X-ray Crystal Structure of Glycinamide Ribonucleotide Synthetase from *Escherichia coli*[†,‡]

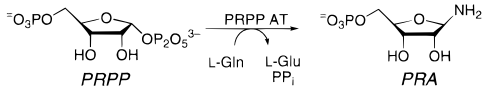
Weiru Wang,[§] T. Joseph Kappock,[||] JoAnne Stubbe,[||] and Steven E. Ealick*[,§]

Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, and Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received June 15, 1998; Revised Manuscript Received September 9, 1998

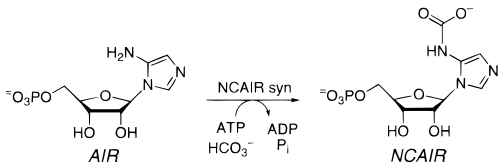
ABSTRACT: Glycinamide ribonucleotide synthetase (GAR-syn) catalyzes the second step of the de novo purine biosynthetic pathway; the conversion of phosphoribosylamine, glycine, and ATP to glycinamide ribonucleotide (GAR), ADP, and P_i. GAR-syn containing an N-terminal polyhistidine tag was expressed as the SeMet incorporated protein for crystallographic studies. In addition, the protein as isolated contains a Pro294Leu mutation. This protein was crystallized, and the structure solved using multiple-wavelength anomalous diffraction (MAD) phase determination and refined to 1.6 Å resolution. GAR-syn adopts an α/β structure that consists of four domains labeled N, A, B, and C. The N, A, and C domains are clustered to form a large central core structure whereas the smaller B domain is extended outward. Two hinge regions, which might readily facilitate interdomain movement, connect the B domain and the main core. A search of structural databases showed that the structure of GAR-syn is similar to D-alanine:D-alanine ligase, biotin carboxylase, and glutathione synthetase, despite low sequence similarity. These four enzymes all utilize similar ATP-dependent catalytic mechanisms even though they catalyze different chemical reactions. Another ATP-binding enzyme with low sequence similarity but unknown function, synapsin Ia, was also found to share high structural similarity with GAR-syn. Interestingly, the GAR-syn N domain shows similarity to the N-terminal region of glycinamide ribonucleotide transformylase and several dinucleotide-dependent dehydrogenases. Models of ADP and GAR binding were generated based on structure and sequence homology. On the basis of these models, the active site lies in a cleft between the large domain and the extended B domain. Most of the residues that facilitate ATP binding belong to the A or B domains. The N and C domains appear to be largely responsible for substrate specificity. The structure of GAR-syn allows modeling studies of possible channeling complexes with PPRP amidotransferase.

The importance of channeling, direct transfer of unstable intermediates between successive enzymes in metabolic pathways, and the requirement of transient protein−protein interactions for this process has been strongly debated over the past decade (*1*). We have used the purine biosynthetic pathway as a vehicle to test this hypothesis. Conversion of phosphoribosyl pyrophosphate (PRPP) to inosine monophosphate (IMP) requires 11 catalytic transformations, two of which generate chemically unstable intermediates: PRPP amidotransferase (PRPP-AT) catalyzes the conversion of PRPP and glutamine to phosphoribosylamine (PRA), glutamate, and PP_i (eq 1) and *N*-carboxyaminoimidazole



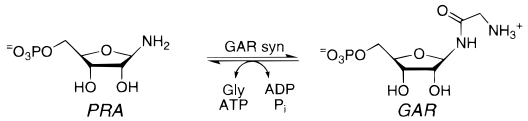
(Eq. 1)

ribonucleotide (NCAIR) synthetase catalyzes the conversion of 5-aminoimidazole ribonucleotide (AIR), HCO_3^-, and ATP to NCAIR, ADP, and P_i (eq 2). In the case of PRPP-AT,



(Eq. 2)

PRA has a half-life of 5 s at 37 °C. Recent kinetic studies in vitro (*2*) have suggested that this intermediate might be channeled directly to the second enzyme in this pathway, glycinamide ribonucleotide synthetase (GAR-syn), which catalyzes the reversible formation of GAR from PRA, ATP, and glycine (eq 3). In an effort to obtain structural evidence



(Eq. 3)

[†] This work was supported by a National Institutes of Health grant to J.S. (GM32191). T.J.K. is supported by NIH cancer training grant CA09112. S.E.E. is indebted to the W. M. Keck Foundation and the Lucille P. Markey Charitable Trust. This work is based upon research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the National Science Foundation under award DMR-9311772, using the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award RR-01646 from the National Institutes of Health.

[‡] The Brookhaven Protein Data Bank code for glycinamide ribonucleotide synthetase is 1GSO.

* Corresponding author. Phone: (607) 255-7961. Fax: (607) 255-4137. E-mail: see3@cornell.edu.

[§] Cornell University

[||] Massachusetts Institute of Technology.

10.1021/bi981405n CCC: $15.00 © 1998 American Chemical Society
Published on Web 10/22/1998

for direct transfer of PRA between these first two enzymes in this pathway, we have crystallized and solved the structure of *Escherichia coli* GAR-syn. An examination of this structure relative to the available structures of the *E. coli* (*3*, *4*) and *B. subtilis* (*5*) PRPP-ATs has provided a basis for designing appropriate mutagenesis and kinetic studies to test the channeling hypothesis in vivo and in vitro.

GAR-syn from *E. coli* is a monomer of M_r 46 kDa. Steady-state kinetic analysis of this system revealed a sequential ordered mechanism with PRA binding before ATP and glycine, and P_i leaving followed by ADP and GAR (*6*). Recent sequence alignments have suggested that GAR-syn and two additional enzymes in the purine biosynthetic pathway, PurK (catalyzing the formation of NCAIR from AIR, eq 2) and PurT (catalyzing the formation of formylglycinamide ribonucleotide (FGAR) from GAR, ATP, and formate), are part of a superfamily of enzymes with a common ATP-binding fold (*7*). This new fold has been called the ATP-grasp domain. Additional enzymes within this family include biotin carboxylase, BNC (*8*), D-alanine:D-alanine ligase, DDLigase (*9*), glutathione synthetase, GSHase (*10*), and most recently synapsin Ia, SynC (*11*). Most of these enzymes are ATP-dependent carboxylate amine ligases. Additional members of this class possess ATP-carboxylate dependent thiol or alcohol ligase activity. Their catalytic mechanisms are postulated to involve acyl phosphate intermediates. The function of SynC, which binds Ca^{2+} and ATP, is still unknown. Recently, the structure of 4-(*N*-succinocarboxamide)-5-aminoimidazole ribonucleotide (SAICAR) synthase, another enzyme in the purine biosynthetic pathway, was also shown to contain some structural similarities with this superfamily (*12*).

The present paper describes the first structure of a GAR-syn and establishes that it is indeed a member of the ATP-grasp superfamily. Furthermore, comparison with structures of additional members of this class (DDLigase and GSHase) with ligands bound in their active sites give us insight into the mode of binding of the GAR-syn substrates and the conformational changes accompanying catalysis. In addition, a structure homology search surprisingly turned up similarities to PurN, a GAR transformylase, and the enzyme after GAR-syn in the purine pathway. This unexpected result suggests that binding specificity as well as chemistry may have played a role in the evolution of enzymes in this pathway. Finally, analysis of the structures of PPRP-AT and GAR-syn allowed us to perform manual docking studies that resulted in a model for a possible channeling complex.

MATERIALS AND METHODS

Protein Expression and Purification. Selenomethionine-incorporated GAR-syn was expressed and purified as previously described (*13*). The purD gene was cloned into pET2a.1 vector and transformed into a competent *E. coli* methionine auxotroph B834(DE3) obtained from Novagen. Cells were grown in M9 minimal media supplemented with 40 mg/L seleno-L-methionine to an optical density of 0.6 at which time they were induced with 1 mM isopropyl β-D-thiogalactopyanoside (IPTG) and grown for an additional 12 h at 37 °C. The protein was purified by chromatography using nickel(II)-nitriloacetic acid column using an imidazole gradient elution. The protein had no catalytic activity resulting from a P294L mutation.

Repair of the P294L Mutation of GAR-syn. The wild-type sequence of GAR-syn was restored by inserting the 898 bp *Nco*I−*Hin*dIII fragment from pJS187 (*6*), which encodes residues 66−430, into the corresponding region of the His-tag GAR-syn expression construct. DNA sequencing of both strands indicated that the sequence of GAR-syn following the 20 amino acid N-terminal His-tag extension was correct.

Enzyme Assays. GAR-syn was assayed using a coupled assay with pyruvate kinase and lactate dehydrogenase monitoring the consumption of NADH at 18 °C as described previously (*6*). PRA was formed from ribose 5-phosphate and NH_4Cl, under basic conditions as previously described (*6*). One unit of activity is defined as the amount of enzyme required to produce 1 μmol of ATP/min at 18 °C.

Crystallization of SeMet GAR-syn. Small initial crystals of SeMet GAR-syn were obtained using the sparse matrix protocol (*14*). Optimized crystallization conditions were in 0.1 M 2-(*N*-morpholino)ethanesulfonic acid (MES), pH 6.3, 0.1 M ammonium sulfate, and 24% poly(ethylene glycol) 5K monomethyl ether (PEG 5K MME) using the hanging drop vapor diffusion method at 18 °C (*13*). Crystals appeared within a few days and reached maximum size of 0.4 mm × 0.6 mm × 1.0 mm over the course of an additional few weeks. Crystals were characterized to be in space group $P2_12_12_1$ with unit cell dimensions $a = 56.22$ Å, $b = 62.42$ Å, and $c = 129.77$ Å. Each asymmetric unit contains one monomeric molecule with 50% solvent content. Crystals were transferred to a cryoprotectant solution composed of the mother liquor and 10% (w/v) glycerol for 2 min and then directly mounted in loops for flash freezing in a gaseous N_2 stream. Flash-frozen crystals were then maintained at the cryogenic temperature (∼−170 °C) for the duration of the experiment.

X-ray Diffraction Data Collection. All X-ray intensity data were collected at the Cornell High Energy Synchrotron radiation Source (CHESS). A nearly complete 1.6 Å resolution monochromatic data set was collected on CHESS station F1 with the X-ray wavelength tuned to 0.9190 Å. One frozen GAR-syn crystal was used to measure 150° of data. Each data frame was collected with a 1° oscillation angle and a 25 s exposure time. The data were recorded with a Quantum 4 mosaic CCD-based X-ray detector (Area Detector System Corporation, ADSC) placed 150 mm from the sample. Individual data frames were corrected for background, spatial distortion, and nonuniformity before final data processing.

A three wavelength MAD (*15*) data set was collected to 2.2 Å resolution using a Princeton Scientific Instruments (PSI) 2K × 2K CCD detector placed 90 mm from the sample and X-rays from CHESS station F2 station. This doubly focused station is equipped with a two crystal monochromator for rapid tuning. The station was calibrated by locating the absorption edge of a Se foil. The calibration scan was repeated periodically to confirm that no significant changes in energy occurred during the measurement of data. The wavelengths for the MAD experiment were chosen by scanning through the absorption edge of a SeMet GAR-syn crystal. The three wavelengths were chosen as the minimum f' (edge), the maximum f'' (peak), and a reference wavelength above the absorption edge (remote). A total of 99° of data was measured using the inverse beam method to ensure that Bijvoet pairs were measured close together in

Table 1: Statistics of GAR-syn SeMet Data Collected at CHESS

data set	native	edge	peak	remote
wavelength (Å)	0.919	0.979 410	0.979 104	0.967 642
resolution (Å)	20−1.60	20−2.2	20−2.2	20−2.2
redundancy	5.2(4.1)[a]	6.2(3.5)	6.1(3.4)	6.2(3.6)
unique reflections	58 385(7256)	22 702(1234)	22 800(1245)	22 820(1279)
completeness of data (%)	95.9(83.9)	97.5(80.9)	97.7(81.7)	97.8(83.9)
I/σ	10(4.1)	22(10.5)	22(10.5)	22(11.2)
R_{sym}[b] (%)	4.3(17.4)	6.5(10.0)	7.5(11.2)	6.7(10.1)

[a] Numbers in parentheses are related to the highest resolution shell, which is 2.25−2.20 Å for MAD data and 1.69−1.60 Å for monochromatic data. [b] $R_{sym} = \sum_{hkl}\sum_i |I_{hkl,i} - \langle I\rangle_{hkl}|/\sum|\langle I\rangle_{hkl}|$.

Table 2: MAD Difference Ratios[a]

wavelength	edge	peak	remote
edge	0.093 (0.081)	0.052	0.083
peak		0.114 (0.080)	0.076
remote			0.097 (0.075)

[a] MAD difference ratios are rms($|\Delta F|$)/rms($|F|$), where ΔF is the Bijvoet difference at one wavelength (diagonal elements) or dispersive difference between pairs of wavelengths (off-diagonal elements) and F is the average value of the structure factor magnitude for that energy. The values in parentheses are ratios for centric reflections, which ideally should be zero. They are indicators of noise in the anomalous signal.

time. Data frames were collected using a 1° oscillation measured for 40 s. Data were measured in groups of three frames following the sequence edge, Bijvoet pairs for edge, peak, Bijvoet pairs for peak, remote, and Bijvoet pairs for remote. Each data frame was corrected for background, spatial distortion, and nonuniformity before final data processing.

Data Processing. Monochromatic data measured with the ADSC Quantum 4 mosaic CCD detector were processed using the CCP4 suite of programs (*16*). The program MOSFLM (*17*) was used to determine and refine the crystal orientation and unit cell parameters, to refine the crystal to detector distance and to determine integrated X-ray diffraction intensities. Data from different frames were scaled using the program SCALA (*18*). The program TRUNCATE (*19*) was used to calculate structure factor magnitudes from averaged intensities. A total of 303 050 measurements were recorded to 1.6 Å resolution giving 58 385 unique reflections after scaling and merging. The overall data completeness is 95.9% with an R_{sym} of 4.3% (see Table 1).

The MAD data from the PSI detector were processed using the computer programs DENZO and SCALEPACK (*20*). Each of the six data sets, three wavelengths each with direct and inverse beam, were integrated and scaled separately. Data processing statistics are shown in Table 1. Further scaling was performed in the MADSYS package (*15*). All six data sets were processed to 2.2 Å resolution because of limitations imposed by the size of the active area of the PSI detector. However, the data extended well beyond this resolution with an I/σ greater than 10 in the highest resolution shell. The typical overall completeness for a data set was about 97% with R_{sym} values ranging 6.5−7.5% (see Table 1). The anomalous signal from the MAD data collection is reported in Table 2.

Location of Se Atom Positions and Phase Calculation. The Se atom positions were located using both direct methods and automated Patterson map searching. The program SnB (*21*) which implements Hauptman's shake-and-bake procedure (*22*) was used for the direct methods approach. First, the anomalous scattering contributions to the structure factor magnitudes, F_a, were estimated using the MADSYS programs with locally scaled data (*15*). A total of 3318 F_a values were used to calculate normalized structure factor magnitudes, E, using the program BAYES (*23*). A total of 1000 SnB trials were performed using 500 E values, 5000 triple-phase relationships, and 20 phase refinement cycles. Out of 1000 trials, one phase set converged to a solution that revealed nine correct Se atom positions. A second SnB run was performed using only the peak wavelength, which contains the maximum average anomalous differences. A total of 41 412 anomalous differences were converted to E values. A total of 1000 SnB trials were performed using the 500 largest E values, 5000 triple-phase relationships, and 30 phase refinement cycles. Three of these trials resulted in solutions revealing 13 of the 14 possible Se atoms positions. Patterson map searching was carried out with the program SOLVE (*24*), which utilizes three wavelength MAD data and the Patterson function to perform an automated search for the positions of the anomalous scattering atoms. In the first attempt using this program, the data set was restricted to 3.5 Å resolution. Under these conditions, SOLVE found 10 Se atom sites. When the resolution was extended to 2.2 Å resolution, SOLVE found 12 Se sites.

Using the combined results from SnB and SOLVE, 13 Se atom positions were identified. The 14th Se atom, corresponding to the N-terminal methionine, was not located by any technique and was assumed to be disordered. The 13 ordered Se atom positions were refined in an iterative manner, using the least-squares method incorporated in ASLSQ and MADABCD of the MADSYS package to yield experimental phases with a mean figure of merit of 0.57. To improve the experimental phases, the heavy atom position refinement was then performed using the maximum likelihood method incorporated in program MLPHARE (*25*). This method gave final experimental phases up to 2.2 Å resolution with an overall figure of merit of 0.69.

Model Building and Refinement. All model building was performed using the graphical interface and refinement tools of the program O (*26*). An electron density map calculated with MLPHARE phases displayed extensive well-defined regions containing long continuous regions of main-chain density with clear density for most side chains and carbonyl oxygen atoms. A representative section of the map is shown in Figure 1. Approximately 70% of the model was built from the initial map.

The initial map was then subjected to eight cycles of solvent flattening (*27*) with 45% solvent content resulting

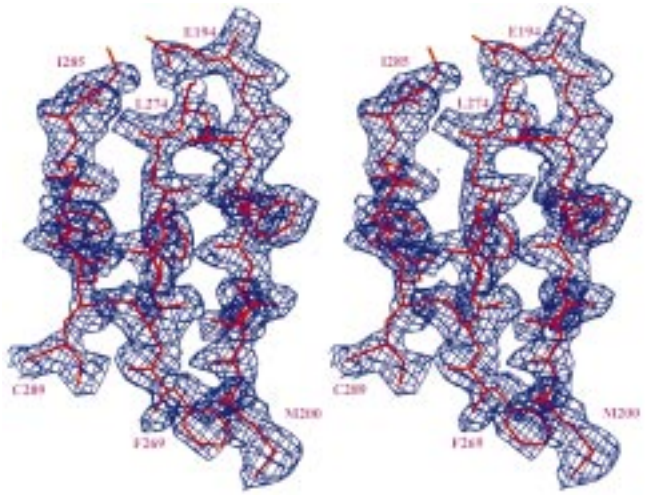


FIGURE 1: Typical electron density of the experimental map at 2.2 Å resolution. The electron density map was calculated using MLPHARE phases and contoured at the 1σ level. The electon density corresponds to a section of β-sheet. The first and last residue in each peptide segment is labeled.

in an overall figure of merit of 0.903 and with a value of 0.828 in the resolution shell 2.27−2.20 Å. This map cleared up several confusing regions in the initial map and enabled us to build about 80% of the molecule. Further map improvement was achieved by combining the experimentally determined MLPHARE phases and calculated phases based on a polyalanine model. The phase combination was performed with the program SIGMAA (*28*). A model containing 416 amino acid residues was derived from progressive improvement of the electron density map using four additional rounds of phase combination and model building. In all models, most of the polyhistidine tag was missing and the loop region 146−153 was not visible.

The preliminary model was then refined to 1.6 Å resolution using the monochromatic data and the program X-PLOR (*29*). After initial rigid body refinement, positional and thermal parameters were refined for all atoms. Occupancies were refined for water molecules only. A total of 10% of the data was randomly picked out for free *R*-factor cross validation. Cycles of refinement combined with difference Fourier maps resulted in model improvements, especially for residues Ala146, Lys153, Ile424, Glu425, and Arg426, which were unclear in earlier electron density maps.

Preparation of Figures. Figures were generated with InsightII (MSI, Inc., San Diego, CA), GRASP (*30*), MOLSCRIPT (*31*), Raster3D (*32*), RIBBONS (*33*), ALSCRIPT (*34*), or ChemDraw 4.5 (Cambridge Soft Corporation, Cambridge, MA).

RESULTS

Protein Expression and Enzyme Characterization. In its native form, *E. coli* GAR-syn is easily purified to homogeneity by ion-exchange chromatography as a soluble monomer of 46 kDa (*6*). However, experiments with homogeneous wild-type material failed to yield useful crystallization conditions (Weaver, T., unpublished results). This result together with the desire to prepare SeMet incorporated GAR-syn lead to an alternative procedure. An expression construct that adds 21 additional N-terminal amino acids (MGSSHHHHHHNSGLVPRGSEF), including a sequence of six consecutive His residues (pHGS), was prepared using PCR techniques. This elongated GAR-syn was overexpressed in an *E. coli* strain auxotrophic for methionine in which SeMet replaced Met. The protein was purified to homogeneity in a single step using nickel-dependent affinity chromatography on an iminodiacetic acid column. This protein was used to find appropriate crystallization conditions (*13*). Because of the His-tag affinity purification, contaminating GAR-syn from the host strain is probably absent from this protein.

Enzymatic activity assays of the SeMet substituted, His-tagged material at 18 °C showed that it contained no detectable activity, ≤0.001 units/mg, relative to an activity of 43 units/mg previously observed for wild-type *E. coli* GAR-syn (*2*). In addition, the pHGS plasmid failed to complement the *pur*D auxotrophic strain TX634 on selective minimal medium, confirming that the GAR-syn construct was essentially devoid of activity. DNA sequencing of this construct established that a point mutation had been acquired at some step during construction of pHGS, which altered the highly conserved proline at position 294 to leucine. This transversion falls within the characteristic sequence RFGD-PExQ, which is found in all known GAR-syns [SwissProt Prosite PS00184 (*35*)], but not in any other protein found in current database sequences.

While it seemed likely that this mutation contributed to the observed loss of function, the addition of a foreign N-terminal sequence and the substitution of Met residues with SeMet posed additional concerns. A plasmid DNA construct with Pro294 restored within the His-tagged protein coding sequence (pJK33) was prepared and used to overexpress the repaired Met- and SeMet-substituted enzymes, which were each purified and assayed for enzymatic activity. These proteins had specific activity values of 19 units/mg and 2.3 units/mg, respectively, up to 3 orders of magnitude higher than the Pro294Leu mutant. Unlike the near-wild type level of activity found for the Met-containing enzyme, SeMet substitution seems to have an additional negative effect on activity that could be due to the comparatively facile oxidation of this side chain. Finally, TX634 cells transformed with pJK33 showed full complementation of the *pur*D auxotrophy.

SDS−PAGE showed that the His-tagged protein migrates slightly slower than the wild-type protein, in a manner consistent with the calculated molecular msss of 48 243 Da. The aggregation state of native His-tagged GAR-syn was assessed by analytical size-exclusion chromatography at comparatively low ionic strength (50 mM Tris, pH 8.0). Both Met substituted, repaired GAR-syn and the SeMet substituted Pro314Leu mutant eluted at the same volume as wild-type *E. coli* GAR-syn, at an apparent $M_r = 49 \pm 3$ kDa. This indicates that the His-tagged proteins containing either Met or SeMet are monomeric and compact in solution.

Structure of GAR-syn. The refined GAR-syn model contains 422 residues and 305 water molecules. The final refinement statistics are shown in Table 3. Several regions of the protein are totally disordered or have high thermal parameters including the poly-His tag, the loop composed of residues 147−152, referred to as the B loop, and three C-terminal residues. A Ramachandran plot, using the program PROCHECK (*36*), reveals that 90.9% of the amino acid residues adopt dihedral angles lying in the most favored regions with an additional 8.8% of the amino acids in less

Table 3: Crystal Data and Refinement Statistics

space group	$P2_12_12_1$
cell dimensions	$a = 56.2$ Å, $b = 62.4$ Å, $c = 129.8$ Å
z	4
protein fraction	50%
V_m (Å^3/Da)	2.47
total no. of residues	422
total non H atoms	3433
no. of water molecules	305
no. of multiple conformer sidechains	9
temperature factors[a]	
protein	22.1 (25.8) Å^2
domain N	14.3 (17.5) Å^2
domain A	19.9 (22.9) Å^2
domain B	43.5 (46.9) Å^2
domain C	20.4 (23.7) Å^2
solvent	36.75 Å^2
resolution range of reflections used	10−1.6 Å
amplitude cutoff	$\|F\| > 3\sigma(F)$
R-factor	20.9%
free R-factor	24.7%
stereochemical ideality	
bond	0.007 Å
angle	1.25°
improper	1.15°

[a] Number refers to main-chain atoms only, number in parentheses refers to side-chain atoms only.

common, but still allowed regions. Four proline residues, 32, 141, 235, and 342, adopt a cis peptide configuration.

The overall monomeric GAR-syn molecule has approximate dimensions of 62 Å × 49 Å × 42 Å. Ribbon representations and a topology plot of the molecular structure are shown in Figures 2 and 3, respectively. The protein displays an α/β structure that contains 16 α-helices and 20 β-strands connected by turns and loops. Table 4 shows a list of the secondary structural elements. There are four structural domains, each with a central β-sheet flanked by α-helices on one or both sides. The four domains labeled A, B, C, and N are shown in Figures 2 and 3. The nomenclature describing these domains is based on that used for BNC (*8*), which is a member of a superfamily of ATP-dependent enzymes (*7*). The N, A, and C domains are clustered into a large core structure, while the B domain is somewhat separated and extended away from the core structure (Figure 2a). The edges of the A and C domain β-sheets are joined by strands β12 and β15 to form a continuous β-sheet structure. This larger β-sheet is oriented roughly perpendicular to the N domain β-sheet. The active site is located between the large core structure formed by the N, A, and C domains and the extended B domain flap. The B domain is connected to the N domain and the A domain by two flexible loops, 114−122 (hinge NB) and 189−191 (hinge BA), respectively, that could facilitate B domain movement relative to rest of the molecule.

Description of the Domains. The N domain (Figures 2b and 3) includes residues 1−118 and consists of a five-stranded parallel β-sheet, composed of strands β1−β5. The sheet topology is β3β2β1β4β5. Two α-helices (α1 and α2) flank the inner surface of β-sheet facing C domain and together with the first three β-strands form a Rossmann fold (*37*). Three α-helices (α3, α4, and α5) flank the side of the β-sheet facing the solvent. Gly79 adopts ϕ, ψ angles of 84.9° and 31.5°, producing a kink between α4 and α5, which is followed by two short α-helices (α6 and α7). The first

Table 4: Secondary Structure Elements in GAR-syn

N domain		A domain	
range	secondary structure (name)	range	secondary structure (name)
1−7	β-strand (β1)	192−202	β-strand (β10)
10−19	α-helix (α1)	206−218	β-strand (β11)
20−21	γ-turn (T1)	220−222	β-turn III′ (T10)
23−24	γ-turn (T2)	223−234	β-strand (β12)
25−32	β-strand (β2)	236−238	β-turn III (T11)
35−39	α-helix (α2)	241−250	α-helix (α10)
40−42	β-turn I (T3)	252−261	α-helix (α11)
43−45	β-strand (β3)	267−277	β-strand (β13)
49−51	β-turn I (T4)	277−279	β-turn I (T12)
53−62	α-helix (α3)	280−288	β-strand (β14)
67−69	β-strand (β4)	293−295	β-turn III (T13)
73−76	α-helix (α4)	296−302	α-helix (α12)
77−79	β-turn I (T4)	307−315	α-helix (α13)
80−86	α-helix (α5)	C domain	
87−88	β-turn IV (T6)		
91−92	β-strand (β5)		
95−98	β-turn III (T7)	range	secondary structure (name)
100−103	α-helix (α6)	319−321	3_{10}−helix(α14)
105−114	α-helix (α7)	330−338	β-strand (β15)
B domain		340−342	β-turn Via (T14)
range	secondary structure (name)	360−362	β-turn II (T15)
		363−366	β-strand (β16)
		369−371	β-strand (β17)
122−125	β-strand (β6)	377−379	β-strand (β18)
130−138	α-helix (α8)	383−391	β-strand (β19)
142−145	β-strand (β7)	394−404	α-helix (α15)
155−158	β-strand (β8)	405−407	β-turn III (T16)
160−168	α-helix (α9)	411−412	β-turn II (T17)
176−178	β-turn II (T8)	413−414	β-strand (β20)
179−181	β-turn I (T9)	418−420	β-turn II (T18)
184−188	β-strand (β9)	421−425	α-helix (α16)

link to the extended B domain is formed by residues 114−122 (hinge NB).

The B domain (Figures 2c and 3) includes residues 119−191 and consists of a four-stranded antiparallel β-sheet (β6−β9) with two α-helices (α8 and α9) flanking an exterior surface. The sheet topology is β6β9β7β8. The inner surface of the B domain β-sheet faces the large structure consisting of the N, A, and C domains. The B domain is joined to strand β10 of the A domain by loop 189−191 which forms the second hinge (BA) to the core N/A/C structure. The overall temperature factor of the B domain is 43.5 Å^2, about two times higher than the average for the rest of the molecule. There is a break in the electron density between strands β8 and β7 corresponding to the B loop (146−152). This conformational flexibility may play an important role in catalysis.

The A domain includes residues 192−329 and consists of a five-stranded antiparallel β-sheet (β10−β14) (Figures 2e and 3). The sheet topology is β12β11β10β13β14. The β-strands of the A domain are considerably longer than those in the other domains, containing 9−13 residues each. One side of one end of the A domain β-sheet faces the B domain β-sheet to form an open cleft. This cleft appears to define most of the enzyme's active site. Two α-helices (α10 and α11) which are connected only by Ile 251 (ϕ, ψ angles −116.4° and −60.4°) run along the solvent-exposed surface of the β-sheet.

The C domain includes residues 330−426 and is the smallest domain of the N/A/C core (Figures 2d and 3). It

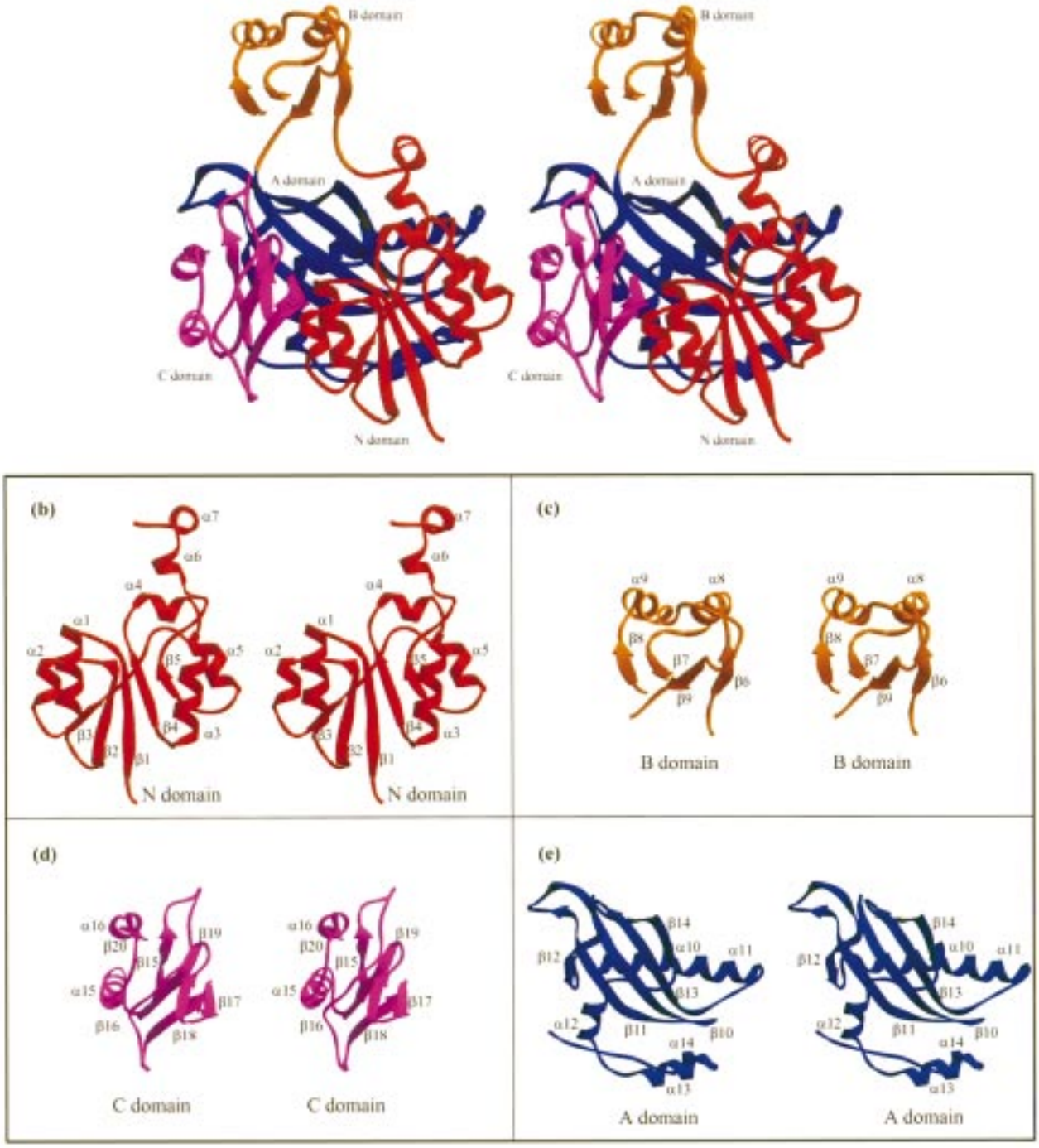


FIGURE 2: Ribbons presentation of GAR-syn structure. (a) overall molecule color coded by domain; (b−e) N domain, B domain, C domain, and A domains, respectively.

contains three strands of antiparallel sheet (β15, β16, and β19) plus a much shorter parallel strand (β20) to form a mixed four-stranded β-sheet. The sheet topology is β16β19β15β20. Two other β-strands (β17 and β18) are very short and cap the central sheet. Two α-helices (α15 and α16) flank the β-sheet on the solvent-exposed side. Comparison of GAR-syn with structurally similar proteins revealed that C domain is missing in DDLigase and GSHase while the N, A, and B domains are well conserved. The entire C domain composes one side of the central groove. As discussed subsequently, the C domain is proposed to play a key role in the putative interactions between GAR-syn and PRPP-AT and is consequential in substrate channeling.

DISCUSSION

Relationship among ATP-Grasp Superfamily Members. A search of the structures deposited in the Brookhaven Protein Data Bank (*38*) for structural homologues to GAR-syn was performed using the DALI algorithm (*39*). This search revealed that GAR-syn is structurally similar to BNC (*8*), DDLigase (*9*), GSHase (*10*), and SynC (*11*), despite low amino acid sequence identities (≤16% in the aligned regions, see Table 5). The computed backbone RMSD between GAR-syn and these four structures are within 3.8−5.0 Å without permutation of the secondary structure element sequence.

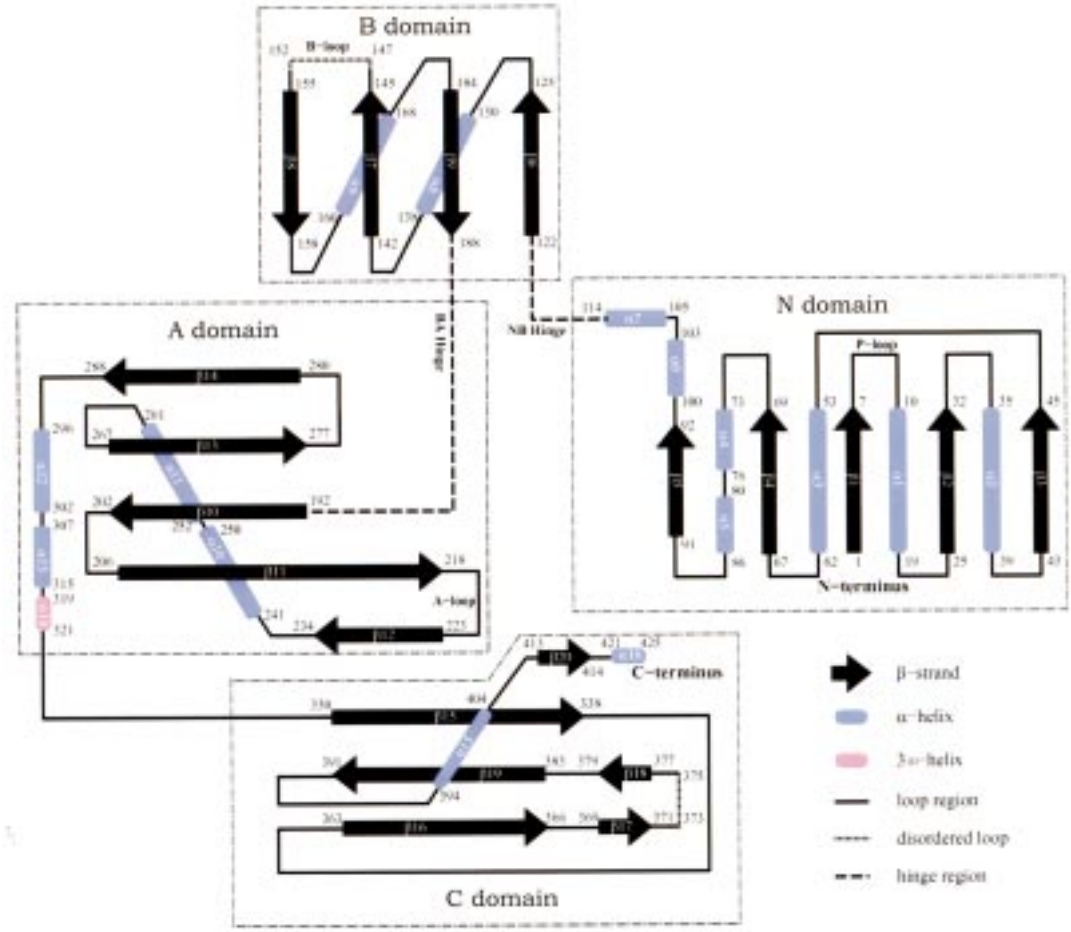


FIGURE 3: Topology plot of the GAR-syn structure. Residue numbers are placed on each end of secondary structure elements. Dashed lines indicate domain boundaries. Residue numbers close to dotted line defines the residue range of the disordered region (see Table 4).

Table 5: Best Alignments to the GAR-syn Structure Found by DALI Search

PDB id (chain)	Z[a]	RMSD[b] (Å)	LALI[c]	LSEQ[d]	%IDE[e]	protein	ref
				Overall structure search			
1bnc-A	26.8	5.0	380	433	16	biotin carboxylase (BNC)	*8*
2dln	16.0	3.8	250	306	16	D-alanine:D-alanine ligase (DDLigase)	*9*
1gsa	12.1	3.8	240	314	10	glutathione synthetase (GSHase)	*10*
1auv-A	9.9	4	205	292	14	synapsin Ia fragment (SynC)	*12*
				A/B domain search			
2dln	10.7	3.0	169	306	16	D-alanine:D-alanine ligase	*9*
1gsa	9.2	2.8	160	314	11	glutathione synthetase	*10*
1bnc-A	8.9	5.3	151	433	17	biotin carboxylase	*8*
1dik	7.6	3.8	155	869	12	pyruvate phosphate dikinase	*42*
1scu-B	7.4	3.5	138	388	21	Succinyl-coa synthetase	*43*
1auv-A	7.4	3.2	148	292	11	synapsin Ia fragment	*12*
				N domain search			
1bnc-A	11.6	2.8	113	433	15	biotin carboxylase	*8*
1uag	8.8	3.4	96	428	15	UDP-*N*-acetylmuramoyl-L-alanine:D-glutamate ligase (MurD)	*44*
1gar-A	7.3	2.6	88	204	9	glycinamide ribonucleotide transformylase (GART-N)	*45*
1ped-A	7.2	2.9	87	351	14	NADP-dependent alcohol dehydrogenase	*46*
2 cmd	7.1	3.8	98	312	17	malate dehydrogenase	*47*
2ohx-A	7.0	2.6	86	374	12	alcohol dehydrogenase, holo form (ADH)	*48*
2dln	7.0	2.3	98	306	10	D-alanine:D-alanine ligase	*9*

[a] Z-score, i.e., strength of structural similarity in standard deviations above expected. [b] Positional root mean square deviation of superimposed CA atoms in angstroms. [c] Total number of equivalenced residues. [d] Length of the entire chain of the equivalenced structure. [e] Percentage of sequence identity over equivalenced positions.

The structural similarity of GAR-syn to BNC, DDLigase, and GSHase confirms a mechanism- and sequence-based prediction that all four are members of the same enzyme superfamily (*7*, *40*, *41*). Each of these enzymes is proposed to use ATP to phosphorylate a carboxylate group in the acceptor substrate (Gly for GAR-syn, bicarbonate for BNC, D-alanine for DDLigase, and γ-glutamylcysteine for GSHase). The resulting labile acyl phosphate intermediate is then

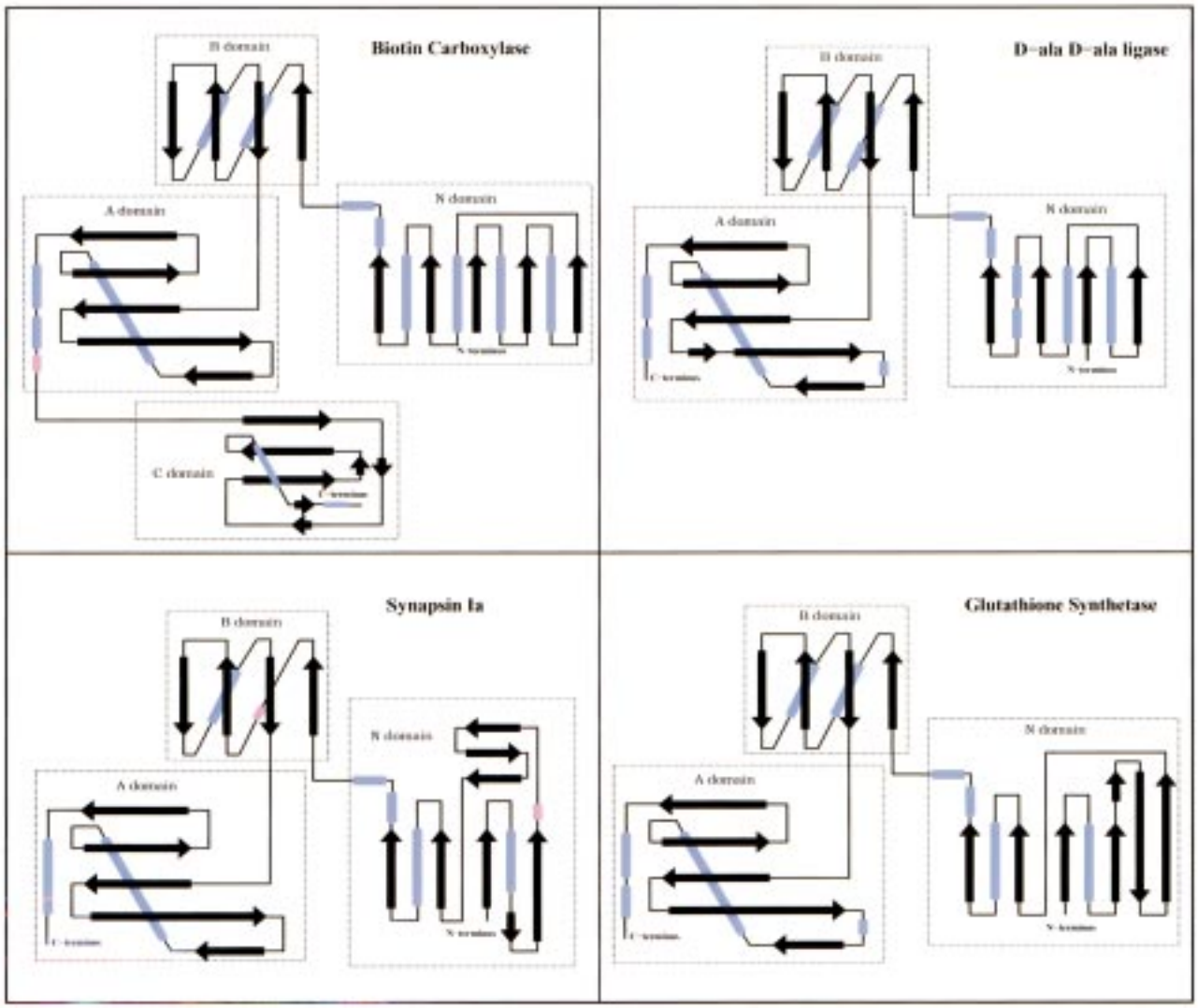


FIGURE 4: Comparison of topology of BNC, DDLigase, SynC, and GSHase. This figure is intended to show the topological similarity among these enzymes and GAR-syn, especially in A/B domain.

attacked by an amine in the donor substrate (GAR, biotin, D-Ala, and Gly, respectively) resulting in formation of an amide bond. This superfamily of enzymes recognizes a wide variety of donor and acceptor substrate pairs, but the mechanisms of the ligation reaction are proposed to be very similar. Crystallographic studies of these proteins indicated a remarkable resemblance in their overall topology (Figure 4), particularly in the ATP-binding region, which presents a conserved constellation of residues in the binding pocket despite low overall sequence conservation (see below). This ATP binding motif (which has been called an ATP-grasp motif because of its resemblance to a left-hand palming an ATP) is also found in a recent ADP-γ-S-bound structure of SynC (PDB code 1aux). An enzymatic function has not been identified for SynC and it is structurally distinct from the ADP-forming ligases in that it binds one Ca^{2+} instead of two Mg^{2+} ions.

All four ADP-forming ligase (GAR-syn, BNC, DDLigase, and GSHase) structures have nearly identical topology in the A and B domains and a conserved parallel β-sheet core in the N domain (Figure 4). Only GAR-syn and BNC have identical N domain folds, and furthermore, only these two members of the superfamily possess a C domain. In both cases, this domain is composed of an antiparallel β-sheet and a flanking α-helix. The A and B domains define the ATP-grasp motif in DDLigase, GSHase, and SynC. A second DALI database search using the A/B domains of GAR-syn showed that the ATP-binding enzymes pyruvate orthophosphate dikinase (*42*) and the B chain of succinyl-CoA synthetase (*43*) are structurally similar (RMSD $\leq$ 3 Å). No domains reminiscent of the N or C domains are present in these proteins. Structural homology in the A/B domains probably reflects the fact that ATP binding is a common feature of these enzymes.

The two hinge regions, 114−122 (NB) and 188−192 (BA), present in GAR-syn (Figure 3) are also observed in BNC, DDLigase, GSHase, and SynC. These two hinges are thought to provide flexibility that allows the B domain to move relative to the other domains. In our unliganded structure of GAR-syn, the cleft between the A and B domains is wider than in the structures of DDLigase, GSHase, and SynC complexed with adenine nucleotides (Figure 5). The unliganded BNC structure is wider still, and thus, the unliganded states appear to be in an open conformation, while the adenine nucleotide-bound structures appear to be in a more closed form. The closed-form structures exhibit the same separation of the B domain relative to the A domain. The open structures of the apoproteins show different degrees

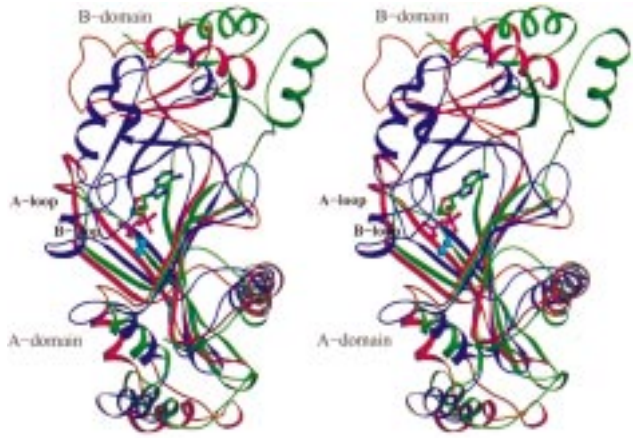


FIGURE 5: Consensus ATP-grasp motif comprised by A/B domain. Ribbons diagram shows A and B domain structural alignment of GAR-syn (red), BNC (green, PDB code 1BNC), and DDLigase complex (blue, PDB code 2DLN). The ball-and-stick model shows the ADP molecule bound to DDLigase; the Mg^{2+} ions are shown as blue balls. The structures are oriented such that the B domain is in the upper half and the A domain is in the lower half. The A domain β-sheet is well aligned among all molecules. The B domain is poorly aligned. In GAR-syn, the cleft between the A and B domains is open wider than that in DDLigase. In BNC, the B domain is even further away from A domain. GAR-syn and BNC are therefore defined as being in an open form, while DDLigase is in a closed form.

of opening. These differences may in part be accounted for by different crystal-packing phenomena. However, they also may be affected by the binding of ATP (Figure 5).

Figure 6 shows the amino acid sequence alignment for GAR-syn, DDLigase, BNC, and GSHase based strictly on structural alignments. The colored regions indicate those parts of the molecules that superimpose well and the domain to which those residues correspond. Highlighted residues are critical for ATP binding. As one might expect, these residues are 90% identical and 100% similar among all five sequences, despite the overall sequence dissimilarity.

Closed Form GAR-syn. By analogy with the available ATP-grasp superfamily structures, we expect the ATP-binding domain to be defined primarily by the A and B domains, particularly the β-sheets of these domains in GAR-syn. To model ATP-binding in GAR-syn, we generated a closed-form GAR-syn model using the DDLigase closed-form structure as a guide. The modeling procedure included two steps. The first step was to calculate a transformation matrix between the A domain of GAR-syn (*192*−*329*) and the A domain of DDLigase (*185*−*306*) using a least-squares comparison method available in the program O. This calculated matrix was then used to position all of DDLigase, including bound MgADP, onto GAR-syn. The second step was to calculate a transformation matrix between the transformed DDLigase B domain (*110*−*184*) and the GAR-syn B domain (*118*−*191*) and apply this matrix to the GAR-syn B domain to create a closed form GAR-syn.

MgADP Model. The X-ray structures of liganded DDLigase, GSHase, and SynC display almost identical conformations for base, ribose, and the α,β-phosphates of ADP. SynC is the only protein that provides structural information on the ATP-binding conformation. However, as noted above, this protein utilizes Ca^{2+}in place of Mg^{2+}. It is thus less likely that the conformation of ATP in this protein is similar to that required for the chemistry of the other members of the superfamily. Our modeling therefore focused on use of the available structural data on GSHase and DDLigase to generate a model for ADP binding shown in Figure 7.

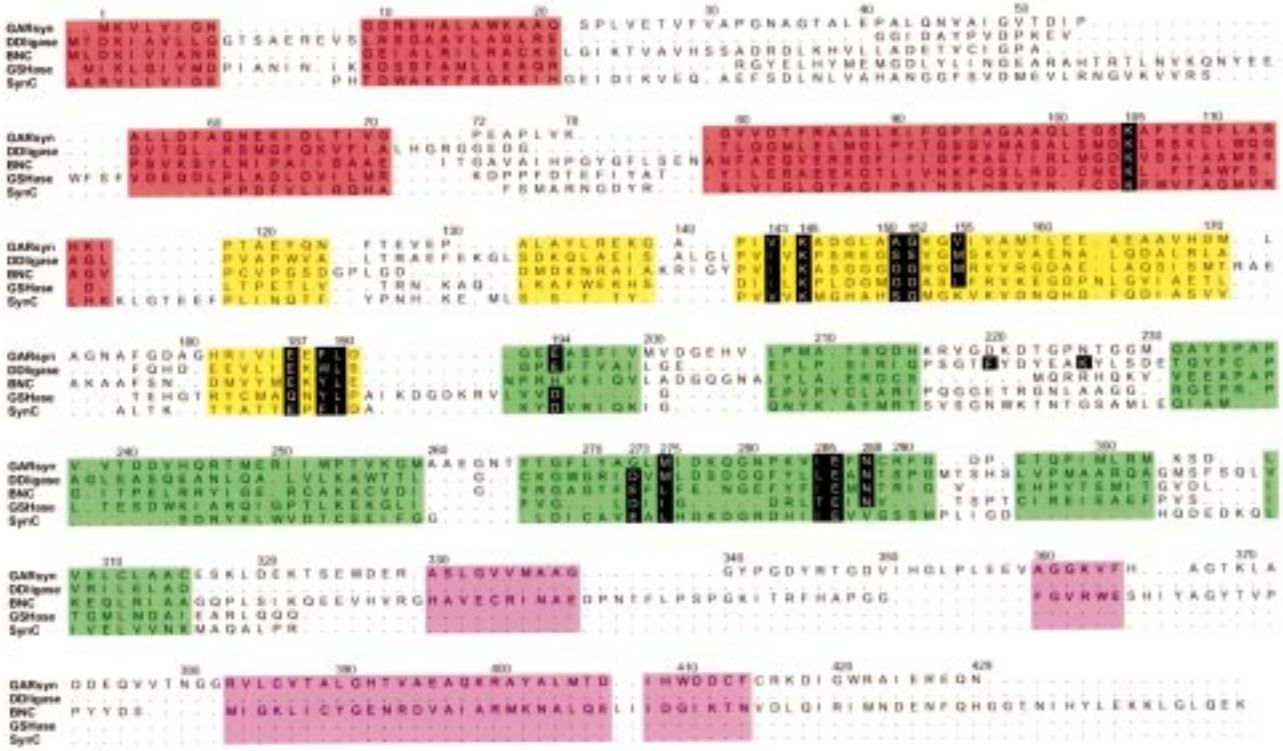


FIGURE 6: Structure-based sequence alignment. The color shaded regions are the segments that structurally align with each other. Color coding is red for the N domain, yellow for the B domain, green for the A domain, and magenta for the C domain. Residues highlighted in black are the ADP/ATP binding residues. The numbering scheme is for *E. coli* GAR-syn.

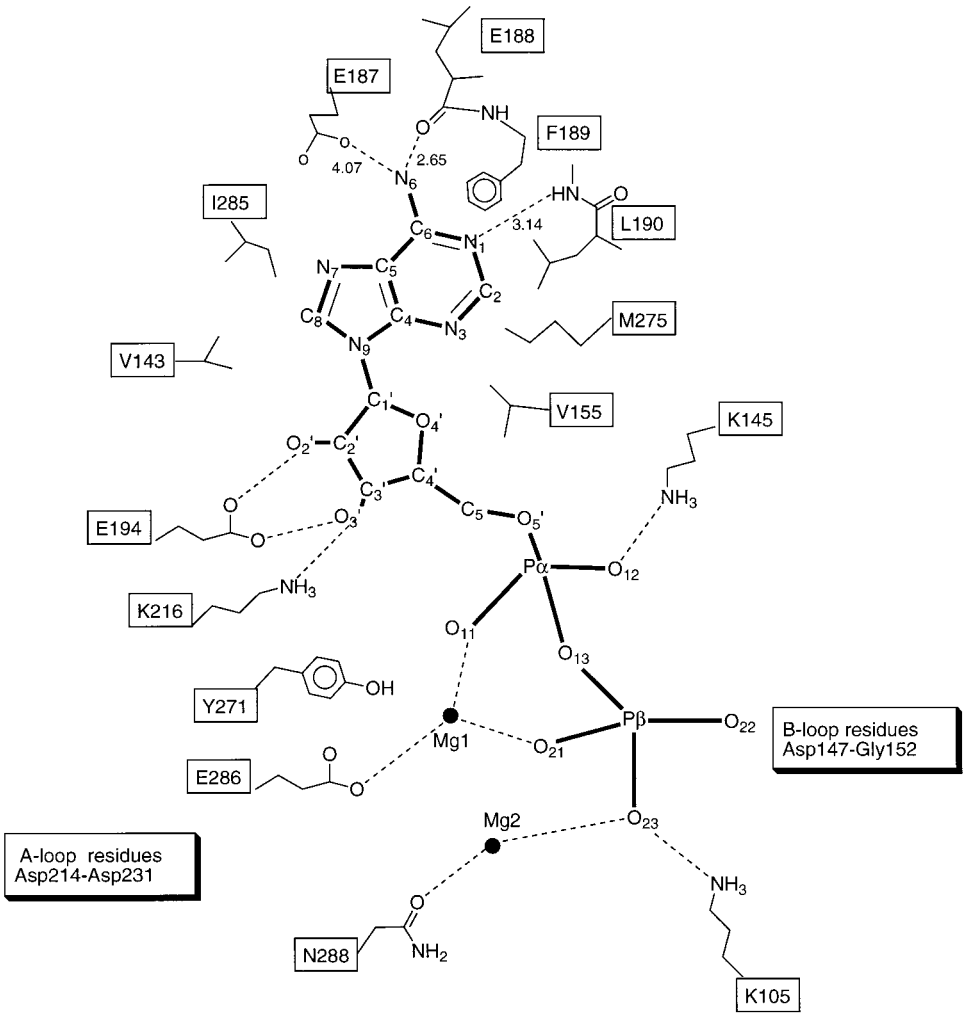


FIGURE 7: A model of the GAR-syn active site with MgADP binding based on DDLigase. The center structure with thick bonds represents ADP. The surrounding thinner structures represent protein active site residues. Dashed lines show possible hydrogen bonds. Mg1 and Mg2 are two magnesium ions. A-loop and B-loop residues are involved in interaction with the pyrophosphate. Their conformations are difficult to predict based on the current unliganded GAR-syn native structure.

ADP is sandwiched between the A and B domains of the closed form GAR-syn model. It binds to the face of the core β-sheets of these domains and lies in the direction that is parallel to the β-strands of the two sheets. The base-binding site is located deep inside the cleft and right next to the AB hinge connecting the A and B domains. The binding specificity for ATP is presumably derived from the interactions of the 6-amino group of adenine. The carboxyl group of the Glu187 side chain and the main-chain carbonyl oxygen of Glu188 are within hydrogen bond distance to this amino group and are proposed to function as hydrogen bond acceptors. Similar hydrogen bond geometry, with a side-chain carboxylate or carboxamide and a backbone carbonyl of the adjacent residue, is observed in DDLigase, GSHase, and SynC. The hydrophobic side chains of Phe189, Leu190, Met275, Ile285, and Val143 make van der Waals contacts with the purine moiety and are proposed to provide the remaining binding energy (Figure 7). Both the structure and the sequence (ExFL) of the adenine base recognition loop, 187−190, is highly conserved among the liganded structures of DDLigase, GSHase, and SynC.

The ribosyl group binding site involves both hydrogen bonding and hydrophobic interactions (Figure 7). In the closed form GAR-syn model, the Val155 side chain forms hydrophobic interactions with the β-face of the sugar ring, while the Glu194 side chain provides two hydrogen bond acceptors to the 2′- and 3′-hydroxyls of the ribosyl group. The Lys216 side chain is also located within hydrogen-bonding distance of 3′-hydroxyl and could serve as a hydrogen bond donor. These interactions resemble the ribose binding motif in DDLigase in which Met154 is the counterpart of GAR-syn's Val155 and Glu187 is the counterpart of GAR-syn's Glu194. In DDLigase, the Tyr210 amide backbone NH supplies a hydrogen bond donor to the 2′-hydroxyl.

ATP-utilizing enzymes including DDLigase, GSHase, and GAR-syn require Mg^{2+} for catalytic activity. Therefore, based on the two Mg^{2+}-binding sites observed in the structures of GSHase and DDLigase, we modeled Mg^{2+} into the GAR-syn closed-form structure. In the GAR-syn model, Mg1 is coordinated with two oxygen atoms from pyrophosphate and an oxygen from the side chain of Glu286. The hydroxyl group of Tyr271, which corresponds to Arg255 and Ile274 in DDLigase and GSHase, respectively, is close to Mg1, although tyrosine does not ordinarily interact with Mg^{2+} ions. Mg2 is coordinated with a β-phosphate oxygen and the side chain of Asn288. The structure and sequence of Glu286 and Asn288 are conserved in all members of this

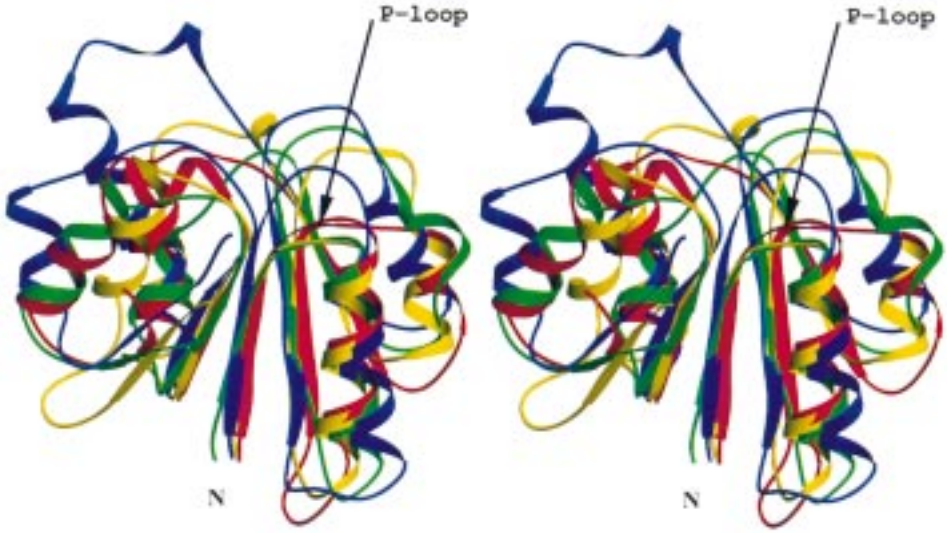


FIGURE 8: Superposition of N-terminal domains of GAR syn (red), GART-N (blue), ADH (yellow), and MurD (green). The least-squares alignment was performed for the first 94 residues of GAR syn and their counterparts in GART-N, ADH, and MurD.

superfamily except SynC. DDLigase utilizes Asp257, and GSHase utilizes Asp273 in Mg^{2+} binding; this position corresponds to Gly273 in GAR-syn. The remaining sites in the octahedral geometry for both Mg^{2+} ions could be occupied by water molecules or by atoms not apparent from the model.

The amino acid residues involved in diphosphate binding are less clear. Loop 214−231 is positioned near the ADP-binding site. However, the corresponding loop in the other superfamily members shows considerable variation. Likewise, the B loop (147−152) is positioned near the phosphate groups but cannot be modeled based on the available structures. Nevertheless, Lys105 and Lys145 are positioned near the ADP β- and α-phosphates, respectively. Lys105 and Lys145 correspond to Lys97 and Lys144, respectively, in DDLigase and are also conserved in BNC, GSHase, and SynC. In addition to these positively charged amino acids side chains, Mg1 and Mg2 are also involved in positioning the α- and β-phosphate groups, respectively.

Prediction of GAR Binding Geometry. Structural and sequence conservation is high in the A and B domains that comprise the ATP-grasp region of the ATP-dependent C−N ligase superfamily. Outside this region, the protein topology is also highly conserved in the N-terminal domain, which is involved in binding the pair of substrates that are ligated together. Structure-based sequence alignments show little conservation of sequence between enzymes in the superfamily, while a given enzyme from different organisms shows a high degree of sequence conservation. This is consistent with different binding motifs for each of the widely different pairs of substrates accepted by this ligase family.

A structure-based database search (DALI, Table 5) revealed that the N domain of GAR-syn adopts a Rossmann fold (*37*), which is common in nucleotide-binding proteins. The phosphate-binding loop (P-loop) corresponds to GAR-syn residues 9−11 (GGR) of the Rossmann fold. The Rossmann-fold motifs in GAR-syn, GAR transformylase (PurN, GART-N), and UDP-*N*-acetylmuramoyl-L-alanine: D-glutamate ligase (MurD) share identical topology, and all three lack the last β-strand and last α-helix found in the canonical ADH domain (*48*).

When the GAR-syn N domain is superimposed on the corresponding domains for GART-N, MurD, and ADH (Figure 8), the P-loops structurally align. This overlap suggests that the phosphate-binding site is defined by the comparatively small P-loop region. The GAR 5′-phosphate is bound to the Rossmann-fold motif in the GART-N complex structure (*45*). In GART-N, the phosphate of GAR is bound to the enzyme via hydrogen bonds to the backbone NHs of Gly11, Ser12, and Gly13. The Ser12 side-chain hydroxyl is also within hydrogen bond distance from phosphate oxygen. These three residues are located at the N-terminal end of a four-turn α-helix which stabilizes phosphate through the helix dipole. An alignment of GART-N sequences (10 of 11 sequences) indicates a SGx[G-[S/T]N hexapeptide motif in this loop region, where the underbars indicate the phosphate-binding residues. Similar phosphate binding is observed in the other liganded structures such as MurD (*44*) and ADH (*48*).

The remarkable similarity of the N-terminal domains of GAR-syn and the folate-dependent GART-N, the next enzyme in the purine pathway, was unexpected given the very different chemistry performed by these two enzymes. GART-N irreversibly converts GAR into FGAR, consuming 1 equiv of N^{10}-formyltetrahydrofolate. As there seemed to be no mechanistic reason for this structural conservation, we wondered whether it might indicate a conserved GAR-binding motif. On the basis of the structural homology with DDLigase and GSHase, we expected that the ATP would bind to the A/B domains and that the N domain of GAR-syn might be crucial for mononucleotide binding.

In the GAR-syn structure, the residues Gly9, Gly10, and Arg11 correspond to the P-loop phosphate-binding residues, and are also found in a region with a favorable helix dipole interaction. The GAR-syn heptapeptide consensus sequence G[S/N][G/A]REH is found in GAR-syn database sequences (20 of 24 sequences), with the putative phosphate-binding residues underlined. Our docking model of GAR is therefore based on the assumption that its phosphate group should bind to this loop as in the GAR/GART-N complex.

Placing GAR in this orientation along the inside face of the GAR-syn N domain positions the ribose ring near the

sequence VGPE (residues 70−73) between β4 and α4 (Figures 2 and 3). In this model, the glycyl moiety of GAR forms two hydrogen bonds with Glu73: one between the glycyl amino group and the Glu73 carboxyl group and the other between the carbonyl of the glycyl group and the amide backbone of Glu73. In GART-N, the analogous sequence LAGFM (residues 85−89) makes contacts to the ribose ring (Gly87 → C5′ and O) and the glycyl side chain of GAR. The other side of the glycyl side chain contacts the segment IHP in GART-N (residues 107−109), just after β5, corresponding to the GAR-syn sequence GPT (residues 93−95). However, this loop is significantly displaced relative to the corresponding loop in GART-N and does not contact the modeled GAR molecule. In the side-chain amine region, the bound folate replaces elements of secondary structure found in GAR-syn (notably α4) that might comprise a portion of the Gly site. The GAR-syn sequence FGDLE (residues 291−295) runs along the ribose portion of the substrate. The sequence FGDPE is invariant in all known GAR-syn sequences and is found in no other known protein sequence. This loop contains the spontaneous P → L mutation in the enzyme that was crystallized.

Further Conservation of Structure in the Purine Biosynthetic Pathway. Two additional purine biosynthetic enzymes were uncovered by sequence analysis of the ATP-grasp superfamily: PurT, the alternative, formate-dependent GART (GART-T), and PurK, the enzyme that utilizes HCO_3^-, ATP, and AIR to generate NCAIR, ADP, and Pi (NCAIR synthetase, eq 2). GART-T forms formyl phosphate, which is attacked by the primary amine N1 group of GAR (*49*), while PurK is proposed to form carboxyphosphate, which is attacked by the exocyclic primary amine in AIR (*50*). In addition to the chemical similarity of these ATP-dependent amide forming reactions, all three purine pathway enzymes recognize monophosphate substrates, for which they are quite specific. As might be anticipated, both enzymes have conserved sequences similar to the monophosphate-binding P-loop found in GART-N and proposed for GAR-syn. The sequences are GSGELG in GART-T (5 of 6 sequences; *E. coli* residues 28−36) and G[G/N]GQL[G/A] in PurK (18 of 19 sequences; *E. coli* residues 8−13). It is probable that all three purine enzymes use both an ATP-grasp motif and a Rossmann fold for nucleotide binding. Therefore, GAR-syn (and perhaps PurT and PurK) seems to have combined two different canonical nucleotide-binding motifs not previously identified in the same molecule.

Very recently, a crystal structure of the ATP-dependent purine pathway enzyme SAICAR synthase was reported (*12*). SAICAR synthase is also an acyl-phosphate-forming C−N ligase, but unlike the other three acyl-phosphate-forming purine pathway enzymes, it phosphorylates the carboxylate group in 4-carboxy-5-aminoimidazole ribonucleotide (CAIR), which is attacked by the amine group of Asp to form SAICAR. The structure of SAICAR synthase lacks the conserved ATP-binding sequences found in members of the ATP-grasp superfamily, but it contains an ATP-binding motif that is similar to the structure of the ATP-grasp motif, with antiparallel β-pleated sheets corresponding to the A/B domain β-sheet of GAR-syn and the other ATP-grasp enzymes. Although SAICAR synthase lacks a Rossmann fold, it preserves the P-loop for binding its monophosphate substrate CAIR.

The ATP grasp domain superfamily provides yet another example in which the chemistry may be driving evolution (*51*). The original protein-catalyzing formation of a phosphoanhydride intermediate could have been altered to accommodate a variety of carboxylate substrates and nucleophilic acceptors. In addition to the chemistry, the binding domain for the intermediate nucleotides in the purine biosynthetic pathway may also have been conserved. The ribose 5-phosphate binding site in PRPP-AT, GAR-syn, and GART-N all appear to contain a common P-loop with the 5′-phosphate being stabilized by the positive end of a helix dipole. Therefore, from the structural data available thus far, the purine pathway provides an example in which a substrate-binding motif may have been evolutionarily conserved. Both chemistry and substrate specificity thus may be involved in the evolution of a pathway.

Conformational Change with Substrate Binding. A critical structural transition in the ATP-grasp family of enzymes is correlated with the ordering of two flexible loops over the bound substrates, thus protecting the reactive acyl phosphate intermediates from hydrolysis (*10*). When these loops are modified by deletion or mutation, decreases in the catalytic activity and the efficiency of coupling of ATP hydrolysis to product formation are observed (*49, 52, 53*). In the two cases for which both liganded and unliganded structures of ATP-grasp enzymes are available, GSHase and SynC, the two loops undergo a disorder → order transition that is dependent upon nucleotide binding.

The first of these flexible loops (B loop), Figure 5 and Table 4, appears to interact directly with bound nucleotide substrates. The B loop bridges β7 and β8 in GAR-syn (residues 145−155; β6 and β7 and residues 146−153 in DDLigase) on the underside of the B domain, extending over the active-site saddle. In the DDLigase B loop, the backbone carbonyl of Glu148 and backbone NH of Ser151 hydrogen bind to the β-phosphate in ADP. In each ATP-grasp enzyme, this loop has a characteristic conserved sequence that does not match the Rossmann-fold consensus. As might be expected from residues involved in backbone contacts, the sequence conservation among enzymes is not high, although the ligating residues have small side chains and the whole sequence is glycine-rich. The B loop is disordered in the unliganded structures of GAR-syn, BNC, GSHase, and SynC and is well-defined in the nucleotide complexes of DDLigase, GSHase, and SynC.

The second of these loops (A loop) is part of the A domain, corresponding to 214−231 in GAR-syn. In GSHase and SynC, the A loop was found to be mobile in the absence of the nucleotide substrates and better defined in their presence. The length and sequence of the A loop varies, as does the conformation of the loop whose structures are clearly defined. The structure of SAICAR synthetase has also been solved both in the absence of substrate and (preliminarily) in the presence of ATP (*12*). It also contains a flexible loop that undergoes a disordered → ordered transition upon ATP binding. While SAICAR synthetase is not a classic ATP-grasp enzyme, its flexible loop occupies the same position as the A loop and appears to isolate the interior of the active site from solvent.

As with the B loop, the A loop in DDLigase and GSHase apparently becomes mobile when there is no substrate bound. This opens up one side of the active site, allowing substrates,

especially ATP/ADP, to pass by and bind at the active site. It therefore seems that the A loop is positioned to regulate ATP access to the active site. When the B loop is also disordered, the ATP binding site has even higher accessibility. When both A and B loops are ordered, they appear to contact each other and seal off the active site.

Interestingly, in both GAR-syn and BNC, the A loop is well ordered in the absence of nucleotides. The C domain of GAR-syn and BNC, which is immediately adjacent to the A loop, appears to immobilize this loop. The other enzymes that have been crystallographically characterized in an unbound state and found to have disordered A loops also lack C domains (Figure 4). All of the previously characterized ATP grasp enzymes are involved in multimeric complexes. It has been suggested that the BNC C domain utilizes its quaternary interactions in a manner not available to the smaller enzymes (*40*). *E. coli* GAR-syn is exclusively monomeric in solution (*54*). With an inflexible A loop in the unliganded state, GAR-syn and BNC structures presumably cannot use the motion of their A loops to regulate access to the ATP-binding site. Since neither of the B loops from these enzymes has been located, it is not possible to determine how the reactive intermediates will be isolated by the two loops during catalysis, or how ATP gains access to the active site. One intriguing possibility is that the two "hinge" segments of coil that attach the B domain to the main body of the protein provide the required flexibility. BNC and GAR-syn, which have otherwise very similar structures, show a large difference in the relative position of their A and B domains, with BNC having a much wider interdomain angle. This seems to arise from a distortion in the two B domain hinges and, while largely due to different crystal contacts in the two enzymes, provides evidence that such flexibility may exist.

Another potential route for the ATP may be a small hole between the second hinge and the A loop, which would provide a back door to the nucleotide-binding site. This hole may not be sealed off entirely during catalysis, as it seems to leave C2 and N3 of the base of ADP exposed to solvent in our model of the nucleotide-binding site.

Implications for Substrate Channeling between GAR-syn and PRPP-AT. As was mentioned in the introductory portion of this paper, GAR-syn is of special interest because it may be involved in channeling of PRA from PRPP-AT, the first enzyme in the purine pathway (eq 1). Kinetic studies in vitro with these two proteins measuring the production of GAR support this proposal (*2*). However, recent studies with *B. subtilis* AT and *E. coli* GAR-syn show kinetics of formation of GAR that are consistent with a free diffusion model predicted from the kinetic parameters of the individual enzymes (Qiao and Stubbe, unpublished results). Our interpretation of these observations is that channeling occurs between the two enzymes from the same organism. Thus, despite our inability to detect protein−protein interactions between these enzymes using conventional methods, kinetic studies suggest that a transient intermediate complex may be important in the transfer of the chemically unstable intermediate PRA (*2, 54*).

Given our structure of GAR-syn and the availability of structural information on both the *E. coli* and *B. subtilis* AT (*3, 4, 55*), we felt that we might be able to generate a docking model of the *E. coli* AT, a putative dimer, with the *E. coli* GAR-syn, a monomer. The requirements of this model would be that the active sites of each enzyme are accessible to one another through a solvent-inaccessible channel. This should further explain why channeling does not appear to occur between the *B. subtilis* AT and the *E. coli* GAR syn.

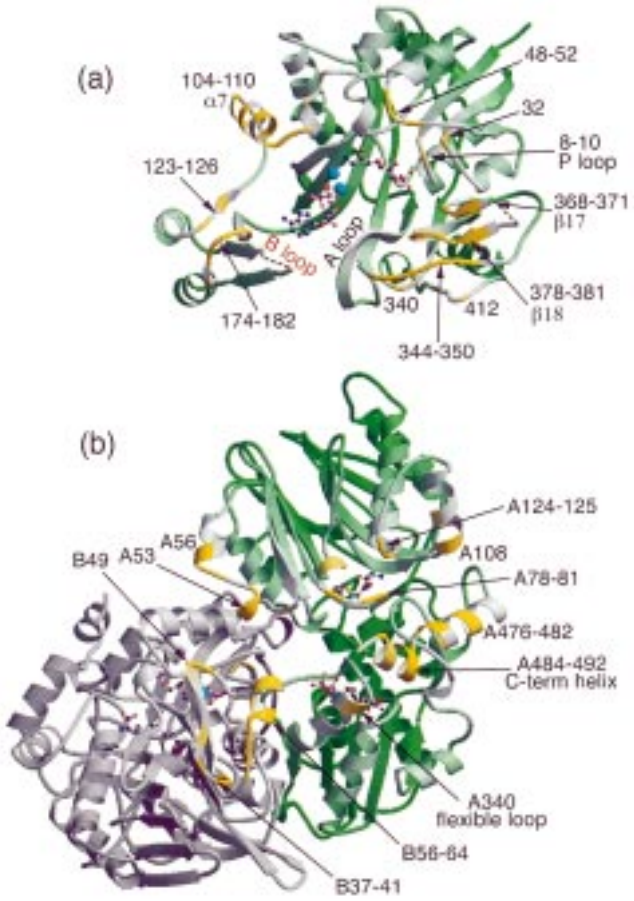


FIGURE 9: Docking model of the proposed channeling complex, identical to Figure 10 in orientation and coloring, indicating important secondary structural features. For GAR-syn (top) the residue numbers (or elements of secondary structure, Table 4) required to understand docking are presented. GAR, ADP, and two Mg atoms (shown as green spheres) are shown in their modeled positions on GAR-syn. For AT (bottom), the docking subunit A of dimeric AT is in blue and the nondocking subunit B is in gray, and appropriate chain and residue numbers are also given. DON and cPRPP are present in the AT. The cPRPP closest to GAR-syn in the docking model is underneath the labeled flexible loop in subunit A of AT.

The studies of Smith, Zalkin, and co-workers have played an important role in the generation of the docking model described below. They have recently published structures of the *E. coli* AT in both "open" and "closed" forms (*3, 4, 55*). Only the structure of an open form of *B. subtilis* AT has been reported (*5*). The overall structures and shapes of the *E. coli* and the *B. subtilis* ATs are quite similar in their open forms, not surprising given the 39% sequence identity (*5, 55*). However, the C-termini differ between the two proteins in that the *E. coli* enzyme contains a helix completely missing in the *B. subtilis* AT. The importance of this helix becomes apparent when the open and closed forms of the *E. coli* enzyme are compared. This helix is elongated in the open form. It becomes kinked in the closed form ("C term helix" in Figure 9) as it contacts a loop-and-helix structure ("flexible loop" in Figure 9) that folds over

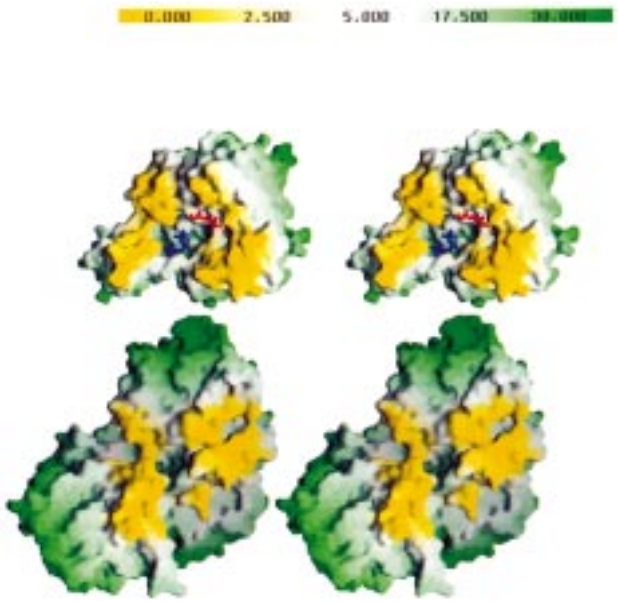


FIGURE 10: Docking model of proposed channeling complex between GAR-syn (top) and AT (bottom) in which the proteins have been pulled apart and rotated 90° so that the docking surfaces are revealed. For GAR-syn, the modeled positions of ADP (blue) and GAR (red) are shown. For AT, the cPRPP-binding site is at the bottom of the funnel. The surfaces are colored with respect to putative docking interactions, with yellow denoting surface contacts. In this model, there is one GAR-syn per dimeric AT.

the PRPP-binding site, isolating it from solution. In the open forms, this flexible loop is longer and contacts a different AT subunit. This flexible loop adopts different conformations in *E. coli* and *B. subtilis* AT (*4*). Smith and co-workers have suggested that this region may comprise a GAR-syn-binding surface because it appears in the presence of a bound nucleotide and it contains all of the major surface structure differences between *E. coli* and *B. subtilis* AT (*55*). Given the observed differences in structure for the open and closed forms and the potential of this transformation to provide access of PRA to solution, this region has been a major focus of our thinking about a docking mechanism.

In our docking models, described subsequently, the closed form of *E. coli* AT (bound to a carbocylic analogue of PRPP, cPRPP, and with the glutaminase site alkylated by diazo-oxonorleucine, DON) has been used, even though, in reality, a form intermediate between the open and closed forms may be more appropriate. The dimeric AT is crescent-shaped with approximate 2-fold symmetry between the subunits, which define a central cleft (Figures 9 and 10, bottom). The two cPRPP-binding domains are located at the bottom of this cleft, near the monomer interface, while the glutaminase domains are on the distal ends of each subunit (Figure 9). Both the glutaminase and nucleotide-binding active sites are isolated from solvent in the closed form AT and are linked to each other by a small channel, the proposed pathway for channeling of glutamine-derived ammonia. Krahn et al. have proposed (*55*) that the product PRA is released into the cleft by a distortion of the flexible loop/C-terminal helix structure.

For GAR-syn, only the structure of the open form (Figures 9 and 10, top) is available for docking. Given that the steady-state kinetic analysis reveals that PRA binds first to the enzyme, this is most reasonably the form of the protein that would dock with the AT (*6*).

While the enzyme stoichiometry of a transient channeling complex is unknown, one AT dimer interacting with one or two GAR-syn monomers is possible. We show the binding of only one GAR-syn monomer per AT dimer for simplicity. However, we favor docking orientations that allow the inclusion of a second equivalent of GAR-syn in the channeling complex.

In the docking process, shape and charge complementarity of the surfaces of the two proteins, location of their active sites relative to one another, and the differences between the *E. coli* and *B. subtilis* ATs were all considered. The two proteins were docked manually as rigid bodies, using InsightII and GRASP (*30*) to guide mating of the two surfaces. The surface regions of each enzyme expected to engage in protein−protein interactions are depicted in Figures 9 and 10. In these figures, regions of each protein predicted to be in direct contact during the docking interaction are colored yellow. Figures 9 and 10 show identical positioning of the proteins: the latter gives the surface representation, while the former gives an annotated ribbon diagram. The surfaces on AT and GAR-syn expected to come into contact with each other when PRA channels between the two active sites are electrostatically compatible.

While in principle there are many orientations that would allow the active sites to connect across a common cavity, we found one general orientation compatible with the channeling criteria and the surface features in both proteins. In this model (Figure 10), GAR-syn makes extensive contacts with the flexible regions of AT, including the kinked C-terminal helix, described above and first proposed by Krahn et al. (*55*). The surfaces surrounding the active-site cavities fit together well, sealing off a protected volume through which PRA might move. This is best seen in Figure 10, in which rings of yellow surround both the central active-site cavity in GAR-syn and the rim of the funnel that leads down to one of the cPRPP sites in dimeric AT.

In the docking model, GAR-syn binds asymmetrically to the top and one side of the AT dimer, with its B domain jutting into the central AT cleft. One side of the AT cleft is blocked by the main body of GAR-syn, leaving a symmetry-related site on the other side for a second GAR-syn to bind. The B domain is visible at the lower left of the GAR-syn model (Figures 9 and 10) extending away from the main body of the enzyme. Loops in this domain (Figure 9, GAR-syn residues 174−182, 123−126, and probably the disordered B loop) contact the glutaminase domains of both AT subunits near its molecular 2-fold axis (Figure 9, AT residues A53, A56, and B49). Above the B domain region in GAR-syn, a shallow saddle separates one side of the B domain from α7 in the N domain, which is near the AT subunit B glutaminase domain (B37−41 and B56−64).

The slightly convex surface of the GAR-syn C domain fits into a concave surface between the glutaminase and nucleotide domains of AT, forming the largest contact region (Figure 10, lower right side of GAR-syn, upper right side of AT). The outer edge of the GAR-syn C domain (340, 344−350, 412) contacts portions of the AT subunit A glutaminase domain (A78−81, A108, A124, and A125). GAR-syn β18 and two loops on the front of the C domain (344−350 and 378−381) contact the first part of the kinked C-terminal helix in the AT subunit A nucleotide-binding domain (A476−482).

A groove on the front face of GAR-syn between the N and C domains is large enough to accommodate the end of the AT kinked C-terminal helix (A484−A492). Contact regions in this groove are β17 (369−371) from the GAR-syn C domain and two loops extending from the top edge of the N domain β-sheet (ref *32*, and P loop residue 8). The end of the C-terminal helix of AT subunit A, which is bound on the inner face of the GAR-syn C domain, would not obscure the GAR-syn N domain active site. While the last 13 residues in AT (both subunits, residues 493−505) are not ordered in the closed form crystal structure, it seems unlikely that they would block the channel.

Finally, the N domain of GAR-syn is nearest the AT nucleotide-binding domain. GAR-syn α3 and the preceding loop (48−55) interact with the flexible loop/helix junction of AT (A340), which covers the bound cPRPP. In the docked complex, the flexible loop/helix region (residues A325−354) poses the only barrier between the cPRPP-binding site on AT and the predicted PRA-binding site on GAR-syn, which is ∼28 Å away. As anticipated by an examination of the structure of the closed form AT (*55*), this loop would have to move out of the way for PRA channeling to occur. If PRA-bound AT resembles closed form, cPRPP-bound AT, this might happen only after GAR-syn binds.

Our docking model indicates that the structure of GAR-syn is sterically and electrostatically compatible with the binding surface previously proposed to be important in AT (*55*). Generally, the linked active sites are capable of channeling PRA (*2*). Specific recognition of the AT C-terminal helix by GAR-syn would explain the observed differences between *B. subtilis* and *E. coli* AT channeling behavior. Additionally, the N domain of GAR-syn in the model is adjacent to the flexible loop region of AT that can adopt several conformations and appears poised to gate PRA release. While entirely hypothetical, the model gives a physical interpretation to the biochemical data (*2*) and provides a guide for experiments to determine which surface regions encode the observed channeling specificity.

ACKNOWLEDGMENT

Thanks to Joe Krahn and Janet Smith for sharing the closed form *E. coli* AT coordinates (PDB code 1ecc) prior to release.

REFERENCES

1. Ovadi, J. (1991) *J. Theor. Biol. 152*, 1−22.
2. Rudolph J., and Stubbe, J. (1995) *Biochemistry 34*, 2241−2250.
3. Kim, J. H., Krahn, J. M., Tomchick, D. R., Smith, J. L., and Zalkin, H. (1996) *J. Biol. Chem. 271*, 15549−15557.
4. Muchmore, C. R. A., Krahn, J. M., Hyun-Kim, J., Zalkin, H., and Smith, J. L. (1998) *Protein Sci. 7*, 39−51.
5. Smith, J. L., Zaluzec, E. J., Wery J. P., Niu, L., Switzer, R. L., Zalkin, H., and Satow, Y. (1994) *Science 264*, 1427−1433.
6. Cheng, Y. S., Rudolph, J., Stern, M., Stubbe, J., Flannigan, K. A., and Smith, J. J. (1990) *Biochemistry 29*, 218−227.
7. Galperin, M. Y., and Koonin, E. V. (1997) *Protein Sci. 6*, 2639−2643.
8. Waldrop, G. L., Rayment, I., and Holden, H. M. (1994) *Biochemistry 33*, 10249−10256.
9. Fan, C., Moews, P. C., Walsh, C. T., and Knox, J. R. (1994) *Science 266*, 439−443.
10. Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J., and Katsuube, Y. (1993) *J. Mol. Biol. 229*, 1083−1100.
11. Esser, L., Wang, C.-R., Hosaka, M., Smagula, C. S., Sudhof, T. C., and Deisenhofer, J. (1998) *EMBO J. 17*, 977−984.
12. Levdikov, V. M., Barynin, V. V., Grebemko, A. I., Melik-Adamyan, W. R., Lamzin, V. S., and Wilson, K. S. (1998) *Structure 6*, 363−376.
13. Weaver, T., Wang, W., and Ealick, S. E. (1998) *Acta Crystallogr., Sect. D* (In press).
14. Jancarik, J., and Kim, S. H. (1991) *J. Appl. Crystallogr. 24*, 409−411.
15. Hendrickson, W. A. (1991) *Science 254*, 51−58.
16. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
17. Leslie, A. G. W. (1992) Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, No. 26, Daresbury laboratory, Warrington U.K.
18. Evans, P. R. (1993) *Proc. CCP4 Study Weekend*, 114−122.
19. French G. S., and Wilson K. S. (1978) *Acta Crystallogr., Sect. A 34*, 517.
20. Otwinowski, Z., and Minor, W. (1996) *Methods Enzymol. 276*, 307−326.
21. Miller, R., Gallo, S. M., Khalak, H. G., and Weeks, C. W. (1994) *J. Appl. Crystallogr. 27*, 613−621.
22. Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. W., and Hauptman, H. A. (1993) *Science 259*, 1430−1433.
23. Blessing, R., personal communication.
24. Terwilliger, T. C., Kim, S.-H., and Eisenberg, D. (1987) *Acta Crystallogr., Sect. A 43*, 1−5.
25. Otwinowski, Z. (1993) *Proc. CCP4 Study Weekend*, 56−62.
26. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A 47*, 110−119.
27. Wang, B.-C. (1985) *Methods Enzymol. 115*, 90−112.
28. Read, R. J. (1986) *Acta Crystallogr., Sect. A 42*, 140−149.
29. Brunger, A. T. (1992) X-PLOR Version 3.1, A System for X-ray Crystallography and NMR, Yale University Press, New Haven and London.
30. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.
31. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
32. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol. 277*, 505−524.
33. Carson, M. (1991) *J. Appl. Crystallogr. 24*, 958−961.
34. Barton, G. J. (1993) *Protein Eng. 6*, 37−40.
35. Bairoch, A., Bucher, P., and Hofmann, K. (1997) *Nucleic Acids Res. 25*, 217−221.
36. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283.
37. Rossmann, M. G., Moras, D., and Olsen, K. W. (1974) *Nature 250*, 194−199.
38. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer. E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, Y., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
39. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
40. Artymiuk, P. J., Poirrette, A. R., Rice, D. W., and Willett (1996) *Nat. Struct. Biol. 3*, 128−132.
41. Fan, C., Moews, P. C., Shi, Y., Walsh, C. T., and Knox, J. R. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 1172−1176.
42. Herzberg, O., Chen, C. C. H., Kapadia, G., McGuire, M., Carroll, L. J., Noh, S. J., and Dunaway-Mariano, D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2652−2657.
43. Wolodko, W. T., Fraser, M. E., James, M. N. G., and Bridge, W. A. (1994) *J. Biol. Chem. 269*, 10883−10890.
44. Bertrand, J. A., Auger, G., Fanchon, E., Martin, L., Blanot, D., van Heijenoort, J., and Dideberg, O. (1997) *EMBO J. 16*, 3416−3425.
45. Almassy, R. J., Janson, C., A., Kan, C.-C., and Hostomska, Z. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6114−6118.
46. Korkhin, Y., Frolow, F., Bogin, O., Pereti, M., Kalb, A. J., and Burstein, Y. (1996) *Acta Crystallogr., Sect. D 52*, 882−886.
47. Hall, M. D., Levitt, D. G., and Banaszak, L. J. (1992) *J. Mol. Biol. 226*, 867−882.

48. Karadaghi, S. A, Cedergren-Zeppezauer, E. S., Petratos, K., Hovmoeller, S., Terry, H., Dauter, Z.; Wilson K. S. (1994) *Acta Crystallogr., Sect. D 50*, 793−807.
49. Marolewski, A. E., Mattia, K. M., Warren, M. S., and Benkovic, S. J. (1997) *Biochemistry 36*, 6709−6716.
50. Mueller, E. J., Meyer, E., Rudolph, J., Davisson, V. J., and Stubbe, J. (1994) *Biochemistry 33*, 2269−2278.
51. Babbitt, P. C., and Gerlt, J. A. (1997) *J. Biol. Chem. 272*, 30591−30594.
52. Tanaka, T., Yamaguchi, H., Kato, H., Nishioka, T., Katsube, Y., and Oda, J. (1993) *Biochemistry 32*, 12398−12404.
53. Kato, H., Tanaka, T., Yamaguchi, H., Hara, T., Nishioka, T., Katsube, Y., and Oda, J. (1994) *Biochemistry 33*, 4995−4999.
54. Cheng, Y. S., Murray, M., Schendel, F., Otvos, J., Wehrli, S., and Stubbe, J. (1987) *Adv. Enzyme Regul. 26*, 319−333.
55. Krahn, J. M., Kim, J. H., Burns, M. R., Parry, R. J., Zalkin, H., and Smith, J. L. (1997) *Biochemistry 36*, 11061−11068.

BI981405N