with their ribose ring atoms. This figure shows the C3′-endo conformation of HI0766 and C2′-endo conformation of catechol O-methyltransferase, along with the different orientations of the adenine rings.

usual knot, we propose that the hypothetical protein MT0001 is also a methyltransferase, with its specificity domain inserted in the middle of the chain that forms the cofactor-binding domain. The all-β OB fold of this putative specificity domain is indicative of oligonucleotide binding, but the exact substrate is unknown.

The *Haemophilus influenzae* genome codes for three long spoU family members, HI0380, HI0860, and HI0424. The postulated substrate-binding domain of HI0380 follows the cofactor-binding domain, indicating activity toward tRNA, whereas the substrate-binding domains of HI0860 and HI0424 precede the cofactor-binding domain, as expected of an rRNA methyltransferase. Other genomes also contain both the short HI0766 homologue and multiple versions of the long spoU homologues. Assuming that each member acts on a different substrate, the superfamily can be divided into at least four subclasses. We propose that the HI0766 sequence family members function together with partner proteins whose identities are yet to be discovered. In our hands, sequence analysis and genome context failed to identify stand-alone proteins with homology to one of the specificity domains of the spoU classes that could form a complex with the respective HI0766 homologues.

While the manuscript was being reviewed, the structure of RNA 2′-O-ribose methyltransferase from *Thermus thermophilus* in the unbound state was published, also revealing a fold with a knot.[43]

ACKNOWLEDGMENTS

We thank John Moult and Eugene Melamud for the use and help with their bioinformatics web site (http://s2f.carb.nist.gov) and the Structural Genomics team at CARB for stimulating discussions. We thank the staff of IMCA-CAT at the Advanced Photon Source for their help during data collection. The IMCA-CAT facility is supported by the companies of the Industrial Macromolecular Crystallographic Association, through a contract with IIT. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract W-31-109-Eng-38. The Keck foundation provided generous support for the purchase of X-ray equipment at CARB. PDB coordinates entry codes: 1j85 and 1mwi.

REFERENCES

1. Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, McCann PP. S-adenosylmethionine and methylation. FASEB J 1996;10:471–480.
2. Swaminathan CP, Sankpal UT, Rao DN, Surolia A. Water-assisted dual mode cofactor recognition by *Hha*I DNA methyltransferase. J Biol Chem 2002;277:4042–4049.
3. Loehrer FM, Tschopl M, Angst CP, Litynski P, Jager K, Fowler B, Haefeli WE. Distributed ratio of erythrocyte and plasma S-adenosylmethionine/S-adenosylhomocysteine in peripheral arterial occlusive disease. Atherosclerosis 2001;154:147–154.
4. Mato JM, Corrales FJ, Lu SC, Avila MA. S-adenosylmethionine: a control switch that regulates liver function. FASEB J 2002;16:15–26.
5. Reguera RM, Balana-Fouce R, Perez-Pertejo Y, Fernandez FJ, Garcia-Estrada C, Cubria JC, Ordonez C, Ordonez D. Cloning expression and characterization of methionine adenosyltransferase in *Leishmania infantum* Promastigotes. J Biol Chem 2002;277:3158–3167.
6. Morana A, Di Lernia I, Carteni M, De Rosa R, De Rosa M. Synthesis and characterization of a new class of stable *S*-adenosyl-L-methionine salts. Int J Pharm 2000;194:61–68.
7. Hoffman JL. Chromatographic analysis of the chiral and covalent instability of *S*-adenosyl-L-methionine. Biochemistry 1986;25:4444–4449.
8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
9. Cheng X, Blumenthal RM, editors. S-adenosylmethionine-dependent methyltransferases: structures and functions. New York: World Scientific; 1999.
10. Skinner MM, Puvathingal JM, Walter RL, Friedman AM. Crystal structure of protein isoaspartyl methyltransferase: a catalyst for protein repair. Structure 2000;8:1189–1201.
11. Zhang X, Zhou L, Cheng X. Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. EMBO J 2000;19:3509–3519.
12. Weiss VH, McBride AE, Soriano MA, Filman DJ, Silver PA, Hogle JM. The structure and oligomerization of the yeast arginine methyltransferase, Hmt1. Nat Struct Biol 2001;7:1165–1171.
13. Schubert HL, Wilson KS, Raux E, Woodcock SC, Warren MJ. The X-ray structure of a cobalamin biosynthetic enzyme, cobalprecorrin-4 methyltransferase. Nat Struct Biol 1998;5:585–592.
14. Anantharaman V, Koonin EV, Aravind L. SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. J Mol Microbiol Biotechnol 2002;4:71–75.
15. Koonin EV, Rudd KE. SpoU protein of *Escherichia coli* belongs to a new family of putative rRNA methylases. Nucleic Acids Res 1993;23:5519.
16. Persson BC, Jager G, Gustafsson C. The *spoU* gene of *Escherichia coli*, the fourth gene of the *spoT* operon, is essential for tRNA (Gm18) 2′-*O*-methyltransferase activity. Nucleic Acids Res 1997;25:4093–4097.
17. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkmess EF,

Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 1995;269:496–512.
18. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lippman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
19. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
20. Terwilliger TC, Berendzen J. Automated structure solution for MIR and MAD. Acta Crystallogr D 1999;55:849–861.
21. CCP4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D 1994;50:760–763.
22. Otwinowski Z. Maximum likelihood refinement of heavy atom parameters. In: Wolf W, Evans PR, Leslie AGW, editors. Isomorphous replacement and anomalous scattering. Warrington, UK: Daresbury Laboratory; 1991. p 80–86.
23. Cowtan K. DM: an automated procedure for phase improvement of density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 1994;31:34–38.
24. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
25. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 1998;54:905–921.
26. Kleywegt GJ, Jones TA. Databases in protein crystallography. Acta Crystallogr D 1998;54:1119–1131.
27. Laskowski RA, MacArthur MW, Moss DS, Thornton J. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
28. Connolly ML. Analytical molecular surface calculation. J Appl Crystallogr 1983;16:548–558.
29. Kraullis PJ. A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
30. Bacon DJ, Anderson WF. A fast algorithm for rendering space-filling molecule pictures. J Mol Graph 1988;6:219–220.
31. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. Methods Enzymol 1997;277:505–524.
32. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. J Mol Biol 1993; 233:123–138.
33. Vldgren J, Svensson LA, Liljas A. Crystal structure of catechol O-methyltransferase. Nature 1994; 368:354–358.
34. Pflugrath JW, Quiocho FA. Sulphate sequestered in the sulphate-binding protein of *Salmonella typhimurium* is bound solely by hydrogen bonds. Nature 1985;314:257–261.
35. Vaney MC, Broutin I, Retailleau P, Douangamath A, Lafont S, Hamiaux C, Prange T, Ducruix A, Ries-Kautt M. Structural effects of monovalent anions on polymorphic lysozyme crystals. Acta Crystallogr 2001;D57:929–940.
36. Saenger W Principles of nucleic acid structure. New York: Springer-Verlag; 1984. ch. 1.
37. Schluckebier G, Kozak M, Bleimling N, Weinhold E, Saenger W. Differential binding of S-adenosylmethionine, S-adenosylhomocysteine, and sinefungin to the adenine-specific DNA methyltransferase M. *Taq*I. J Mol Biol 1997;265:56–67.
38. Dixon MM, Huang S, Matthews RG, Ludwig M. The structure of the C-terminal domain of methionine syntase: presenting S-adenosylmethionine for reductive methylation of B_{12}. Structure 1996;4:1263–1275.
39. Taylor WR. A deeply knotted protein structure and how it might fold. Nature 2000;406:916–919.
40. Takusagawa F, Kamitori S. A real knot in protein. J Am Chem Soc 1996;118:8945–8946.
41. Ye S, Goldsmith EJ. Serpins and other covalent protease inhibitors. Curr Opin Struct Biol 2001;11:740–745.
42. Bechthold A, Floss HG. Overexpression of the thiostrepton-resistance gene from *Streptomyces azureus* in *Escherichia coli* and characterization of recognition sites of the 23S rRNA A1067 2′-methyltransferase in the guanosine triphosphatase center of 23S ribosomal RNA. Eur J Biochem 1994;224:431–437.
43. Nureki O, Shirouzu M, Hashimoto K, Ishitami R, Terada T, Tamakoshi M, et al. An enzyme with a deep trefoil knot for the active-site architecture. Acta Crystallogr 2002;D58:1129–1137.
44. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.