Archives of Biochemistry and Biophysics
Vol. 369, No. 1, September 1, pp. 30–41, 1999
Article ID abbi.1999.1344, available online at http://www.idealibrary.com on IDEAL®

Prediction of the Active-Site Structure and NAD^+ Binding in SQD1, a Protein Essential for Sulfolipid Biosynthesis in *Arabidopsis*[1]

Bernd Essigmann,[2] Brandon M. Hespenheide,[2] Leslie A. Kuhn,[3] and Christoph Benning
Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824-1319

Received March 11, 1999, and in revised form June 8, 1999

Sulfolipids of photosynthetic bacteria and plants are characterized by their unique sulfoquinovose headgroup, a derivative of glucose in which the 6-hydroxyl group is replaced by a sulfonate group. These sulfolipids have been discussed as promising anti-tumor and anti-HIV therapeutics based on their inhibition of DNA polymerase and reverse transcriptase. To study sulfolipid biosynthesis, in particular the formation of UDP-sulfoquinovose, we have combined computational modeling with biochemical methods. A database search was performed employing the derived amino acid sequence from *SQD1*, a gene involved in sulfolipid biosynthesis of *Arabidopsis thaliana*. This sequence shows high similarity to other sulfolipid biosynthetic proteins of different organisms and also to sugar nucleotide modifying enzymes, including UDP-glucose epimerase and dTDP-glucose dehydratase. Additional biochemical data on the purified SQD1 protein suggest that it is involved in the formation of UDP-sulfoquinovose, the first step of sulfolipid biosynthesis. To understand which aspects of epimerase catalysis may be shared by SQD1, we built a three-dimensional model of SQD1 using the 1.8 Å crystallographic structure of UDP-glucose 4-epimerase as a template. This model predicted an NAD^+ binding site, and the binding of NAD^+ was subsequently confirmed by enzymatic assay and mass spectrometry. The active-site interactions together with biochemical data provide the basis for proposing a reaction mechanism for UDP-sulfoquinovose formation.

Key Words: **lipid biosynthesis; epimerase; dehydratase; UDP-glucose; protein structure–function relationships; protein structural modeling; sulfoquinovosyl diacylglycerol.**

Sulfoquinovosyl diacylglycerol (SQDG)[4] is one of four glycerolipids that provide the bulk of the lipid matrix of thylakoid membranes. For this reason, SQDG is one of the most abundant sulfur-containing organic compounds and plays an important role in the global sulfur cycle (1). This anionic lipid is characterized by its unique sulfonic acid headgroup, 6-deoxy-6-sulfo-glucose (sulfoquinovose). The association of SQDG with photosynthetic membranes, together with the unusual structure of the headgroup, has led to much speculation about a role for SQDG in photosynthetic organisms (2). Recent results of the analysis of a sulfolipid null-mutant of the cyanobacterium *Synechococcus* sp. PCC7942 indicate that the loss of SQDG has no major effects on oxygenic photosynthesis (3). However, SQDG appears to be involved in maintaining a constant ratio of anionic lipids in the thylakoid membrane. Studies on the regulation of thylakoid lipid composition in *Arabidopsis thaliana* demonstrate that SQDG biosynthesis is enhanced and the level of phospholipids is reduced under phosphate limiting conditions. Moreover, there

[1] This work was supported in part by grants to C.B. from the Deutsche Forschungsgemeinschaft (BE 1591/1-2) and NSF and to L.K. from the Research Excellence Fund for Protein Structure, Function, and Design.

[2] Contributed equally to this work.

[3] To whom correspondence should be addressed. Fax: 517-353-9334. E-mail: kuhn@agua.bch.msu.edu. World Wide Web: http://www.bch.msu.edu/labs/kuhn.

[4] Abbreviations used: SQDG, sulfoquinovosyldiacylglycerol; APS, adenosine-5′-phosphosulfate; PG, phosphatidyl glycerol; PAPS, 3′-phosphoadenosine-5′-phosphosulfate; NCBI, National Center for Biotechnology Information; BLAST, Basic Local Alignment Search Tool; PDB, Protein Data Bank; HPLC, high-performance liquid chromatography; MALDI-MS, matrix-assisted laser desorption/ionization mass spectrometry; RMSD, root-mean-square deviation in atomic positions; DAG, diacylglycerol; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; yADH, yeast alcohol dehydrogenase.

0003-9861/99 $30.00

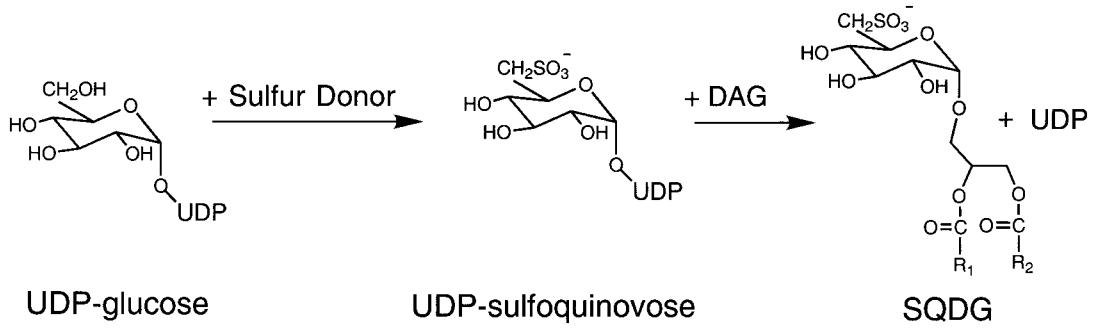

FIG. 1. Model of the sugar-nucleotide pathway for sulfolipid biosynthesis. In the last step, the UDP-sulfoquinovose moiety is transferred to diacylglycerol (DAG).

seems to be an inverse relationship between the concentrations of the two anionic lipids SQDG and phosphatidyl glycerol (PG), such that their sum remains constant (4). This suggests that SQDG substitutes for the anionic PG to maintain the function of the thylakoid membrane. Besides its role in the thylakoid membrane, it has been reported that SQDG is a potent inhibitor of eukaryotic DNA polymerase and HIV reverse transcriptase type 1 (5–7).

Despite the importance of SQDG, a genetic approach has only recently provided a breakthrough for studying its biosynthesis (8). In particular, little is known about the reaction mechanism responsible for the formation of the C-S bond in the sulfoquinovose moiety. A. A. Benson, who discovered SQDG in 1959, first suggested a sulfoglycolytic pathway, involving 3′-phosphoadenosine-5′-phosphosulfate (PAPS), sulfopyruvate, and other sulfonated analogs of glycolysis intermediates (9). Sulfopyruvate had been proposed to be converted via sulfolactaldehyde into 6-sulfoquinovose (10). Data in support of this pathway in plants were obtained by labeling experiments in alfalfa, which indicated that cysteic acid was incorporated into SQDG with higher efficiency than was sulfate (11).

An alternative pathway was suggested by Barber: The 6-sulfoquinovose in SQDG might be derived from an activated intermediate such as UDP-4-keto-5,6-glucoseen. This substance could serve as a sulfur acceptor and could be produced during the biosynthesis of 6-deoxyhexoses by sugar nucleotide hexose-4,6 dehydratases (12). Pugh *et al.* (13) observed the incorporation of UDP-[^{14}C]glucose into SQDG by broken pea chloroplasts and detailed a possible mechanism for the conversion of UDP-glucose into UDP-sulfoquinovose. Additionally, Pugh *et al.* reported a stimulatory effect of α-methyl glucosenide on sulfate incorporation into SQDG by chloroplasts (14).

Strong evidence for this alternative sugar nucleotide pathway is derived from the analysis of genes involved in sulfolipid biosynthesis (reviewed in 8). The functions of two genes, *sqdD* (15) and *sqdC* (16), involved in SQDG biosynthesis in *Rhodobacter sphaeroides* were determined by insertional mutagenesis. In these studies, naturally occurring UDP-sulfoquinovose was identified as an intermediate in SQDG biosynthesis (15). With the demonstration of UDP-sulfoquinovose accumulation in the *R. sphaeroides* null mutant and the characterization of UDP-sulfoquinovose:diacylglycerol sulfoquinovosyltransferase from spinach (17, 18), the last step of SQDG biosynthesis has been clarified (Fig. 1).

Recently, we isolated a cDNA from the higher plant *Arabidopsis* that displays high sequence similarity to the *sqdB* genes of bacteria (4). This *Arabidopsis SQD1* gene and the bacterial *sqdB* genes are the only sulfolipid genes known to be conserved between different organisms. Another notable finding is the amino acid sequence similarity between these enzymes and a diverse family of sugar nucleotide-modifying enzymes, including UDP-glucose epimerases and nucleotide-glucose dehydratases (19). These results, together with the observation that in bacterial *sqdB* null mutants no detectable sulfur-containing compound accumulates, led to the suggestion that SQD1 is involved in the first step of the biosynthetic pathway (Fig. 1), the formation of UDP-sulfoquinovose (8).

The high degree of sequence similarity between SQD1 and the glucose epimerases and dehydratases implies both functional and structural similarity (20). In the present study, we have constructed and analyzed a three-dimensional atomic model of SQD1 based on sequence similarity to the high-resolution crystallographic structure of UDP-glucose/galactose epimerase from *Escherichia coli* (21). This enzyme catalyzes the interconversion of UDP-glucose and UDP-galactose (22). Both SQD1 and UDP-glucose epimerase are also sequence-similar to *E. coli* dTDP-dehydratase, whose structure has recently become available (Protein Data Bank entry 1bxk, deposited by J. B. Thoden, A. D. Hegeman, P. A. Frey, and H. M. Holden). The structural model of the SQD1 active site has provided a basis for analyzing possible mechanistic similarities

between SQD1, epimerases, and dehydratases, and also predicted NAD^+ as a previously unknown cofactor for SQD1, which has been confirmed by biochemical and physical assays.

EXPERIMENTAL PROCEDURES

Database searching and sequence alignments. The National Center for Biotechnology Information (NCBI) nonredundant database was searched for sequences similar to SQD1 using BLAST (23) on the NCBI server (www.ncbi.nlm.nih.gov). The pairwise alignments with the epimerase and dehydratase amino acid sequences were performed with the program SIM (24) on the ExPASy Molecular Biology server (www.expasy.ch/sprot/sim-prot.html).

Molecular modeling. Swiss-Model (25), an automated protein modeling server (Glaxo Wellcome Experimental Research, Geneva; www.expasy.ch/swissmod/SWISS-MODEL.html), was used to obtain a preliminary three-dimensional structural model of SQD1 using a series of known UDP-glucose 4-epimerase structures from *E. coli.* The complete SQD1 amino acid sequence was submitted to Swiss-Model; however, due to small gaps in the sequence alignment of SQD1 with the known structure, an incomplete model was returned. To overcome this problem, seven short regions which did not align with the UDP-glucose epimerase amino acid sequence were deleted from the SQD1 sequence. The modified sequence was submitted to Swiss-Model, which then returned a complete structural model for SQD1 based on the structure of *E. coli* epimerase. All subsequent manipulation of protein structures was performed with the molecular modeling software Insight II (Molecular Simulations, Inc.).

The structural model of SQD1 was superimposed onto the structure of UDP-glucose epimerase in complex with NAD^+ and UDP-glucose (21), entry 1xel in the Brookhaven Protein Data Bank (Refs. 26 and 27; www.pdb.bnl.gov). NAD^+ was modeled in the *syn* conformation present in 1xel, consistent with the observation that this is the active conformation in the epimerase (28) and places the NAD^+ in the proper position for a B-side specific hydride transfer from the glucose C-4 to the nicotinamide C-4 (21, 28). The residues that bind NAD^+ and UDP-glucose in 1xel were identified using Ligand–Protein Contacts analysis (29) and were used to create an active-site template. The corresponding active-site residues in SQD1 were found by visual analysis of the superimposed structures and modeled in the same conformation as in the epimerase template. Side chains which differed between the two proteins were modeled in a conformation which optimized both structural similarity and chemical interactions between the protein and NAD^+.

The stereochemical quality of the SQD1 model, including (Φ, Ψ) angles, van der Waals interactions, and nonproline *cis* peptide bonds, was validated using the program Procheck (30, 31) through the stages of model building. All structural problems reported by Procheck involving residues near the active site were resolved with minor modifications to the side-chain dihedral χ angles of the residues. Two of the seven regions excluded from the original Swiss-Model input (residues 143–148 and 364–374) were added to the model using the search-loop-protein routine of Insight II, followed by side-chain torsional angle optimization to alleviate steric overlaps. The active site residues were energy minimized using 100 steps of backbone-constrained steepest-descent minimization with the cff91 force field using the Discover 3 module of Insight II, in order to alleviate any remaining unfavorable interactions. The program Dali (Ref. 47; www2.ebi.ac.uk/dali/dali.html) was used to optimally superimpose the main chains of UDP-glucose epimerase and dTDP-glucose dehydratase onto the SQD1 structural model for comparison.

Cloning, expression in E. coli, and purification of recombinant SQD1. The cloning into the pQE30 vector (Qiagen) and expression and purification of SQD1 were as previously described (4). Protein from control cultures containing the pQE30 vector with no insert, pQE30 containing the SQD1 coding region, and the purified SQD1 protein were boiled in sample buffer and run on a 12% SDS–polyacrylamide gel (32) for 1.5 h at 100 V using the Mighty Small II gel system (Hoefer).

NAD^+ enzyme assay. For quantitative determination of NAD^+, an alcohol dehydrogenase-based enzyme assay was carried out. Yeast alcohol dehydrogenase (yADH) was obtained from Sigma Chemical Co. The reaction buffer contained 50 mM Tris–HCl buffer, pH 8.5, 0.2 M ethanol, 1 mM EDTA, 10 units yADH ml^{-1}, and an aliquot of 50 μl of the supernatant of the denatured SQD1 protein. The total reaction volume was 1 ml. The reaction was initiated by the addition of yADH, and the absorbance was recorded at 340 nm against a blank cuvette containing the same constituents except for the yADH. All assays were performed at 25°C in an Uvicon 930 double-beam spectrophotometer (Kontron Instruments).

HPLC analysis. The analysis of nucleotides, including NAD^+, was accomplished by reverse-phase high-performance liquid chromatography (HPLC) using a 4.6 × 250-mm RP-18 column (Ultrasphere from Beckman) with a particle size of 5 μm. For the mobile phase, a mixture was used of 30 mM KH_2PO_4 and 2 mM tetrabutylammonium hydroxide, adjusted to pH 6.0 with KOH (solvent A) and acetonitrile (solvent B). Gradient elution was performed at a flow rate of 1 ml min^{-1} for the first 30 min, using a linear gradient from 0 to 30% of solvent B in A. Elution was continued 5 min longer with this solvent ratio. After 35 min, the column was washed with 100% solvent B for 10 min, followed by equilibration with 100% solvent A for 20 min.

Mass spectrometry. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) was carried out on a PerSeptive Biosystems Voyager Elite instrument equipped with a nitrogen laser emitting at 337 nm with a pulse duration of 2 ns. A saturated solution of 2,5-dihydroxybenzoic acid dissolved in acetonitrile:water (1:1, by volume) was used as the MALDI matrix. To prepare the target, 1 μl of the sample was mixed with 1 μl of matrix solution on the sample plate cell and air-dried. The instrument was operated in the positive ion mode.

RESULTS

Similarity of SQD1 to epimerases and dehydratases. A search of the nonredundant translated GenBank database (NCBI) using the derived amino acid sequence of the *SQD1* gene from *A. thaliana* revealed considerable sequence similarity to sugar nucleotide modifying enzymes (Fig. 2A). The first 84 amino acids of the SQD1 protein are absent in the orthologous bacterial SQDB proteins and in epimerase and dehydratase sequences from different organisms, and are believed to encode for a transit sequence targeting SQD1 to the chloroplast. The presence of a chloroplast transit sequence in SQD1 has been shown by *in vitro* translation and chloroplast import experiments (4). The sequence identity between SQD1 and the bacterial SQDB proteins varied from 40 to 68%. The sequence identity to the next most closely related proteins, UDP-glucose 4-epimerases and dTDP-glucose 2,4-dehydratases, approaches 28% depending on the organism. Closer comparison of the N- and C-terminal sequences of SQD1 with those of the epimerase and dehydratase enzymes from *Arabidopsis* showed that the region of highest sequence identity, 26%, to the epimerase is located in the N-terminal portion of SQD1 (Fig. 2B). In contrast, the highest sequence similarity to the

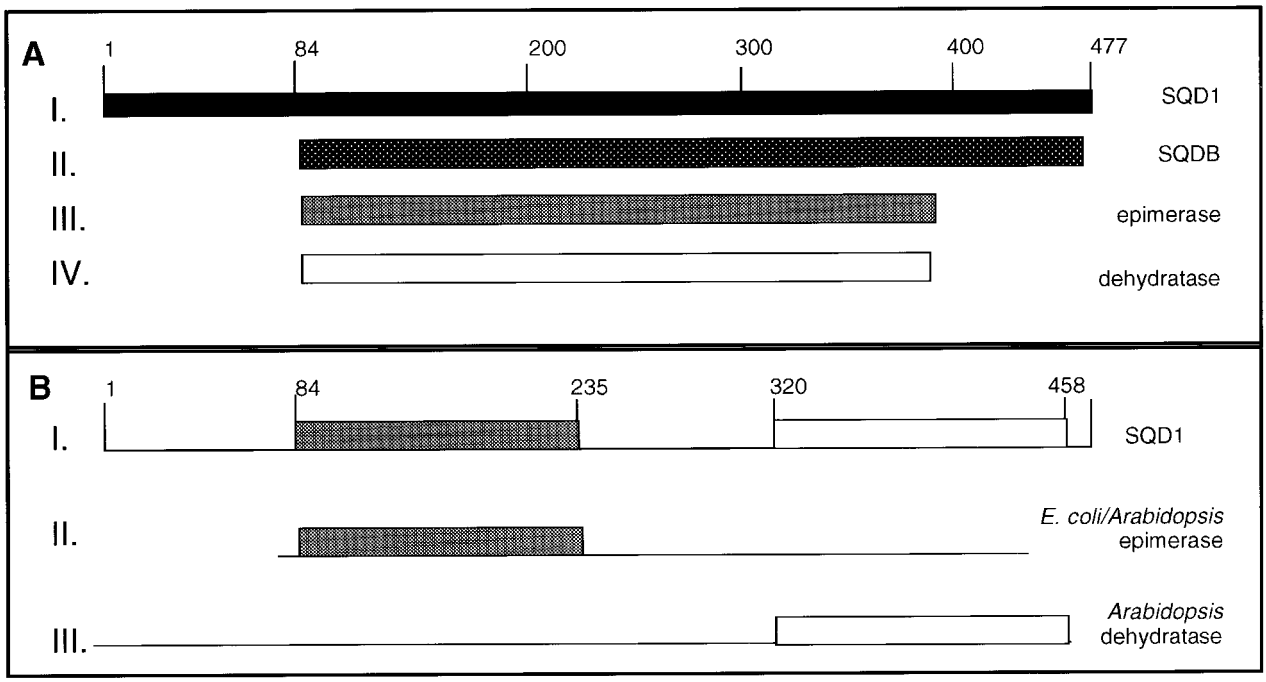

FIG. 2. Summary of SQD1 sequence similarity. (A) Comparison of the amino acid sequence predicted from the SQD1 cDNA of *A. thaliana* and the most similar amino acid sequences from the nonredundant database (using the SIM sequence alignment program with BLOSUM62 matrix). (I) SQD1 sequence. (II) Region of highest sequence identity between SQD1 and bacterial SQDB proteins (68% identity for *Synechocystis* sp., 41% for *R. sphaeroides*, and 40% *Synechococcus* sp.). (III) Region of similarity between SQD1 and epimerases from diverse organisms (28% identity with UDP-glucose epimerase from *Pasteurella haemolytica*, 27% with *Aquifex aeolicus*, 27% with *Saccharomyces cerevisiae*, 27% with *Neisseria meningitidis*, 26% with *Haemophilus influenzae*, 26% with *Yersinia enterocolitica*, 26% with *Lactobacillus casei*, 26% with *Kluyveromyces lactis*, 25% with *Bacillus subtilis*, and 27% identity with CDP-tyvelose epimerase from *Yersinia pseudotuberculosis*). (IV) Region of SQD1 identity with dehydratases from diverse organisms (26% for dTDP-glucose 4,6-dehydratase from *Salmonella typhimurium*, 26% for *E. coli*, 25% for *Rhizobium* sp. NGR234, 25% for *Haemophilus influenzae*, and 24% for *Neisseria meningitidis*). (B) Comparison of regions of the amino acid sequence of SQD1 from *Arabidopsis* with the amino acid sequences of *E. coli* and *Arabidopsis* UDP-glucose epimerase as well as the *Arabidopsis* dTDP-glucose 4,6-dehydratase (using the SIM sequence alignment program with BLOSUM30 matrix). (I) SQD1 sequence regions showing the highest similarity to epimerases and dehydratases. (II) Region of the *E. coli* and *Arabidopsis* UDP-glucose epimerases showing 26% identity to SQD1. The identity over the entire sequence (residues 84–469) is 23% for *E. coli* and 22% for *Arabidopsis* epimerase. (III) Region in the *Arabidopsis* dTDP-glucose 4,6-dehydratase showing 32% identity to SQD1 (identity for this region in *E. coli* dehydratase, 21%). The identity over the entire *Arabidopis* sequence (residues 84–469) is 25% (23% for *E. coli*).

dehydratase, 32%, is located in the C-terminal domain. These results indicate that SQD1 is similar to both types of enzymes, which are in turn paralogs (27% sequence identity). A study of more than 300 protein structures representing a range in sequence identity showed that 24.8% or greater sequence identity over 80 or more residues between two proteins indicates that the two proteins are structurally homologous (have main chains that superimpose to within 2.5 Å RMSD in C_α positions; Ref. 20). Given the >24% identity over 355 residues between SQD1 and the *E. coli* UDP-glucose epimerase (29% identity over the 164 N-terminal residues), SQD1 and the epimerase are structural homologs.

Structural model of SQD1. A three-dimensional structural model of the *Arabidopsis* SQD1 protein was constructed based on the amino acid sequence alignment and identity with the known structure of *E. coli* UDP-glucose epimerase (Fig. 3), to evaluate the functional consequences of their similarity. The UDP-glucose epimerase structure is in complex with UDP-glucose and NAD^+, so the analysis focused on the conservation of structure and chemistry in the ligand-binding sites and their relevance to SQD1 function. A (Φ, Ψ) plot of the final structural model of SQD1 is shown in Fig. 4; all of the residues fall into sterically favored regions of conformational space. Furthermore, the software package Procheck (30, 31), used to validate the stereochemical quality of crystallographic structures, gave an overall quality factor (G-factor) for the SQD1 structural model within the range expected for a 1.8 Å resolution structure.

Analysis of the potential NAD^+ binding site. We refined the NAD^+ binding site of the SQD1 model using a template derived from the 1.8 Å resolution crystallographic structure available for UDP-glucose epimerase, PDB code 1xel. In the epimerase, 31 residues are within 5.0 Å of NAD^+ (Table I). Of the corresponding residues in SQD1, 42% displayed exact identity to the

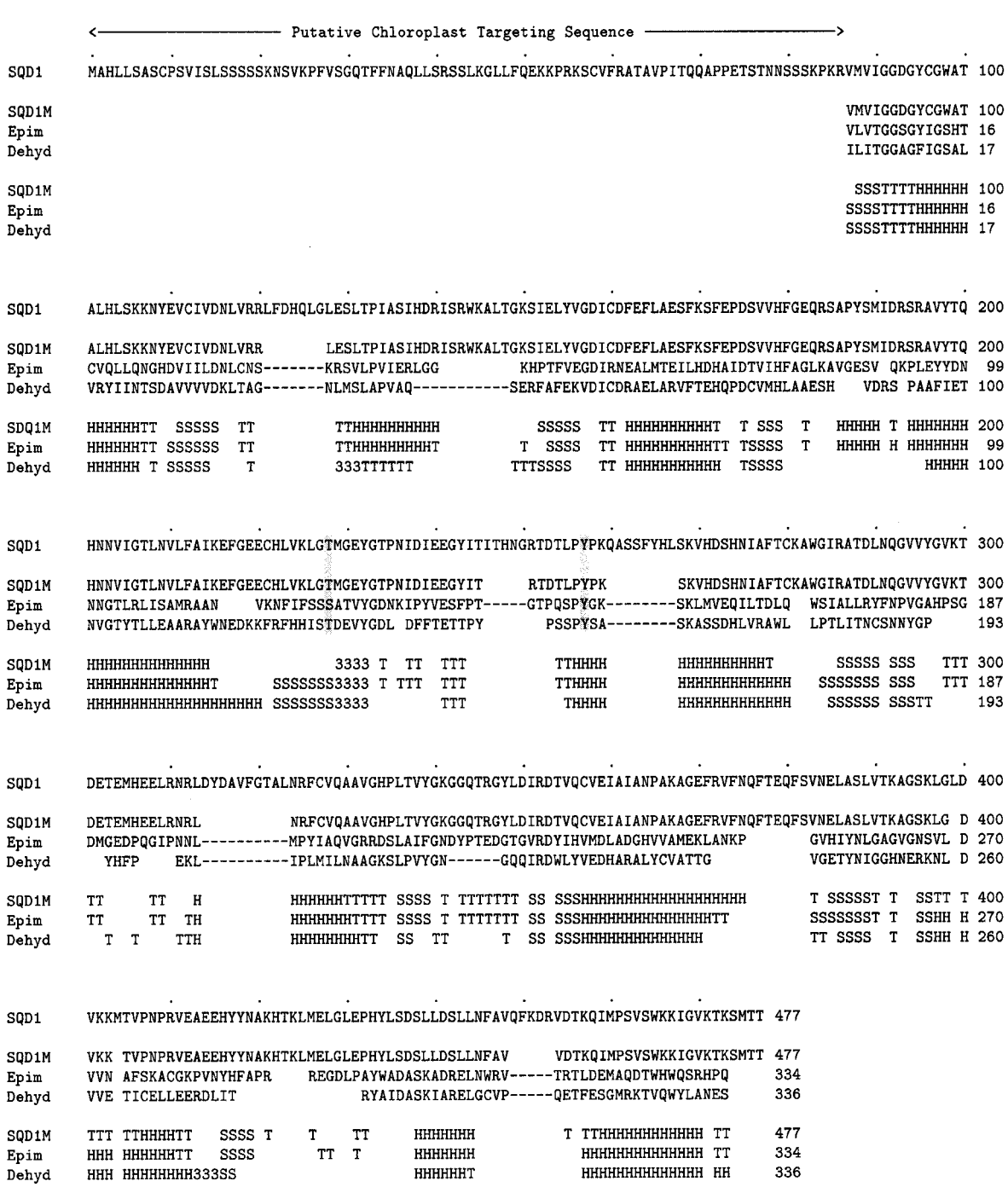

FIG. 3. Sequence alignment for SQD1, the three-dimensional modeled region of SQD1 (SQD1M), the *E. coli* UDP-glucose epimerase upon which the model was based (Epim), and *E. coli* dTDP-glucose dehydratase (Dehyd), with secondary structural annotations for each. SQD1 numbering begins at the putative chloroplast targeting sequence, and numbering for the epimerase and dehydratase sequences is from PDB entries 1xel and 1bxk. Two catalytically essential residues (Fig. 9B) in the epimerase (39) have conserved side-chain chemistry in the three proteins and are shaded in grey. There are two possibilities for aligning residues between 246 and 267 in SQD1: either RTDTLPYPK or KQASSFYHL in SQD1 aligns reasonably with GTPQSPYGK in the epimerase, in both cases conserving a tyrosine residue two positions from the end. Thus, either Tyr257 or 265 in SQD1 corresponds to the catalytically essential Tyr149 residue in the epimerase. The sequence alignment shown is based upon Dali structural alignment of the epimerase and dehydratase structures onto SQD1M (47). Secondary structures of the aligned residues were assigned by DSSP (48) based on their structural coordinates and are annotated below the amino acid sequences; H indicates helix; S, strand; T, reverse turn; 3, 3(10) helix; and blank, no regular secondary structure. For SQD1M, spaces indicate regions with low sequence identity to the epimerase that were not modeled; spaces in the epimerase and dehydratase sequences indicate regions that did not align structurally with SQD1. Dashes in the epimerase and dehydratase sequences connect residues that are contiguous in the protein structure. Altogether, 326 residues aligned between the SQD1M and epimerase structures, with 0.9 Å C_α RMSD, and 293 residues aligned between SQD1M and the dehydratase structure (2.3 Å RMSD).

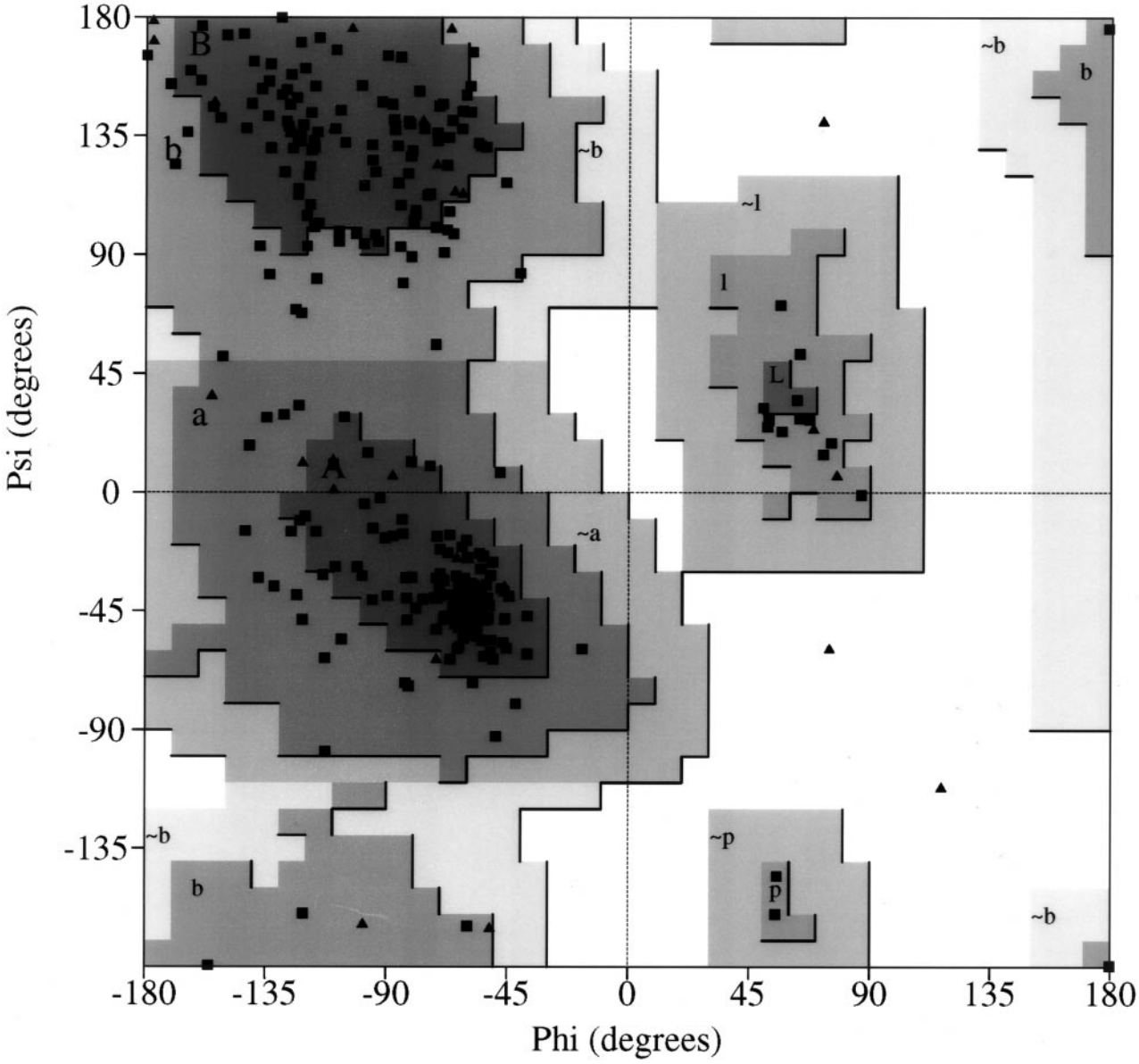

FIG. 4. Main-chain (Φ, Ψ) angle plot for the SQD1 model. All nonglycine residues (filled squares) fall into most favorable (dark grey) to favorable (medium grey) regions of the plot; glycine residues, whose (Φ, Ψ) angles are sterically unconstrained, appear as filled triangles. The distribution was calculated and plotted using Procheck (30, 31).

epimerase, and an additional 13% were highly conservative substitutions. Two of the remaining residues interact with NAD^+ via main-chain atoms. In fact, this Rossmann fold (33) associated with nucleotide binding is a canonical example of structural and functional conservation despite weak sequence conservation; there is no specific sequence pattern in the Prosite database (expasy.hcuge.ch/sprot/prosite.html) for identifying NAD^+ binding sites or Rossmann folds, and thus the NAD^+ site was not predicted from earlier sequence analysis.

A backbone superposition (RMSD, 0.33 Å) of the NAD^+ binding residues of the SQD1 model with those of the *E. coli* epimerase (Fig. 5B, Table I) allowed us to evaluate the potential of SQD1 to bind the cofactor. The mode of NAD^+ binding in the orthologous dTDP-dehydratase structure (PDB entry 1bxk; depositors Thoden, Hegeman, Frey, and Holden) was found to be virtually identical to that in the epimerase. SQD1 showed side-chain chemistry and hydrogen-bonding patterns in the NAD^+ site similar to those in the epimerase (PDB entry 1xel (21)). All but one of the hydrogen

TABLE I

Corresponding Residues in the Ligand Binding Sites of the *Arabidopsis* SQD1 and *E. coli* Epimerase Structures

NAD^+ site		UDP-glucose site	
Epimerase	SQD1	Epimerase	SQD1
GLY:7	**GLY:**91	LYS:84	ARG:184
GLY:10	**GLY:**94	VAL:86	ALA:186
TYR:11	**TYR:**95	**THR:**124	**THR:**228
ILE:12	CYS:96	ALA:125	MET:229
GLY:13	**GLY:**97	THR:126	GLY:230
LEU:30	VAL:114	**TYR:**149	**TYR:**257
ASP:31	**ASP:**115	TYR:177	LEU:290
ASN:32	**ASN:**116	PHE:178	ASN:291
CYS:34	VAL:118	ASN:179	GLN:292
ASN:35	ARG:119	TYR:203	PHE:326
GLY:57	**GLY:**157	**LEU:**215	**LEU:**336
ASP:58	**ASP:**158	ALA:216	THR:337
ILE:59	**ILE:**159	ILE:217	VAL:338
ARG:60	CYS:160	PHE:218	TYR:339
PHE:80	**PHE:**180	GLY:229	ASP:350
ALA:81	GLY:181	**ARG:**231	**ARG:**352
GLY:82	GLU:182	TYR:233	THR:354
LEU:83	GLN:183	VAL:269	LEU:397
LYS:84	ARG:184	ASP:295	GLY:420
ASN:99	GLN:200	**LEU:**296	**LEU:**430
SER:122	LEU:226	TYR:299	HIS:433
SER:123	GLY:227		
SER:124	THR:228		
TYR:149	**TYR:**257		
LYS:153	**LYS:**269		
TYR:177	LEU:290		
PHE:178	ASN:291		
PRO:180	GLY:293		
GLU:191	**GLU:**304		
ASN:199	ARG:312		

Note. Within the NAD^+ binding site, 42% of the residues are identical (boldface) and an additional 13% are highly conservative substitutions. 24% of the residues contacting UDP-glucose are identical, and an additional 29% are conservative substitutions. SQD1 makes all but one of the hydrogen bonds observed between the epimerase and each ligand, and contributes one new hydrogen bond for each.

bonds made by the epimerase to NAD^+ were reproduced in the SQD1 model, and SQD1 forms an additional hydrogen bond via Arg119. An examination for van der Waals overlaps between atoms of NAD^+ and SQD1 showed no steric hindrance, indicating that NAD^+ can adopt the same *syn* conformation in SQD1 as in the epimerase (28).

Analysis of the potential UDP-glucose binding site. As with the NAD^+ binding site, potential UDP-glucose binding residues within SQD1 were identified by analyzing interactions with UDP-glucose positioned according to the epimerase complex. (This site is preserved in the dTDP-glucose dehydratase structure.) There were 21 residues located within 5.0 Å of any UDP-glucose atom (Table I), and a backbone superposition of the SQD1 residues binding UDP-glucose with corresponding residues in the epimerase (RMSD, 0.31 Å; Fig. 5C) shows that SQD1 reproduces 9 of the 10 hydrogen bonds the epimerase makes to UDP-glucose, and there is one additional hydrogen bond not found in the epimerase. The conformation of UDP-glucose from 1xel (21) fits with no steric overlaps in the SQD1 model, and the binding of UDP-glucose to SQD1 is supported by stabilization of the SQD1 protein by addition of UDP-glucose. The reaction of SQD1 requires a sulfur donor (whose identity is currently under investigation), which may mean that the active-site side chains also reflect complementarity to a sulfur containing ligand.

Determination of NAD^+ by HPLC and characterization by MALDI-MS. For the biochemical analysis of SQD1, highly purified protein was required. The recombinant protein was at least 95% pure following chromatography on nickel resin, as determined by Coomassie staining of a SDS gel (Fig. 6). It was soluble and stable in 50 mM Mes buffer, pH 6.5, at 2 mg ml^{-1} protein concentration for 2 to 3 days at 4°C. Addition of UDP-glucose (10 mM) stabilized the protein for at least 10 days under the same conditions (data not shown), suggesting interaction between UDP-glucose and SQD1. To investigate whether the previously unsuspected cofactor, NAD^+, was indeed a ligand of SQD1, we employed HPLC to resolve nucleotides according to their constituent bases. For this purpose, the SQD1 protein was purified, dialyzed to remove the purification buffer, denatured, and then precipitated, and the supernatant was analyzed by HPLC. The chromatogram displayed one major compound eluting at 22 min (UV absorbance at 254 nm). Injection of an NAD^+ standard (99.9% pure from Boehringer-Mannheim) revealed a single compound at the same retention time (Fig. 7), suggesting that the observed compound in the supernatant represents NAD^+ formerly bound to the enzyme. To confirm the structure of this compound, we analyzed the collected material by MALDI-MS. The MALDI mass spectrum of the sample contained a predominant signal representing the molecular ion (MH^+) at a mass-to-charge ratio of 664.2 (Fig. 8), nearly identical to that expected for NAD^+ (calculated molecular mass, 664.3). As a control, we also analyzed the purified NAD^+ standard by MALDI-MS, producing an identical spectrum. Taken together, these results demonstrate that NAD^+ is associated with SQD1, and the requirement for denaturing conditions to retrieve the cofactor from SQD1 suggests it is tightly bound.

Determination of NAD^+:SQD1 stoichiometry by enzyme assay. In addition to the HPLC characterization, we determined the stoichiometric ratio of NAD^+ to SQD1 using an enzymatic assay. The purified and subsequently denatured protein was removed by cen-

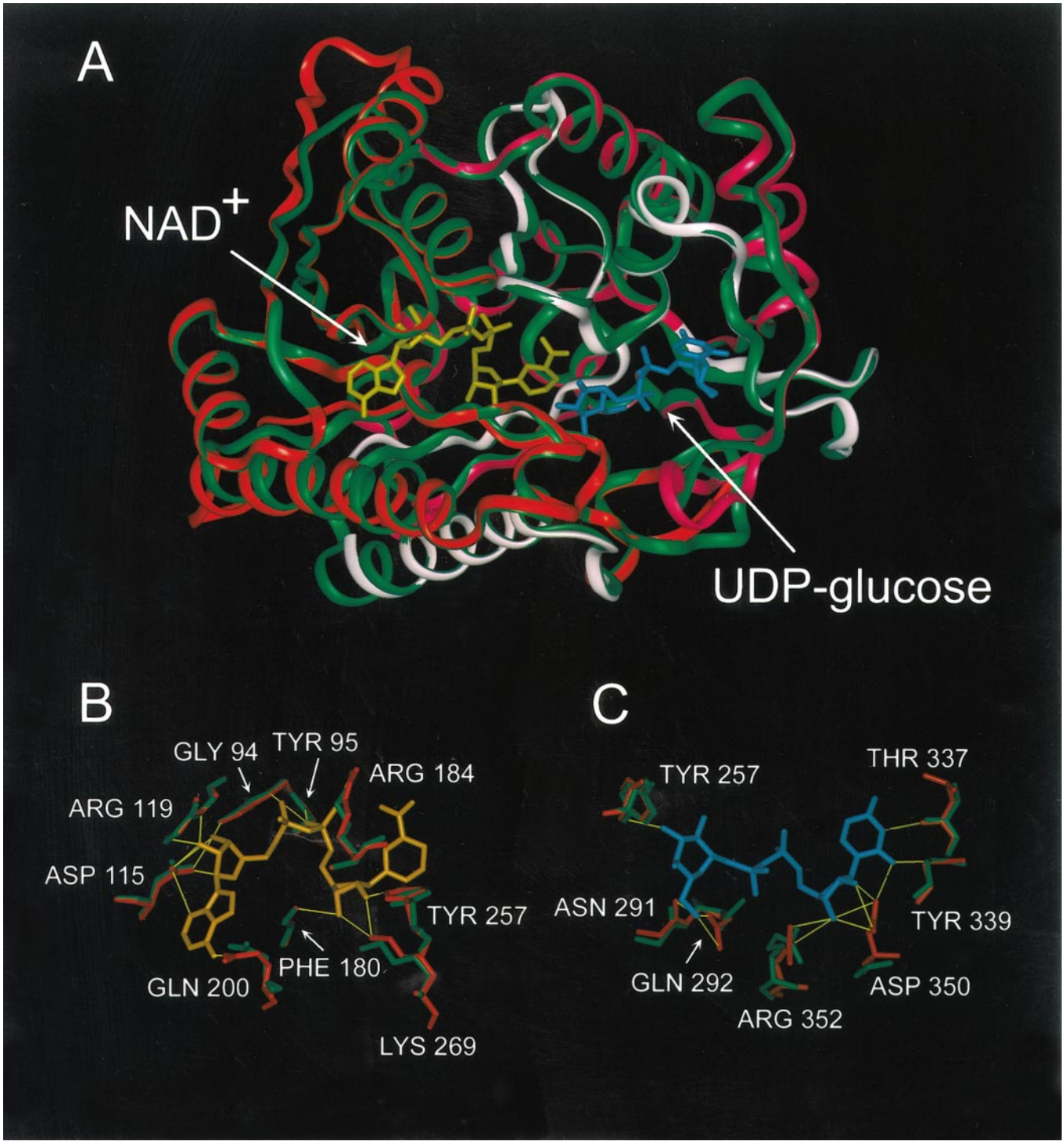

FIG. 5. Superposition of SQD1 and UDP-glucose epimerase structures. (A) Main-chain superposition of the SQD1 model (residues 87–239, red; 240–323, light pink; 324–441, magenta) with UDP-glucose epimerase (PDB entry 1xel, green). The bound UDP-glucose is colored blue, with the NAD^+ cofactor colored orange. (B) Close-up view of the NAD^+ (gold) binding site of SQD1 (red), displaying only those residues which form hydrogen bonds (yellow lines) to the cofactor. Corresponding residues from the epimerase are shown in green. (C) The UDP-glucose (blue) binding site of SQD1 (red), displaying only those residues forming hydrogen bonds to the ligand, with corresponding residues from the epimerase shown in green.

trifugation, and the supernatant was assayed (see Experimental Procedures). Averaging measurements on independent protein preparations, we determined a ratio of 0.5 ± 0.1 equivalents NAD^+ per mole of SQD1 protein. In contrast, when nondenatured protein was used in the same assay, no NAD^+ was observed.

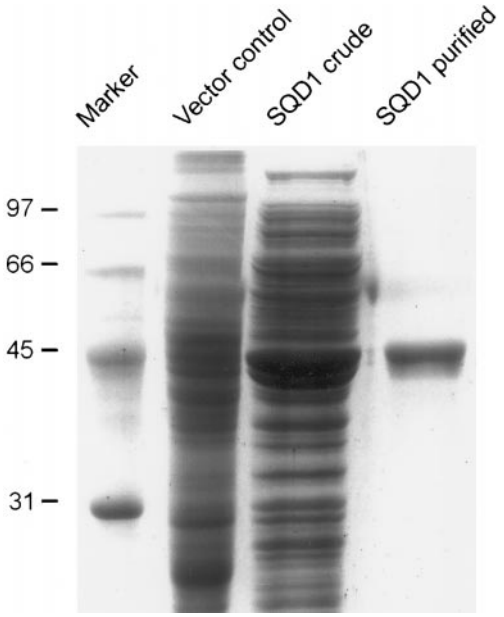

FIG. 6. SDS–PAGE showing the purification of the recombinant SQD1 protein. The marker lane indicates the masses of molecular markers in kDa. Soluble protein extracts are shown for pQE30 vector control, containing no insert; pQE30 + SQD1 before purification; and the purified SQD1 protein.

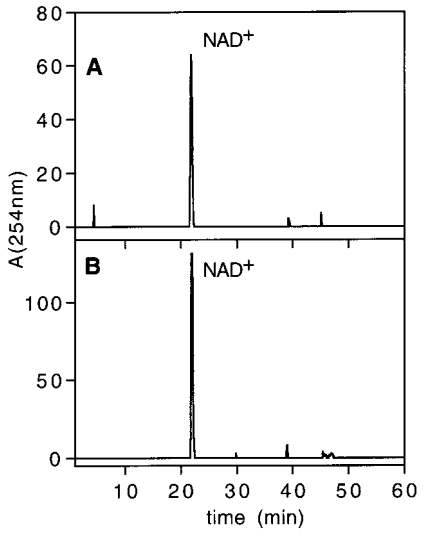

FIG. 7. Determination of NAD^+ as cofactor in the SQD1 protein using HPLC. (A) UV trace following injection of supernatant of the denatured SQD1 protein. (B) UV trace of NAD^+ standard, in the protein buffer (without SQD1).

DISCUSSION

Structural implications of SQD1 similarity to epimerases. A database search and subsequent sequence alignment revealed significant similarity between SQD1 and dehydratases and hexose-nucleotide epimerases, members of the dehydrogenase/reductase superfamily (34). Comparison of the SQD1 structural model of epimerases further predicted the presence of an NAD^+ binding site formed by the enzyme's Rossmann fold (35), which was not identified by sequence analysis, including a Prosite pattern search. Subsequently, we have shown via enzymatic assay, HPLC, and mass spectrometry that SQD1 does indeed bind the cofactor NAD^+. In determining the cofactor/enzyme ratio, our results indicate a ratio of 0.5 ± 0.1 equivalents NAD^+/mol of SQD1. While variability between preparations is small, as indicated by the small standard deviation in the stoichiometry, it is possible that a consistent fraction of the nucleotide cofactor is present in reduced form as NADH, which would elude enzymatic quantification. For CDP-glucose 4,6-dehydratase, half of the cofactor sites may be occupied by NADH (36). Original reports for UDP-glucose epimerase suggested the presence of only one NAD^+ molecule per enzyme dimer, although a recently obtained X-ray structure shows one equivalent of NAD^+/mol of the monomeric enzyme (28). The similarity in structure and our biochemical data suggest that SQD1 binds one nucleotide per monomer.

Catalytic implications. Interpretation of the structural model in conjunction with the biochemical data allowed us to propose a reaction mechanism for UDP-sulfoquinovose formation (Fig. 9A), in which SQD1 catalyzes the first step in sulfolipid biosynthesis: the conversion of UDP-glucose to UDP-sulfoquinovose. While a functional assay for SQD1 activity is not yet possible due to the unknown identity of the sulfur donor, we believe UDP-glucose is the substrate for SDQ1 based on the following observations. First, UDP-glucose stabilizes SQD1; second, labeling experiments suggest that UDP-glucose acts as a precursor in sulfolipid biosynthesis (13); finally, SQD1 exhibits high sequence and structural similarity to nucleotide-glucose-modifying enzymes, including epimerases and dehydratases. Analysis of the catalytic site of the structural model (based on structural superposition of SQD1, UDP-glu-

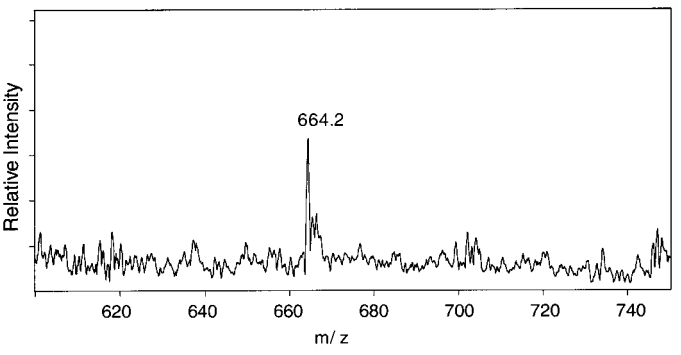

FIG. 8. Identification of NAD^+ by MALDI-MS. High mass region of MALDI-MS for the collected HPLC peak.

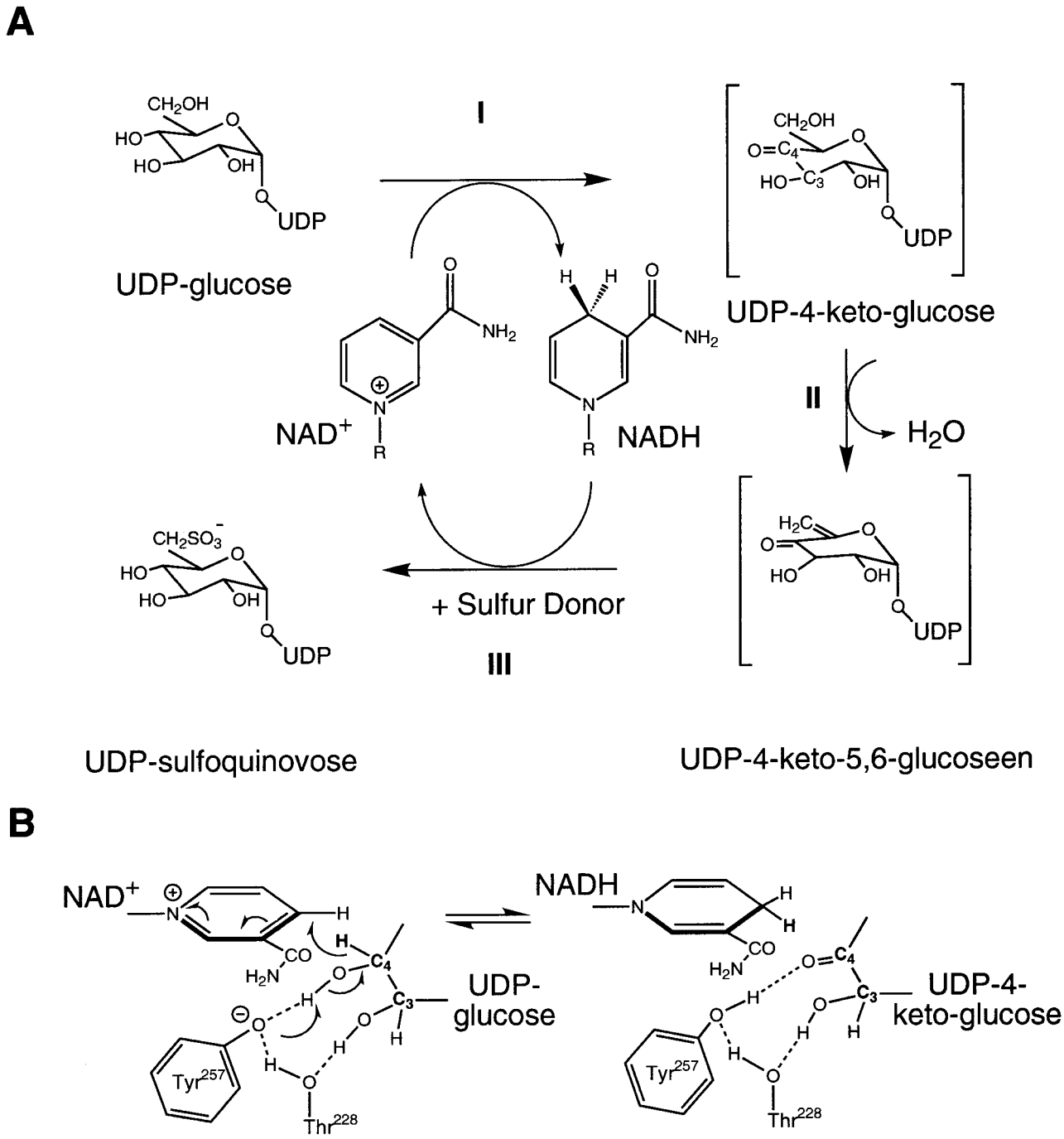

FIG. 9. Proposed SQD1 reaction mechanism. (A) The conversion of UDP-glucose to UDP-sulfoquinovose. (B) Structural chemistry proposed for formation of the UDP-4-keto-glucose intermediate. Based on the two possible sequence alignments for residues 246–267 in SQD1 with the epimerase (discussed in the legend for Fig. 3), both Tyr257 or 265 in SQD1 are candidates for the catalytic phenolate (Tyr149 in the epimerase).

cose epimerase, and dTDP-glucose dehydratase) shows that Thr228 in SQD1 is conserved in the dehydratase and corresponds to Ser124 of the epimerase, providing the isosteric hydroxyl group shown to be catalytically important (37). Also, this threonine residue is absolutely conserved and lies in a highly conserved region of orthologous SQDB bacterial proteins. Importantly, there is a Ser124 → Thr mutant of the epimerase for which the structure has been solved, and this mutant structure retains NAD^+ in its active *syn* conformation,

as does the dehydratase structure. We propose that Thr228 of SQD1 performs the same role as Ser124 in the epimerase, which has the proper chemistry and orientation to facilitate formation of the 4-keto-glucose intermediate (Fig. 9), a step shared with dehydratases (36, 38).

The second catalytically important residue, tyrosine, is conserved among SQD1, UDP-glucose epimerase, and dTDP-glucose dehydratase (Fig. 3). Tyr149 of the *E. coli* UDP-galactose-4-epimerase has been proposed to be involved in charge-transfer complexation with NAD^+ (39); the pH dependence of charge transfer supports that it is strongest when Tyr149 is ionized to the phenolate ($Tyr\text{-}O^-$) and weakened by protonation of Tyr149 and in a Tyr149 → Phe mutant. The pK_α of the epimerase Tyr149 is 6.08, supporting its ionizability (39). Based on crystallographic data on the epimerase (21), it has been proposed that Tyr149 and Ser124 can interact directly with NAD^+ and UDP-glucose, and their role could include general acid–base catalysis of hydride transfer from the C-4 hydroxyl of glucose and/or orienting the sugar ring through hydrogen bonding to its C-3 hydroxyl group (39). Two possibilities have been proposed for the specific roles of the Tyr and Ser side chains in mediating epimerase-catalyzed hydride transfer between UDP-glucose and NAD^+ (39): (i) the phenolate group of Tyr149 abstracts a proton from the hydroxyl group of Ser124, which in turn abstracts a protein from the C-4 hydroxyl group of UDP-glucose, leading to hydride transfer from C-4 of glucose to NAD^+, which converts NAD^+ to NADH; or (ii) the Ser124 hydroxyl group hydrogen-bonds to the C-3 hydroxyl group of glucose and to the phenolate of Tyr149, positioning it to directly abstract the proton from the C-4 hydroxyl group of UDP-glucose, followed by C-4 hydride transfer as described in (1). Based on the simpler of these two possibilities (2), we propose that the corresponding phenolate in SQD1 accepts a proton from the C-4 hydroxyl group of the glucose moiety, concurrently with NAD^+ accepting the glucose C-4 carbon hydride (Fig. 9B). This forms the UDP-4-keto-glucose intermediate in step I of the proposed mechanism (Fig. 9A).

Structural interpretation of the sequence alignment (Fig. 2B) implies that the high sequence identity in the N-terminal domain of SQD1 to the epimerases (26%; red ribbon in Fig. 5A) is due to the conserved NAD^+ binding site, whereas the high sequence identity in the SQD1 C-terminal domain to the *Arabidopsis* dehydratase (32%; magenta ribbon in Fig. 5A) reflects functional similarity with respect to UDP-glucose binding and chemistry. Therefore, we propose that step II (Fig. 9A), by analogy to dehydratases, is a β-elimination reaction of water, whereby UDP-4-keto-5,6-glucoseen is formed as a reactive intermediate from UDP-4-keto-glucose. A similar reaction mechanism with a sugar nucleotide 4-keto-5,6-en as an intermediate was discovered and reported for TDP-glucose dehydratase (40, 41), and proposed for GDP-mannose (42) and CDP-glucose (36) in dehydratase reactions. His423 in SQD1 is positioned appropriately to serve as a general base in the β-elimination reaction, abstracting the C-5 proton from UDP-4-keto-glucose.

With regard to the third step, formation of UDP-sulfoquinovose (Fig. 9A), several scenarios are possible. The suggestion of an intermediate such as UDP-4-keto-5,6-glucoseen is reasonable because it has been shown that a similar substrate, glucose-5,6-en, is able to accept a sulfite in an aqueous radical reaction at room temperature, pH 5.6 (43, 44). Another possibility is the nucleophilic addition of a sulfite to a mesomeric form of UDP-4-keto-5,6-glucoseen, with a carbonium ion at the C-6 position and carbanion at the C-5 position (41). Sulfite may be released through the reduction of activated sulfur compounds such as PAPS or APS. The enzymes that catalyze the reduction of these compounds, PAPS-reductase and APS-reductase, respectively, are not well characterized in plants, and the reaction products have not been determined for either enzyme. However, recent results support the presence of an APS-dependent pathway in plants, which proceeds via free sulfite (45, 46). Confirmation of the proposed reaction mechanism for SQD1 will ultimately be provided by solving the crystal structure of SQD1 and trapping the intermediates, as well as identifying the sulfur donor of the reaction.

ACKNOWLEDGMENTS

We are grateful to Drs. Gerald Babcock and Peter Dörmann for their valuable comments on the manuscript. We also gratefully acknowledge Dr. Douglas Gage for the analysis of NAD^+ by mass spectrometry.

REFERENCES

1. Heinz, E. (1993) *in* Sulfur Nutrition and Assimilation in Higher Plants (De Kok, L. J., Ed.), pp. 163–178, SPB Acad., Den Haag, Netherlands.
2. Barber, J., and Gounaris, J. L. (1986) *Photosynth. Res.* **9,** 239–247.
3. Güler, S., Seelinger, S., Härtel, H., Renger, G., and Benning, C. (1996) *J. Biol. Chem.* **271,** 7501–7507.
4. Essigmann, B., Güler, S., Narng, R. A., Linke, D., and Benning, C. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 1950–1955.
5. Otha, K., Mizushina, Y., Hirata, N., Takemura, M., Matsukage, F., Yoshida, A., and Sakaguchi, K. (1998) *Chem. Pharm. Bull.* **46,** 684–686.
6. Mizushina, Y., Watanabe, I., Otha, K., Takemura, M., Sahara, H., Takahashi, N., Gasa, S., Sugawara, F., Matsukage, A., Yoshida, S., and Sakaguchi, K. (1998) *Biochem. Pharmacol.* **55,** 537–541.
7. Gustafson, K. R., Cardellina, J. H., Fuller, R. W., Weislow, O. S., Kiser, R. F., Snader, K. M., Patterson, G. M. L., and Boyd, M. R. (1989) *J. Natl. Cancer Inst.* **81,** 1254–1258.

8. Benning, C. (1998) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49,** 53–75.
9. Benson, A. A. (1963) *Adv. Lipid. Res.* **1,** 387–394.
10. Haines, T. H. (1973) *in* Lipids and Biomembranes of Eukaryotic Microorganisms (Erwin, J. A., Ed.), pp. 197–232, Academic Press, New York.
11. Harwood, J. L. (1975) *Biochim. Biophys. Acta* **398,** 224–230.
12. Barber, G. A. (1963) *Arch. Biochem. Biophys.* **103,** 276–282.
13. Pugh, C. E., Roy, A. B., Hawkes, T., and Harwood, J. L. (1995) *Biochem. J.* **309,** 513–519.
14. Pugh, C. E., Hawkes, T., and Harwood, J. L. (1995) *Phytochemistry* **39,** 1071–1075.
15. Rossak, M., Tietje, C., Heinz, E., and Benning, C. (1995) *J. Biol. Chem.* **270,** 25792–25797.
16. Rossak, M., Schäfer, A., Xu, N., Gage, D. A., and Benning, C. (1997) *Arch. Biochem. Biophys.* **340,** 219–230.
17. Seifert, U., and Heinz, E. (1992) *Bot. Acta* **105,** 197–205.
18. Tietje, C., and Heinz, E. (1998) *Planta* **206,** 72–78.
19. Benning, C., and Somerville, C. R. (1992) *J. Bacteriol.* **174,** 6479–6487.
20. Sander, C., and Schneider, R. (1991) *Proteins* **9,** 56–68.
21. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) *Biochemistry* **35,** 5137–5144.
22. Caputto, R., Leloir, L. F., Cardini, C. E., and Paladini, A. C. (1950) *J. Biol. Chem.* **184,** 333–350.
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
24. Huang, X., and Miller, W. (1991) *Adv. Appl. Math.* **12,** 337–357.
25. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18,** 2714–2723.
26. Abola, E., Bernstein, F., Bryant, S. H., Koetzle, T. F., and Weng, J. (1987) *in* Crystallographic Databases—Information Content, Software Systems, Scientific Applications (Allen, F., Bergerhoff, G., and Sievers, R., Eds.), pp. 107–132, Bonn/Cambridge/Chester.
27. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
28. Thoden, J. B., Frey, P. A., and Holden, H. M. (1996) *Prot. Sci.* **5,** 2149–2161.
29. Sobolev, V., Wade, R. C., Vriend, G., and Edelman, M. (1996) *Proteins* **25,** 120–129.
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* **26,** 283–291.
31. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) *Proteins* **12,** 345–364.
32. Laemmli, U. K. (1970) *Nature* **227,** 680–685.
33. Rossmann, M. G., and Argos, P. (1978) *Mol. Cell. Biochem.* **21,** 161–182.
34. Holm, L., Sander, C., and Murzin, A. (1994) *Nat. Struct. Biol.* **1,** 146–147.
35. Kutzenko, A. S., Lamzin, V. S., and Popov, V. O. (1998) *FEBS Lett.* **423,** 105–109.
36. He, X., Thorson, J. S., and Liu, H. (1996) *Biochemistry* **35,** 4721–4731.
37. Thoden, J. B., Gulick, A. M., and Holden, H. M. (1997) *Biochemistry* **36,** 10685–10695.
38. Bonin, C. P., Potter, I., Vanzin, G. F., and Reiter, W.-D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 2085–2090.
39. Liu, Y., Thoden, J. B., Kim, J., Berger, E., Gulick, A. M., Ruzicka, F. J., Holden, H. M., and Frey, P. A. (1997) *Biochemistry* **36,** 10675–10684.
40. Melo, A., and Glaser, L. (1968) *J. Biol. Chem.* **243,** 1475–1478.
41. Gabriel, O., and Lindquist, L. C. (1968) *J. Biol. Chem.* **243,** 1479–1484.
42. Oths, P. J., Mayer, R. M., and Floss, H. G. (1990) *Carbohyd. Res.* **198,** 91–100.
43. Lehmann, J., and Weckerle, W. (1972) *Carbohyd. Res.* **22,** 23–35.
44. Lehmann, J., and Weckler, W. (1975) *Carbohyd. Res.* **42,** 275–295.
45. Hell, R. (1997) *Planta* **202,** 138–148.
46. Bick, J. A., and Leustek, T. (1998) *Curr. Opin. Plant Biol.* **1,** 240–244.
47. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138.
48. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637.