J Mol Model (2004) 10:317–327
DOI 10.1007/s00894-004-0197-6

ORIGINAL PAPER

Lars Bräuer · Wolfgang Brandt ·
Ludger A. Wessjohann

Modeling the *E. coli* 4-hydroxybenzoic acid oligoprenyltransferase (*ubiA* transferase) and characterization of potential active sites

Received: 12 March 2004 / Accepted: 26 May 2004 / Published online: 27 August 2004

Abstract 4-Hydroxybenzoate oligoprenyltransferase of *E. coli*, encoded in the gene *ubiA,* is an important key enzyme in the biosynthetic pathway to ubiquinone. It catalyzes the prenylation of 4-hydroxybenzoic acid in position 3 using an oligoprenyl diphosphate as a second substrate. Up to now, no X-ray structure of this oligoprenyltransferase or any structurally related enzyme is known. Knowledge of the tertiary structure and possible active sites is, however, essential for understanding the catalysis mechanism and the substrate specificity.
With homology modeling techniques, secondary structure prediction tools, molecular dynamics simulations, and energy optimizations, a model with two putative active sites could be created and refined. One active site selected to be the most likely one for the docking of oligoprenyl diphosphate and 4-hydroxybenzoic acid is located near the N-terminus of the enzyme. It is widely accepted that residues forming an active site are usually evolutionary conserved within a family of enzymes. Multiple alignments of a multitude of related proteins clearly showed 100% conservation of the amino acid residues that form the first putative active site and therefore strongly support this hypothesis. However, an additional highly conserved region in the amino acid sequence of the *ubiA* enzyme could be detected, which also can be considered a putative (or rudimentary) active site. This site is characterized by a high sequence similarity to the aforementioned site and may give some hints regarding the evolutionary origin of the *ubiA* enzyme.

Semiempirical quantum mechanical PM3 calculations have been performed to investigate the thermodynamics and kinetics of the catalysis mechanism. These results suggest a near S_N1 mechanism for the cleavage of the diphosphate ion from the isoprenyl unit. The 4-hydroxybenzoic acid interestingly appears not to be activated as benzoate anion but rather as phenolate anion to allow attack of the isoprenyl cation to the phenolate, which appeared to be the rate limiting step of the whole process according to our quantum chemical calculations. Our models are a basis for developing inhibitors of this enzyme, which is crucial for bacterial aerobic metabolism.

Keywords Transferases · Prenylation · Homology modeling · Biocatalysis mechanism

Abbreviations *3D-PSSM:* 3D protein secondary structure prediction · *4-HB:* 4-hydroxybenzoic acid (*para*-hydroxybenzoat) · *BLAST:* basic local alignment search tool · *BLOSUM:* blocks substitution matrix · *DMAPP:* dimethylallyl diphosphate · *GOLD:* genetic optimization ligand docking · *GPP:* geranyl diphosphate · *HIV:* human immunodeficiency virus · *MOE:* molecular operating environment · *OP-PP:* octaprenyl diphosphate · *PAM:* percent accepted mutation · *PDB:* protein data bank · *pHB: para*-hydroxybenzoic acid · *PM3:* parametrized method 3 · *PP or DPP:* diphosphate · *PROSA:* protein structure analysis · *TAFF:* tripos associated force field

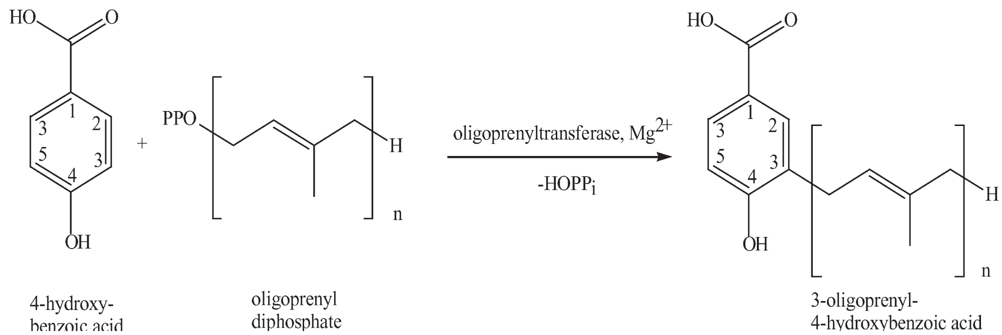
Introduction

4-Hydroxybenzoic acid oligoprenyltransferase from *E. coli* is encoded by the gene *ubiA*. It is a key enzyme in the biosynthetic pathway to ubiquinones, which are essential electron carriers in the respiratory chain of procaryotic and eucaryotic organisms. The two-substrate enzyme catalyzes the prenylation of 4-hydroxybenzoic acid in the 3-position utilizing an oligoprenyl diphosphate as activated component. This reaction is illustrated in Scheme 1.

The enzyme was discovered 1972 by Young et al. [1] in *E. coli* cell extracts. The corresponding gene was located on the physical map of *E. coli* by Nishimura et al. [2] and cloned by Heide et al. [3] and Nichols et al. [4] It consists of 290 amino acids and is membrane bound. Solubilization of the enzyme in detergents leads to fast

L. Bräuer · W. Brandt · L. A. Wessjohann (✉)
Department of Bioorganic Chemistry,
Leibniz Institute of Plant Biochemistry,
Weinberg 3, 06120 Halle, Germany
e-mail: wessjohann@ipb-halle.de
Tel.: +49(0) 345-55821 301

Scheme 1 The enzymatic prenylation of 4-hydroxybenzoic acid using oligoprenyl diphosphates (n>1)



and usually irreversible loss of activity. Furthermore, magnesium ions are essential for the catalytic activity. No X-ray structure of a 4-hydroxybenzoate oligoprenyltransferase or any homologous protein is known. Aromatic prenyltransferases are still largely white spots for structural biology. In order to understand the reaction mechanism and to support the design of selective inhibitors, knowledge of the tertiary structure and the active site is essential since 3-oligoprenyl-4-hydroxybenzoic acid is a direct precursor of ubiquinones (coenzymes Q_n). The specific inhibition of the biosynthesis of this compound should disrupt cellular respiration and thus affect bacterial growth. Additionally, the enzyme has been used successfully as a biocatalyst to form different C–C bonds under mild reaction conditions. [5] Modeling may also aid the design of alternative clones with a different, non-natural substrate spectrum.

Here we will describe a first 3D structural model of a member of this family of membrane bound prenyltransferases, including the characterization of two putative active sites and studies of the enzymatic catalysis mechanism.

Materials and methods

All calculations were performed on Silicon Graphics Octane workstations and personal computers. Structures were graphically displayed, modified and evaluated using SYBYL and STEREOGRAPHIC stereoglasses. [6] MOE, a molecular modeling program package, was used to prepare different models of 4-hydroxybenzoate oligoprenyltransferase. The homology modeling procedure inside MOE is automated. The template(s) and the substitution matrix and other parameters, e.g. gap penalties, must only be defined for an appropriate alignment between the template and target sequence. Briefly, the modeling procedure consists of the following steps: comparable enzymes with known X-ray structures from the PDB were aligned with the sequence of 4-hydroxybenzoate oligoprenyltransferase. [7]

For all calculations and structural refinements of the models without any ligand, the AMBER all atom force field, included in the modeling packages described above, was used. [8] The AMBER force field is very well suited for calculations of proteins, but due to the lack of parameters for the ligands the TRIPOS associated force field (TAFF) was used for calculations of protein–ligand complexes. [9] During all these simulations the backbone atoms of the protein were fixed.

Ten slightly different models were created with MOE and subsequently minimized by means of the AMBER force field including electrostatic interactions based on AMBER partial charge distributions. The final structure refinement (including lateron the ligands) of the resultant models was done using the TAFF implemented in SYBYL. The quality of the structures obtained was checked with PROSAII. [10] The program calculates the energy potentials for the atomic interactions of all amino acid residue pairs as a function of the distance between the corresponding atoms. The energies of all conformations that exist in an integrated data base with respect to the given sequence are calculated using the potential of mean force derived by statistical analysis of a set of natively folded proteins. Negative energies of a PROSAII plot indicate that the modeled structures may represent a native fold, whereas sequences with positive energies must be inspected critically. Due to the membrane location of the enzyme, only the C_α and C_β interactions are inspected, instead of including surface potentials, which would only be relevant for water-soluble proteins.

Different substrates (e.g. 4-hydroxybenzoate, geranyl diphosphate, octaprenyl diphosphate) were docked inside suggested active sites of the models. The docking of the substrates was done manually as well as automated. The program used for automated docking was GOLD (Genetic Optimized Ligand Docking), which uses genetic algorithms to find appropriate docking arrangements of ligands into receptor binding sites. [11] Additionally, the docking arrangements were refined using molecular dynamic simulations with subsequent energy minimization using the TAFF. To simulate an induced fit, the ligands and the side chains of the protein were considered flexible whereas all backbone atoms of the protein were fixed. Secondary structure predictions, alignments, multiple alignments, and other diagnostics were done by means of methods available on the internet. To align the sequences we used program routines implemented in SYBYL/COMPOSER. As substitution matrix we applied PMUTATION. The gap penalty was set to 8. The calculated identity score (% identity) is the number of identical residues in the two sequences divided by the length of the shortest sequence (without gaps). Semiempirical quantum mechanical calculations were carried out using the program package Spartan. [12] The program parameters and options for these calculations are given in detail in the relevant section of this article.

Results

Search for homologous proteins and structural similarities

Since there are no hard structural data available yet for any aromatic oligoprenyltransferase or closely related protein, the only chance to obtain first impressions of the 3D structure of the *ubiA* enzyme is to use homology modeling techniques. For this purpose, it was necessary to screen databases for proteins with a known three-dimensional structure, which display a maximum similarity to the *ubiA* enzyme. First we applied a BLAST search (Basic Local Alignment Search Tool), but no significant simi-

Table 1 Proteins with highest similarity scores to 4-hydroxybenzoic acid oligoprenyltransferase and X-ray structural data available. The calculated overall hydrophobicity of the X-ray structures is given by log *P*

PDB code	Name and source of the protein	Identity score	Overall hydrophobicity (log *P*)
5EAS	5-Epi-aristolochene synthase (*Nicotiana tabacum*)	23.2%	−578.9
1DCE	Geranylgeranyltransferase (*Rattus norvegicus*)	21.0%	−100.1
1PRC	Photosynthetic reaction centre (*Rhodopseudomonas viridis*)	20.9%	+45.7
1UBW	Farnesyl pyrophosphate synthetase (*Gallus gallus*)	20.2%	−20.45
1FX8	Glycerol facilitator (*Escherichia coli*)	18.1%	+25.6
3ERD	Human estrogen receptor (*Homo sapiens*)	18.1%	−13.9
1OYA	Old yellow enzyme (*Saccharomyces pastorianus*)	17.2%	−21.6
1AD1	Dihydropteroate synthetase (*Staphylococcus aureus*)	13.1%	67.8

larities were found. [13] Thus, we performed a search using the modeling programs MOE and Insight which include an internal protein database and specific search algorithms implemented in these programs. [14, 15] MOE comprises a non-redundant protein database, containing approximately 7,000 entries. Its substitution algorithm (substitution matrix) and the structural alignment causes a 10% higher crop compared to the BLAST search. The main difference between Insight and BLAST is the search algorithm. The search for similarities using BLAST is restricted to significant regions of the sequence only. This procedure makes this search tool very fast, but not all similar sequences can be found. On the other hand, Insight comprises the FASTA algorithm for similarity search, a methodology that enables the program to find more similarities. [16, 17] As a result of these searches, we identified eight proteins that show at least some limited similarity to the *ubiA* enzyme. The proteins identified and the corresponding values for their identities are listed in Table 1.

In order to generate a reasonable model, it is not only crucial to find template proteins with high similarity, but additionally the predicted secondary structure elements of the target protein must correspond with respect to the secondary structural moieties in the template protein. We used two web-based methods, phd, and 3D-pssm for the prediction of the secondary structure of the *ubiA* enzyme. [18, 19] The predicted secondary structures resulting from these two methods are compared in Table 2.

The bold letters designate amino acids of α-helical regions, which were predicted by both programs, phd and 3D-pssm. All other residues show loop regions, elements with indifferent secondary structure. In no case, sheet elements were predicted.

Comparing the predicted secondary structure of the *ubiA* enzyme with those of all eight proteins of maximum identity (Table 1) revealed only two proteins with useful similarity. However, we have to mention that currently secondary structure prediction methods have a reliability of some 75%. Another main criterion for a relevant template structure is the hydrophobicity of the protein surface. Due to the fact that the transferase is embedded in a biological membrane, we must detect hydrophobic amino acid residues, which may represent the membrane spanning region of the protein. Although there is no simple physical model available for this effect, the overall hydrophobicity of a molecule can be measured by its partition coefficient log *P*. The calculated log *P* values are given in Table 1. Only proteins with a positive log *P* were considered useful as templates for the modeling of this membrane-bound enzyme. The combined requirements of identity score, fitting, and hydrophobicity were fulfilled by only two candidates of Table 1: the photosynthetic reaction center from *Rhodopseudomonas viridis* (pdb code: 1prc) and the glycerol facilitator from *E. coli* (pdb code: 1fx8). The next step of our approach was to identify the putative catalytically active site(s). For this purpose, multiple alignments were carried out. These were done using 15 sequences of 4-hydroxybenzoic acid octaprenyltransferases from different microorganisms taken from the non-redundant biological database SWISSPROT. [20] The result of the alignment is shown in Table 3. Heide et al. [21] suggested a possible site for docking oligoprenyl diphosphate and 4-hydroxybenzoate near the N-terminus of the enzyme. However, Ashby and Edwards [22] identified two domains possessing a direct repeat of the consensus sequence xDDxxD and proposed these sites as putative substrate binding sites. The sequential location of the first consensus motive, predicted by Ashby and Edwards, is identical to the one proposed by Heide and coworkers. Our multiple sequence alignments could confirm the proposal of another independent sequence motive showing a high evolutionary conservation of Asp^{191} and Asp^{195} in addition to Asp^{71}and Asp^{75}. Therefore, the former must be considered as a binding site as well, in order to obtain a full picture.

Furthermore, we found other conserved amino acid residues (e.g. arginine^{137} and asparagine^{67}) in the neighborhood of the conserved aspartic acids. The following sequences of the two highly conserved regions were defined as putative active sites 1 (site 1) and 2 (site 2) for discussions below:

putative active site 1
−58 ˜MRAAGCVVNDYADRKFDGHVKRT ˜82

putative active site 2
−182 ˜YDTQYAMVDRDDDVKIG ˜200

Summarizing the results obtained up to here, we have two possible independent templates for homology modeling of the tertiary structure of the *ubiA* enzyme and could identify two putative active sites. Therefore, alto-

Table 2 Amino acid sequence alignment of the photosynthetic reaction centre protein (1prc) compared to predicted positions of secondary structural elements of the *ubiA*-enzyme, using phd and 3D-pssm

Number	10 20 30 40 50 60 70 80
1prc (exp.)	--------ALLSFERKYRVRGGTLIGGDLFDFWVG—PYFVGFFG**VSA—IFFIFL-GVSLIGYAA**SQGPTWD--PFAIS
ubiA-phd	MEWSL**TQNKLLAFHR—LMR**TDKPIGA-LLLL**WPTLWALWVATPGVPQLWILAVFVAG**VWLMRAAGCVVNDYADRKFDGH
ubiA-3D-pssm	ME**WSLTQNKLLAFHR—LM**RTDKPIGA-LLLLWPTLWALWVATPGVPQ**LWILAVFVAGVWLMRAAGCVVNDYADRKF**DGH
Number	90 100 110 120 130 140 150 160
1prc (exp.)	INPPDLKYGLGAAPLLEG**GFWQAITVCALGAF-ISWMLREVEI---SRKL**GIGW**HVPLAFCVPIFMFCVL-QVF---RPL**
ubiA-phd	VKRTANR-PLPSGAVT**EKEARALFVVLVLISFLLVLTLNTMTILLSIAALALAWV**YPFMKRYTHLP**QVVLGAAFGW**SIP**M**
ubiA-3D-pssm	VKRTANR-PLPSGAVT**EKEARALFVVLVLISFLLVL**TLN**TMTILLSIAALALAWV**YPFM**KR**YTHL**PQVVLGAAFGWSIPM**
Number	170 180 190 200 210 220 230 240
1prc (exp.)	**LL**GSWGHAFPYG----IL**SHLDWVNNFGYQ**YLNWHYN**PGHMSSVSFLFVNAMALGLHGGLILSVAN**PGDGDKVKT**AEHEN**
ubiA-phd	**AFAAV**SESVPLS**CWLMFLANILWAV**AYD**TQYAMV**DR**DDDVKIGIK---STAILF**GQY**DKLIIGILQIGVLALMAIIGELN**
ubiA-3D-pssm	**AFAAV**SESVP**LSCWLMFLANILWAVAYDTQYA**MVD**RDDDVK**IGIK---ST**AIL**FG**QYDKLIIGILQIGVLALMAIIGELN**
Number	250 260 270 280 290 300 310 320
1prc (exp.)	----**QYFRDVV**GYSIG**ALSIHRLGLFLAS**N**IFLTGAFGTIA**SGPFWTRG**WPEW**W**GWWL**DIPFWS----------------
ubiA-phd	GLG**WGYYWSIL** **VAGALFVYQQKLIA**NRE**R** **EACFKAFM**NNNY VG**LVLFLG** **LAMSYW**HF---------------
ubiA-3D-pssm	GL**GWGYYWSIL** **VAGALFVYQQKL**IANRE**R** **EACFKAFMNNNY** **VGLVLFLG** **LAMSYW**HF---------------

Bold highlighted are helical secondary structures either experimentally in 1prc or predicted for the *ubiA*-enzyme by phd and 3D-pssm. No significant β-sheet-elements are predicted for the *ubiA*-enzyme. The calculated identity score is 20.9 %.

gether four active sites had to be inspected and characterized for their relevance to allow docking of both substrates, 4-hydroxybenzoic acid and a prenyl diphosphate, close to each other. Due to our docking results, and their correlation with the known substrate specificity, we could exclude the model derived from the template glycerol facilitator of *E. coli* (pdb code: 1fx8) from being a reasonable one for this aromatic prenyltransferase. [23] As mentioned above, two highly conserved sequence motives were identified as putative active sites of the enzyme. For the model obtained from the glycerol facilitator of *E. coli*, one of these putative active sites could not be investigated due to strong steric hindrances between possible substrates and active amino acid residues around the catalytically active region. Consequently, only the model based on the photosynthetic reaction center (1prc) will be analyzed in detail and discussed in the following. The alignment and secondary structure comparisons with the *ubiA* enzyme are shown in Table 2.

Table 3 Alignment of partial sequences of 4-hydroxybenzoic acid octaprenyltransferases

Organism	amino acid residue	
	~ 71 75	~191 195
Campylobacter pylori	~GFNRLV**D**RDI**D**KDNPRT~	~FDLLYSLQ**D**MEF**D**KER~
Helicobacter pylori	~GFNRLV**D**RDI**D**KDNPRT~	~FDLLYSLQ**D**MEF**D**KER~
Archaeoglobus fulgidus	~TFNRII**D**REI**D**AKNPRT~	~FDMIYGLQ**D**VDF**D**RSN~
Aquifex aeolicus	~AFNRLI**D**EPY**D**RLNPRT~	~FDVLYALQ**D**YEF**D**KEV~
Bacillus pseudofirmus	~SLNRVI**D**EKI**D**KYNPRT~	~FDVIYATQ**D**ADY**D**RER~
Chlamydia trachomatis	~IVNQVV**D**CAI**D**KRNPRT~	~NDIIYALQ**D**VEF**D**QKE~
Sulfolobus solfataricus	~TNDNLA**D**IEI**D**AKNPRT~	~FDLYNHIP**D**AEF**D**KKM~
*Escherichia coli**	~-VVNYA**D**RKF**D**GHVKRT~	~YDTQYAMV**D**RDD**D**VKIG~
Pasteurella multocida	~VINDYA**D**RHI**D**GAVKRT~	~YDTQYAMV**D**RDD**D**LRIG~
Rickettsia conorii	~IINDIF**D**RKF**D**KHVART~	~YDTIYGYM**D**IRD**D**KKIG~
Rickettsia prowazekii	~IINDIF**D**RKF**D**KYVERT~	~YDTIYGYM**D**IKD**D**KKIG~
Wolbachia sp.	~IINDIF**D**RKI**D**AHVERT~	~YDTIYAHQ**D**KKD**D**EKLG~
Synechocystis sp.	~-VVNDW**D**RDI**D**PQVERT~	~FDTVYAMA**D**RED**D**RRIG~
Sulfolobus solfataricus	~VINDVY**D**VEI**D**KINK–~	~REIVKGIE**D**YNG**D**SLNN~
Pyrococcus abyssi	~TINDYF**D**VEI**D**RVNR–~	~REIMKDIE**D**IEG**D**MKM~

Bold letters represent the most conserved amino acid residues, in this case aspartates (D). Only the fully conserved, and for our suggested enzymatic reaction most relevant residues are labeled.

Homology modeling and structure refinement

The photosynthetic reaction center of *Rhodopseudomonas viridis* is a protein composed of four subunits. For homology modeling we used only chain L, which consists of

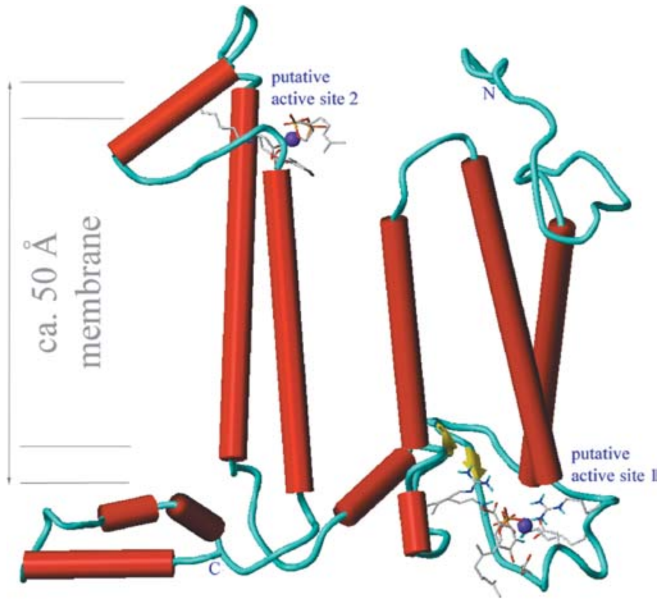


Fig. 1 Structure of the model of *ubiA* oligoprenyltransferase derived from the photosynthetic reaction center (1PRC). Putative active amino acid residues and substrates are shown as *capped sticks* to describe their location and geometry in the putative active sites. The *violet* spheres identify Mg^{2+}

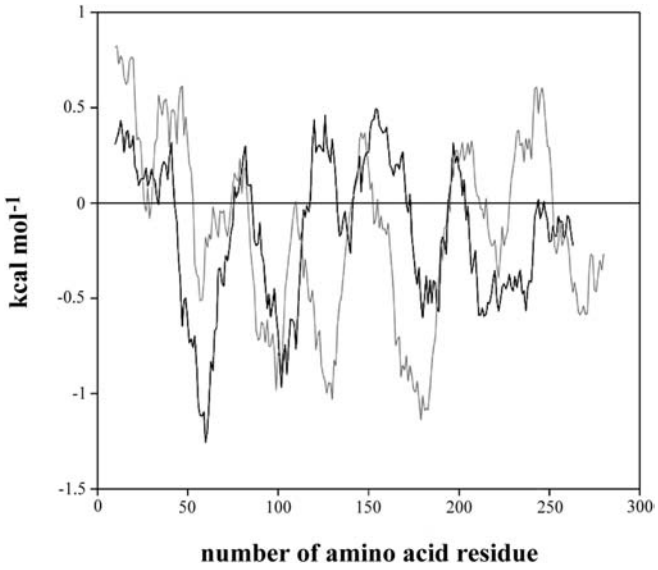


Fig. 2 PROSAII energy plot for the *ubiA* protein model based on 1PRC. The potential of mean force for the developed model is given by the brighter line. In comparison the potential of mean force for the template (1prc) is displayed as the dark line. A major discrepancy between model and template protein is located between the amino acid residues 120–145. However, in this region the potential of mean force is better for the model than for the template protein

273 amino acid residues. The template protein and the target sequence of *ubiA* were superimposed on the basis of alignments in MOE. Moreover, MOE was used to generate and calculate ten independent homology models. These models are built using a Boltzmann-weighted randomized modeling procedure adapted from Levitt, combined with a specialized logic for the proper handling of insertions and deletions. [24, 25] Each of these intermediate models is evaluated by a residue-packing quality function, which is sensitive to the degrees to which nonpolar side chain groups are buried and hydrogen bonding opportunities are satisfied. The best of these intermediates was selected for further optimization and improvement. Molecular dynamics simulations and simulated annealing calculations for some loops were done to enhance the quality of the models. The resulting structure model is characterized by five putative transmembrane helices and is shown in Fig. 1.

Structure evaluation

The quality and stereochemistry of the three-dimensional structure of the model was evaluated and analyzed using PROCHECK at a theoretical resolution of 2.0 Å. [26] The dihedral angles (Phi and Psi) are localized only in most favored and additionally allowed regions of the Ramachandran plot.

All other criteria such as peptide bond planarity, hydrogen bond energy etc. are within values statistically expected for proteins with a resolution below 2.0 Å. Energy profiles for C_α carbons (potentials of mean force), calculated with PROSAII, show small deviations in some regions of the models, compared with the template structures (Fig. 2).

Docking studies and characterization of the active sites

As shown in Fig. 1 the active site 1 is situated in a loop region, probably outside of the cell membrane, but directly at the surface of the cell membrane. Because of the hydrophobicity of longer prenyl side chains, we expect these substrates to be located in the membrane, whereas the diphosphate sticks out and is able to dock into the active site cavity. The other presumed active site 2 is located on the opposite site, in the interior of the enzyme and within the membrane area, but perhaps still reachable from the cytosol.

From experimental studies, it is known that magnesium ions play a crucial role in the activity of the enzyme. For this reason, a magnesium ion was assumed to bind to at least one of the conserved aspartates in the active sites. Docking studies to active site 1 using GOLD clearly show that this site may indeed represent a catalytically active domain (Fig. 3). [27]

At this site, a magnesium ion forms a tetrahedral metal complex with the two oxygen atoms of Asp^{75} and two oxygens of the diphosphate substrate. The theoretically remaining two ligand sites to form an octahedral ligand sphere at the metal are not occupied. The binding of the magnesium ion is very similar to the arrangement found in the active site of farnesyl diphosphate synthetase (pdb code: 1UBY), where a magnesium ion is simultaneously

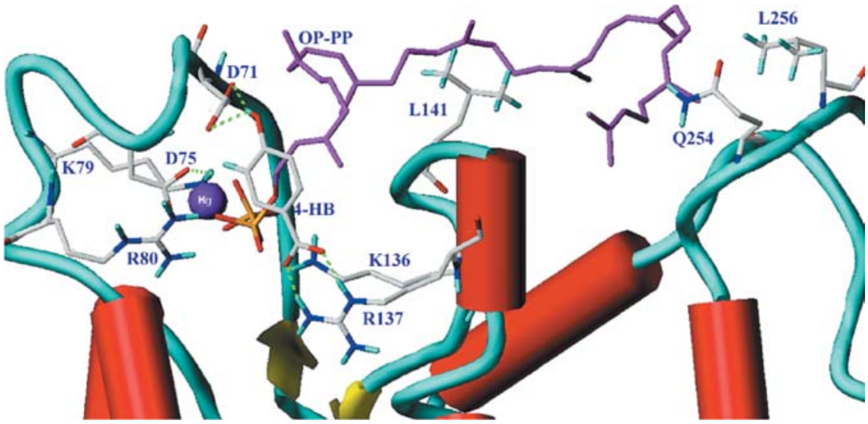


Fig. 3 Docking arrangement of 4-hydroxybenzoic acid and octaprenyl diphosphate including a magnesium ion in the active site 1 of the *ubiA* protein model based on the photosynthetic reaction center of *Rhodopseudomonas viridis* (1prc)

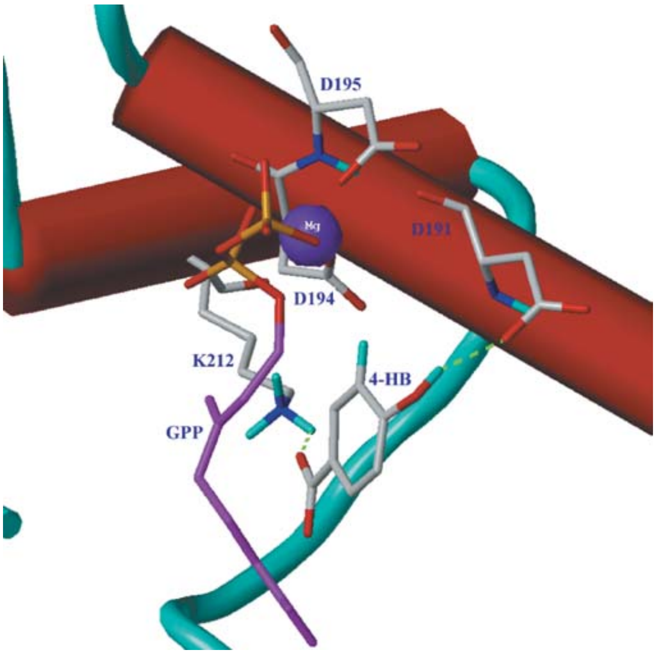


Fig. 4 In the putative active site 2 of the *ubiA* enzyme model the position of 4-hydroxybenzoate would be directed by hydrogen bonds of Asp^{191} and Lys^{212} to the hydroxy and carboxylate group, respectively. In this model the magnesium ion is bound by five donor atoms, i.e. with one oxygen each of Asp^{194} and Asp^{195}, and three oxygen atoms of the diphosphate unit of geranyl diphosphate. The remaining position to form an octahedral magnesium complex may be occupied by a water molecule

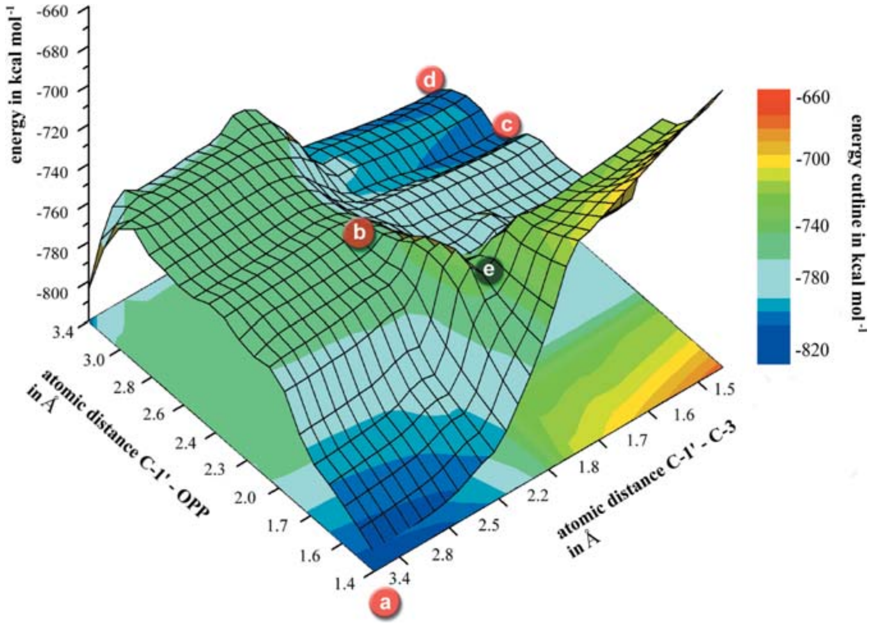
complexed by an aspartate and a dimethylallyl diphosphate. Furthermore, in our model 4-hydroxybenzoate forms a salt bridge of its carboxyl group to the side chain of Arg^{137} and the phenolic hydroxy group of the benzoic acid derivative forms a hydrogen bond to Asp^{71}. In addition to the binding of the diphosphate moiety via complexation of the magnesium ion, the docking arrangement is stabilized by hydrophobic interactions of the organic oligoprenyl chain, especially with Leu^{141} and Leu^{256} of the enzyme. In this way, the two substrates are bound close to each other in the active site. In order to test the relevance of the model, (2-*Z*)-octaprenyl diphosphate was also docked to this active site. From experimental investigations it is known that 2-*cis* isomers are not substrates of the oligoprenyltransferase. [28] The resulting docking arrangement showed that the calculated affinity of the *cis* isomer to the enzyme (p$K_{d(calc)}$=8.84) is reduced in comparison to the all *trans* isomer (p$K_{d(calc)}$=10.11). However, more relevant is that the prenyl chain cannot be docked close to the 4-hydroxybenzoic acid due to significant steric interactions or even overlap with protein side chains. This result—also true for the other active site discussed—nicely explains why the (2-*Z*)-isomers of oligoprenyl diphosphates are not substrates.

In active site 2 (cf. Fig. 4) the binding of the 4-hydroxybenzoic acid takes place via hydrogen bonds to Asp^{191} and Lys^{212}. The magnesium ion in this model is complexed by Asp^{194} and Asp^{195}. The diphosphate unit of geranyl diphosphate is bound to the magnesium ion and in this way close to the hydroxybenzoic acid derivative. The complexation of the magnesium ion shows a high similarity to the arrangement of reactive amino acid residues in the active site of HIV integrase (pdb code: 1BIU), where a magnesium ion is complexed by two aspartate residues. The interaction of octaprenyl diphosphate with the enzyme is stabilized by several hydrophobic amino acid residues such as Leu^{123}, Leu^{120}, Leu^{235} and Tyr^{239}. For this putative active site, and the affinity of (2-*Z*)-octaprenyl diphosphate is reduced to only 6.13.

A suggested mechanism for the prenylation of 4-hydroxybenzoic acid with oligoprenyl diphosphate catalyzed by 4-hydroxybenzoate oligoprenyltransferase (ubiA)

Based on our experimental substrate models and on the structure of active site 1 of the developed protein model, detailed investigations of the reaction mechanism were performed using semiempirical quantum mechanical PM3 calculations to analyze the catalytic mechanism. [23] For these calculations, all amino acid residues were simplified by using acetic acid or acetate to mimic aspartic acid or aspartate, respectively, lysine was represented by methylammonium, and arginine by methyl-

Fig. 5a–e Interpolated energy (hyper) surface of the aromatic prenylation reaction (cf. Scheme 2b). Energies for each point on the grid result from calculations with the semiempirical quantum mechanical method PM3. **a** Starting geometry (−816.18 kcal mol^{-1}). **b** Transition state (−755.3 kcal mol^{-1}). **c** Sigma complex of the aromatic substitution (−800.6 kcal mol^{-1}). **d** Final geometry (−832.68 kcal mol^{-1}). **e** Just for clarity, this point is ca. 20 kcal mol^{-1} higher in energy than the transition state (b) (optical distortion of 3D image)



guanidinium. Calculation distance constraints were set fixed for all carbon atoms of the methyl groups of the amino acid residue mimics to simulate the relative rigidity of the protein backbone and to ascertain the arrangement of the residues to each other. In analogy to common Mg^{2+} aspartate binding motifs, the magnesium ion forms a tetrahedral metal complex with Asp^{75} and the diphosphate of the dimethylallyl diphosphate (as a model for all prenyl diphosphates), and the carboxylic acid group of 4-hydroxybenzoate forms a salt bridge to the side chain of Arg^{137}. Additionally, the orientation of the phenolic hydroxy group of the benzoic acid derivative is fixed by a hydrogen bond to Asp^{71}.

A complete energy grid was calculated by stepwise cleavage of the diphosphate from the dimethylallyl unit and attack of the carbocation formed on the benzoate (cf. Fig. 5). Both reaction coordinates (cf. Fig. 5) were calculated in steps of 0.1 Å. At each grid point the two reaction coordinates were fixed, whereas the remaining system was completely optimized and the corresponding energy was calculated. These calculations should help to decide if the preferred reaction path is S_N1-like or S_N2-like.

Figure 5 shows the calculated energy landscape, with the different energy values (sum of the heats of formation for the whole system) encoded as different colors. The reaction starts at an energy value of −816.2 kcal mol^{-1} in an arrangement as shown in Fig. 3 and in more detail in Fig. 6a. The distance of the phosphoric ester bond between O and C-1' is 1.40 Å, whereas the distance of the bond forming prenyl-C-1' and C-3 of the 4-hydroxybenzoic acid is longer than 3.5 Å. The lowest energy path leads to a transition state geometry where the C-1' of the allyl cation approaches the C-3 atom of the 4-hydroxybenzoic acid with a distance of 2.48 Å whereby the C-1'-OPP bond is almost cleaved (distance = 2.4 Å). The heterogeneous cleavage of this bond is an energy-consuming step requiring some 61 kcal mol^{-1} (heats of formation of the whole system = −755.3 kcal mol^{-1}). The structure of the transition state (cf. Fig. 6b) clearly indicates that an S_N1 mechanism is preferred. Each pathway with shorter distances between the C-1' atom and the C-3 atom of the 4-hydroxybenzoic acid with still connected diphosphate leads to a much higher activation barrier.

After passing through the transition state, the reaction forms a σ complex intermediate (−800.6 kcal mol^{-1}, cf. Fig. 6c). The C-1'–OPP distance for this optimized structure is 2.8 Å, the one from C-1' to C-3 is 1.53 Å. A very small barrier of 4.7 kcal mol^{-1} has to be overcome to reach the final structures (−832.68 kcal mol^{-1}, cf. Fig. 6d). Finally the C-1' and diphosphate are separated by a distance of over 3.50 Å and the bond between C-1' and C-3 is established with a distance of 1.49 Å. The transfer of the C-3 proton to the oxygen of the α-phosphate, which became accessible during the cleavage of the diphosphate, supposedly takes place simultaneously to the formation of the new C-1'–C-3 bond. This proton translocation requires no additional energy and no new energy barrier has to be passed.

The complete formal mechanism of the prenylation reaction, including the intermediates without the amino acid residues that do not take part directly in the reaction are given in Scheme 2. The energies listed in this scheme correspond to the heats of formation of all amino acid side chains and substrates during the quantum mechanical calculations. Intermolecular interaction energies between all the compound have not been taken into consideration in this case. Values on the arrows represent the reaction enthalpies for each row in this scheme.

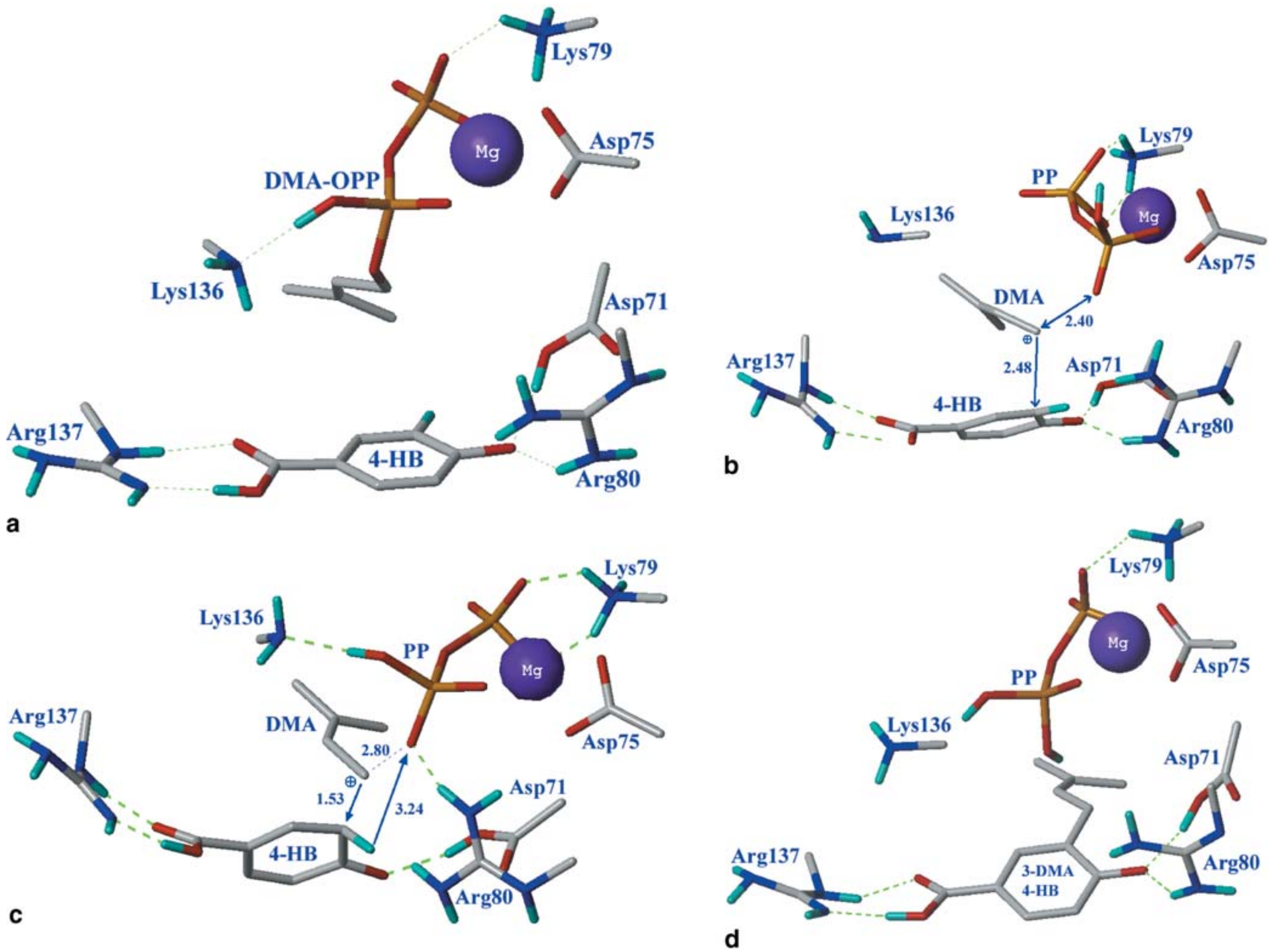


Fig. 6 a Initial geometry of the reactive species in active site 1. Dimethylallyl diphosphate (DMA-OPP) and Asp[75] are complexed by a tetrahedral magnesium ion. Two hydrogen bonds between Lys[79], Lys[136], and the diphosphate are formed. **b** Transition state geometry in active site 1. The carboxylic group of the 4-hydroxybenzoic acid (4-HB) is twisted out of the plane and causes an isolation of its π electrons from the aromatic ring. In this way an interaction between the allylic cation and the C3 atom becomes favorable. Due to a newly formed intramolecular hydrogen bond within the diphosphate, Lys[136] is loosing contact to the diphosphate. The newly formed hydrogen bond is more stable and preferred than the salt bridge between Lys[136] and the diphosphate. **c** Geometry of the sigma complex in active site 1. The bond between C1' and the oxygen of the diphosphate is already cleaved. A small change in the orientation of the diphosphate and the aromatic ring causes a decrease of the distance between the hydrogen on C3 and the α oxygen of the diphosphate and thus a transfer of the proton becomes possible. **d** Geometry in active site 1 after the prenylation. The proton from C-3 is already transferred to the accessible oxygen of the α phosphate. 3-dimethylallyl-4-hydroxybenzoic acid and the cleaved pyrophosphate (PP_i) are still residing in their positions.

Analysis of the substrate specificity by means of quantum mechanical investigations

The substrate specificity of the prenyl unit has already been discussed above. However, several hydroxybenzoate derivatives have also been tested experimentally as substrates for *ubiA* (cf. Table 4). [29]

It could be shown that some of these are viable substrates whereas others are not. Docking of some of the compounds to the active site 1 of the enzyme showed that in principal all could bind in an appropriate manner to allow catalysis. The only ligand that might bind in a slightly different orientation is 4-hydroxyphthalic acid. The difference is caused by the two neighboring carboxylic acids. This leads to an uncertain interaction with arginine[137] and through this to the prenyl unit. Apart from this exception, it is obvious that electronic properties might play a dominant role for the substrate specificity. To explain the specificity, we performed semiempirical quantum mechanical calculations (PM3). First of all, the heats of formations of the phenols in comparison to the phenolates (first step in Scheme 2) were calculated. The results are listed in Table 4. From the thermodynamic point of view, it became evident that within the enzyme a phenolate anion and a neutral carboxylic acid unit are favored to the intuitively expected inverse deprotonation to a carboxylate and a phenol. The lowest energy gain is

Scheme 2a–d Suggested reaction pathway for the prenylation of 4-hydroxybenzoic acid with geranyl diphosphate including probable intermediates (all values in kcal mol^{-1}). The reaction starts with the deprotonation (activation) of the phenol by aspartate (**a**). Geranyl disphosphate is activated for diphosphate leaving as Mg^{2+} complex and attacked by the phenolates *ortho* carbon at C-1' (**b**) to form the σ complex from which the proton at C-3 is transferred to the oxygen of the diphosphate (**c**). The products of the reaction are formed and aspartate is regenerated after protonation of the product (**d**). In this scheme, a uniform numbering of the phenolic moiety was chosen so that the prenylation position is always assigned as C-3 although IUPAC nomenclature might be different, depending on the substitution pattern.

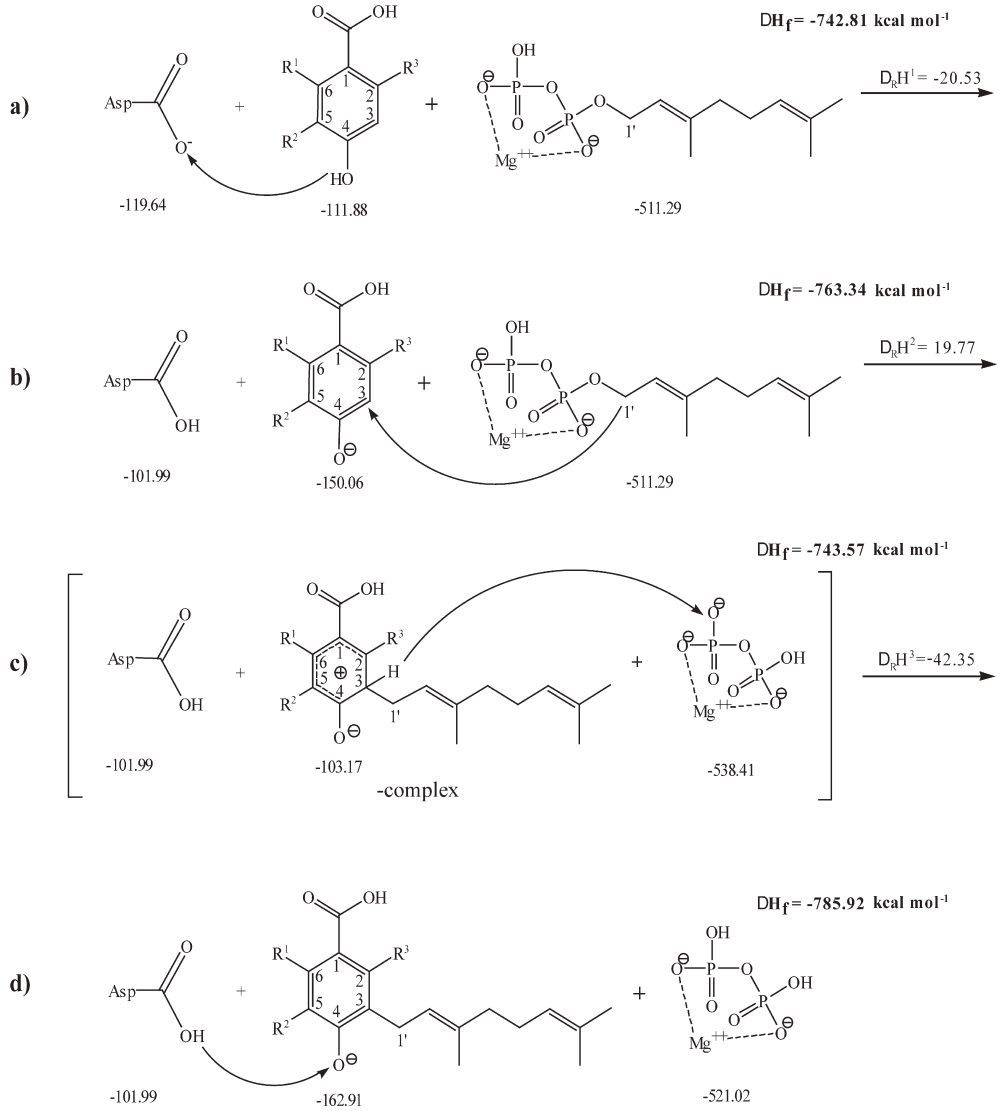


Table 4 Calculated $\Delta_R H$ and $\Delta_R H^{\#}$ values for benzoic acid substrates in kcal mol^{-1}

Compound	Educts (A)	Deprot (B)	Sigma (C)	Products (D)	$\Delta_R H^{\#}$	$\Delta_R H$	Exp. (%[a])
3-Chloro-4-hydroxybenzoic acid	−748.72	−773.17	−752.5	−795.57	−3.78	−46.85	21
2,4-Dihydroxybenzoic acid	−791.25	−813.7	−794.13	−835.83	−2.88	−44.58	20
2,3,4-Trihydroxybenzoic acid	−830.59	−855.85	−833.63	−877.82	−3.04	−47.23	9
4-Hydroxybenzoic acid	**−742.81**	**−763.34**	**−743.57**	**−785.92**	**−0.76**	**−43.11**	**100**
3-Nitro-4-hydroxybenzoic acid	*−753.03*	*−789.38*	*−749.16*	*−807.55*	*3.87*	*−54.52*	*0*
3-Chloro-4-aminobenzoic acid	*−706.11*	*−716.39*	*−700.38*	*−737.05*	*5.73*	*−30.94*	*0*
4-Hydroxyphthalic acid	*−827.43*	*−860.73*	*−827.75*	*−873.4*	*−0.32*	*−45.97*	*0*
2,4,6-Trihydroxybenzoic acid	*−838.31*	*−861.69*	*−833.4*	*−875.97*	*4.91*	*−37.66*	*0*

[a] Experimental conversions in % by HPLC / DAD [19]
The preferred substrate in vivo is displayed in bold. Substrates displayed above are also converted by the enzyme in experimental investigations. [18] Derivatives in italics are not converted by the enzyme, which coincides with the calculated results of $\Delta_R H^{\#}$, except for the 4-hydroxyphthalic acid which probably fails because of a different binding orientation. $\Delta_R H^{\#}$ values are defined as $\Delta_R H^{\#}$ = sigma(C) − educts (A)

calculated for 4-amino-3-chloro-benzoic acid (formation of an imide anion).

This result, however, is of immediate importance for the catalysis, since a deprotonation of the phenolic group enhances the negative charge at the *meta* position of the 4-hydroxybenzoic acid by mesomeric stabilization and therefore will considerably support the reaction of the positively charged prenyl C-1' with this position.

Finally, the energies of all intermediates of the pHB derivatives were calculated for the reaction as shown in Scheme 2. From the calculated energy difference between the σ complex and the educts it becomes obvious that the viable substrates are 4 to 9 kcal mol^{-1} favored in comparison to the derivatives which are not accepted as substrates. Since, the formation of prenylation products is thermodynamically favorable in all cases (last column in Table 4), the reactivity difference between substrates and non-substrates must have kinetic reasons caused by different transition state energies, or by steric interactions to fine too notice in our preliminary model. The special case of 4-hydroxyphthalic acid has already been discussed above.

Furthermore, despite the uncertainties of the 3D model structure, some aspects of the substrate specificity could be explained based on the model. Based on the semiempirical quantum mechanical calculations it could be shown that some experimentally tested hydroxybenzoic acid derivatives are unsuitable as substrates due to high activation barriers during the catalysis. It could be shown that the whole process is thermodynamically favored and most likely proceeds via an S_N1-type mechanism.

These results help to understand basic principals of the catalytic process exhibited by the *ubiA* enzyme and may serve as basis for the development of new ligands or inhibitors for this enzyme with its crucial importance for aerobic processes.

Discussion

A first 3D model of an oligoprenyltransferase has been developed by homology modeling based on an X-ray structure of the photosynthetic reaction center of *Rhodopseudomonas viridis* (1prc). Because of the very low amino acid identity, it is evident that this model is only a first crude approximation, hopefully with relevance for the real native structure, and thus has to be considered with care. Nevertheless, there is currently no better model available for aromatic prenyltransferases. The one presented here for the first time can help to understand principal mechanisms of substrate recognition and catalysis exhibited by this class of enzymes.

By means of multiple alignments, it could be shown that two putative active sites must be taken into consideration. Which of these, or if indeed both, are catalytically active must be validated by means of site-directed mutagenesis. However, in a review of Liang et al. [30] the effect of mutagenesis of the evolutionarily conserved aspartate residues in farnesyl diphosphate synthase is described. In a site-directed mutagensis experiment with yeast farnesyl diphosphate synthase, Song and Poulter [31], after a substitution of Asp by Ala, found k_{cat} values with rates 4–5 orders of magnitude slower than the wild type. In the second xDDxxD motif of yeast, the first and second Asp to Ala mutations resulted in k_{cat} values 4–5 orders of magnitude smaller. The third Asp of this motive seems to be less important to catalysis as its mutation to Ala only resulted in 6 to 16-fold lower k_{cat} values. [32] Based on these results, Liang et al. conclude that all Asp residues in the two xDDxxD motifs except the last one in the second motif are important for catalysis. Our investigations are indirectly justified by this and it can be assumed that 4-hydroxybenzoate oligoprenyltransferase of *E. coli* is characterized by a similar mode of catalysis. However, since the *ubiA* enzyme only needs one diphosphate binding site, only one is likely active. The rudimentary second active site is a strong hint for the evolutionary origin of this aromatic prenyltransferase. It may have common roots with chain-elongation prenyl diphosphate synthases, which always need at least one more prenyl diphosphate binding site.

References

1. Young IG, Leppik RA, Hamilton JA, Gibson F (1972) J Bacteriol 110:18–25
2. Nishimura K, Nakahigashi K, Inokuchi H (1992) J Bacteriol 174:5762
3. Siebert M, Bechthold A, Melzer M, May U, Berger U, Schröder G, Schröder J, Severin K, Heide L (1992) FEBS Lett 3:347–50
4. Nichols BP, Green JM (1992) J Bacteriol 16:5309–5316
5. Wessjohann LA, Sonntag B, Dessoy MA (1999) Enzymatic C–C coupling: the development of aromatic prenylation for organic synthesis. In: Diederichson U, Lindhorst TK, Westermann B, Wessjohann LA (eds) Bioorganic chemistry—reviews and perspectives. Wiley-VCH, Weinheim
6. SYBYL (2001) Tripos Associates Inc, 1699 S. Hanley Road, Suite 303, St. Louis, MO 63144, USA
7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
8. Weiner SJ, Kollman P, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) J Am Chem Soc 106:765–787
9. Clark M, Cramer III RD, van Opdenbosch N (1989) J Comput Chem 10:982–1012
10. Sippl MJ (1990) J Mol Biol 213:859–883
11. GOLD (1998) Copyright: Cambridge Crystallographic Data Centre, Cambridge, UK
12. Wavefunction Inc., 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612, USA
13. Altschul S, Madden T, Schäffer A, Zheng Zhang J, Miller W, Lipman D (1997) Nucleic Acids Res 25:3389–3402 (http://www.ncbi.nlm.nih.gov/blast/)
14. MOE (2001) Chemical Computing Group headquarters, Montreal, Canada
15. Insight (2001) Accelrys Inc., Molecular Simulations, San Diego, California, USA
16. Pearson, Lipmanog (1988) PNAS 4:2444–2448
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) J Mol Biol 215:403–410
18. http://www.dodo.cpmc.columbia.edu/predictprotein
19. http://www.sbg.bio.ic.ac.uk/~3dpssm/
20. Bairoch A, Apweiler R (2000) Nucleic Acids Res 28:45–48
21. Heide L, Melzer M (1994) Biochim Biophys Acta 1212:93–102
22. Ashby MN, Kutsunai SY, Ackermann S, Tzagoloff A, Edwards PA (1992) J Biol Chem 267:4128–4136
23. Wessjohann LA, Sonntag B (1996) Angew Chem Int Ed Engl 35 108:1821–1823
24. Levitt M (1992) J Mol Biol 226:507–533
25. Fechteler T, Dengler U, Schomberg D (1995) J Mol Biol 253:114–131
26. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Cryst 26:283–291

27. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267:727–748
28. Fulhorst M, Wessjohann LA (2004) Org Lett (submitted)
29. Sonntag B (1995) In: Wessjohann LA (ed) Diploma thesis "Untersuchungen zur Substratspezifität des Enzyms Polyprenyldiphosphat: 4-Hydroxybenzoat Polyprenyltransferase aus *E. coli* Überproduzenten". Ludwig-Maximilians-Universität München, Germany
30. Liang PH, Ko TP, Wang HJ (2002) Eur J Biochem 269:3339–3354
31. Song L, Poulter CD (1994) Proc Natl Acad Sci USA 91:3044–3048
32. Koyama T, Tajiama M, Sano H, Doi T, Koike-Takeshita A., Obata S, Nishino T, Ogura K (1996) Biochemistry 35:9533–9538