Oxidosqualene Cyclase Second-Sphere Residues Profoundly Influence the Product Profile

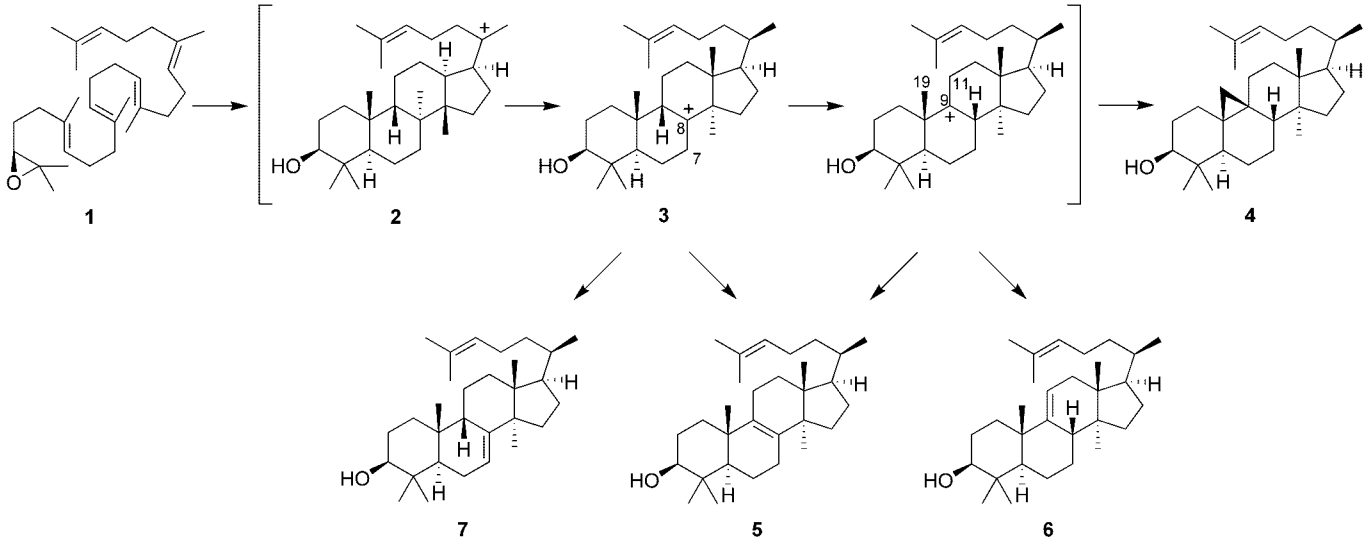
Silvia Lodeiro,[a] Michael J. R. Segura,[a] Martin Stahl,[b] Tanja Schulz-Gasch,*[b] and Seiichi P. T. Matsuda*[a]

Oxidosqualene cyclases convert oxidosqualene (**1**) to cyclic triterpene alcohols through cationic cyclization, rearrangement, and deprotonation reactions.[1] These enzymes control the reactivity of carbocations with a precision unrivalled by nonenzymatic catalysts, but how they utilize steric bulk and polar groups to guide carbocation reactivity remains poorly understood. Cycloartenol synthase is an oxidosqualene cyclase that cyclizes oxidosqualene to the protosteryl cation (**2**), guides rearrangement to the lanosteryl cation (**3**), and promotes specific deprotonation from C-19 to form cycloartenol (**4**; Scheme 1). Lanosterol synthase is a mechanistically related enzyme that catalyzes the same cyclization and rearrangement reactions, and abstracts a proton from C-8 to form the tetrasubstituted olefin in lanosterol (**5**). We describe herein mutagenesis experiments and computer modeling that establish that second-sphere oxidosqualene cyclase residues are a critical component of the catalytic distinction between cycloartenol synthase and lanosterol synthase.

[a] *S. Lodeiro, Dr. M. J. R. Segura, Prof. Dr. S. P. T. Matsuda*
Departments of Chemistry and Biochemistry and Cell Biology
Rice University
6100 S. Main Street, Houston, TX 77005 (USA)
Fax: (+1) 713-348-5154
E-mail: matsuda@rice.edu

[b] *Dr. M. Stahl, Dr. T. Schulz-Gasch*
F. Hoffmann–La Roche Ltd. Molecular Design, PRBD-CS 92/2.10D
4070 Basel (Switzerland)
Fax: (+41) 61-6886459
E-mail: tanja.schulz-gasch@roche.com

 DOI: 10.1002/cbic.200400086



Scheme 1. Ath*CAS1 mutants cyclize oxidosqualene (**1**) to cycloartenol (**4**), lanosterol (**5**), parkeol (**6**), and 9β-Δ7-lanosterol (**7**).*

Tyr410, His477, and Ile481 are catalytically important residues in *Arabidopsis thaliana* cycloartenol synthase (*Ath*CAS1).[2] These residues are strictly conserved in cycloartenol synthases, but animal and fungal lanosterol synthases maintain Thr, Cys or Gln, and Val at the corresponding positions (Figure 1). Mutating these residues in *Ath*CAS1 to the corresponding lanosterol synthase residues allows some degree of lanosterol biosynthesis.[3] The *Ath*CAS1 Ile481Val mutant produces 25% lanosterol, along with cycloartenol and parkeol (**6**; Table 1),[3a] and the *Ath*CAS1 Tyr410Thr mutant forms 65% lanosterol with 9β-lanosta-7,24-dien-3β-ol (9β-Δ7-lanosterol, **7**) and parkeol as by-products (Table 1).[3b] Of all the *Ath*CAS1 residues studied to date, His477 has the strongest effect on the product profile. The *Ath*CAS1 His477Asn mutant biosynthesizes lanosterol more accurately (88% lanosterol, 12% parkeol) than any other described *Ath*CAS1 mutant,[4] whereas the His477Gln mutation strongly favors parkeol formation (73% parkeol, 22% lanosterol, and 5% 9β-Δ7-lanosterol).[4]

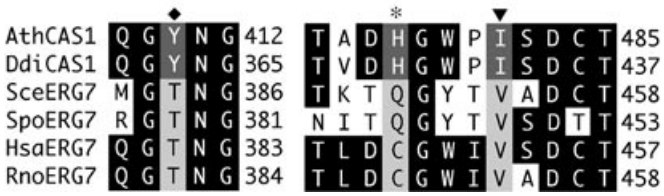


Figure 1. *Representative cycloartenol synthases from* Dictyostelium discoideum*[14] (DdiCAS1) and* A. thaliana *(Ath*CAS1*), with representative lanosterol synthases from the fungi* Saccharomyces cerevisiae*[15] (SceERG7) and* Schizosaccharomyces pombe*[16] (SpoERG7) and the animals* Homo sapiens*[17] (HsaERG7) and* Rattus norvegicus*[18] (RnoERG7). Tyr410 (◆), His477 (*), and Ile481 (▼) are labeled above the sequences.*

Table 1. *Products of* Ath*CAS1 mutants.*

*Ath*CAS1 mutant	**4**	**5**	**6**	**7**
Native	99%	0%	1%	0%
I481V	54%	25%	21%	0%
Y410T	0%	65%	2%	33%
H477N	0%	88%	12%	0%
H477Q	0%	22%	73%	5%
Y410T/I481V	0%	78%	<1%	22%
Y410T/H477N/I481V	0%	78%	0%	22%
Y410T/H477Q/I481V	0%	78%	0%	22%

Mutations at positions 410 and 481 have synergistic effects. The *Ath*CAS1 Tyr410Thr Ile481Val double mutant biosynthesizes lanosterol more accurately than either single mutant (Table 1).[3b,5] We anticipated that the three mutations that promote lanosterol biosynthesis (His477Asn, Tyr410Thr, and Ile481Val) would act synergistically to form lanosterol even more accurately than the Tyr410Thr Ile481Val double mutant. In contrast, we expected the parkeol-forming properties of the His477Gln mutant to diminish lanosterol biosynthesis in the Tyr410Thr Ile481Val double mutant background. We generated the triple mutants and expressed them in the yeast strain LHY4.[3a] In vitro assays coupled with GC-FID, GC-MS, and NMR analyses established that both mutants generate essentially the same product profile as the Tyr410Thr Ile481Val double mutant (Table 1). The His477Asn mutation does not increase lanosterol biosynthesis in the Tyr410Thr Ile481Val background; the triple mutant actually produces less lanosterol than does the His477Asn single mutant. Similarly, the Tyr410Thr and Ile481Val mutations completely abolish the influence of His477Gln, so that parkeol is not even produced as a measurable by-product. Although His477Asn and His477Gln mutations impart radically different catalytic properties on native *Ath*CAS1, this influence is preempted by the Tyr410Thr and Ile481Val mutations. This catalytic hierarchy is unusual. Previous studies show synergistic effects between catalytically relevant oxidosqualene cyclase mutations; combining catalytically relevant mutations generates a multiple mutant with catalytic properties distinct from either parent.[3b,6] We applied homology modeling studies to investigate how changing the structure

through mutations affects product specificity. The *Ath*CAS1 structure was modeled as described previously for the closely related (44% identical) human lanosterol synthase.[7] A key feature of the active site is a hydrogen-bonding network of residues that include Tyr118, His257, Tyr410, Asp483, Tyr532, Tyr616, Tyr734, and Tyr737 (Figure 2). These residues are strictly conserved in the known cycloartenol synthases, and all but Tyr410 are also present in the known lanosterol synthases. Tyr410 and Tyr532 are essential for cycloartenol biosynthesis;[3b,c] Asp483 and His257 correspond to the catalytically essential yeast lanosterol synthase residues that were proposed to be the active-site acid and base, respectively.[8] This hydrogen-bonding network would order the active-site structure and may shuttle a proton from C-19 to the active-site acid (Figure 2).

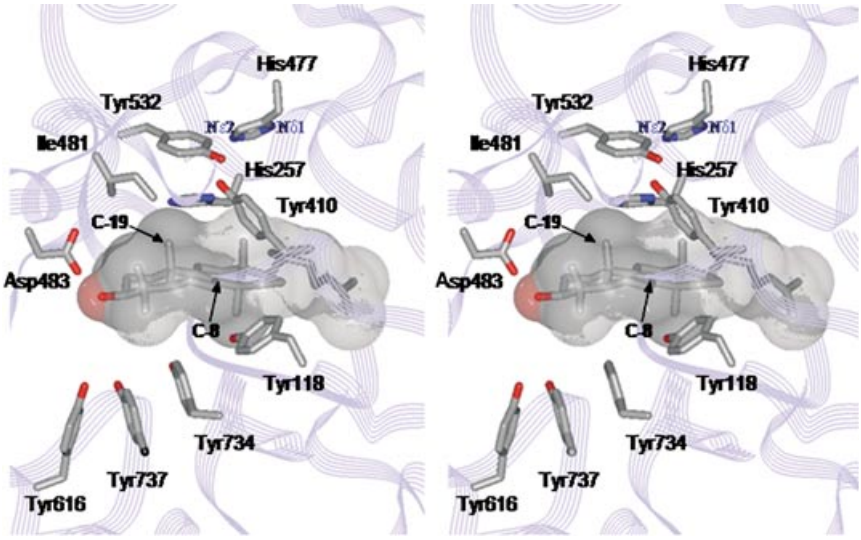


Figure 2. *Stereo representation of the* Ath*CAS1 homology model showing mutated residues and residues crucial for the deprotonation step.*

The Ile481 side chain is at the top of the active site and, by interacting with the A ring, it orients the substrate in the binding site to avoid early reaction termination.[3a,7] Lanosterol synthase has smaller residues at positions corresponding to *Ath*CAS1 Ile481 and Tyr410. The active site enlarges when these residues are mutated to their lanosterol synthase counterparts; this may compromise the mutants' ability to fix the position of the lanosteryl cation to promote deprotonation from C-19.[3a,b]

His257 and Tyr410 are a H-bonded pair positioned near the C-19 angular methyl group where deprotonation yields cycloartenol. Distances after molecular dynamics simulations (Moloc, Gerber Molecular Design) with constraint backbone and flexible active-site side chains were 2.5 Å for the Tyr410 oxygen and 2.7 Å for the His257 Nε2 to either hydrogen atom of the C-19 angular methyl group. Although Tyr410 appears to be slightly closer to accept the proton, His257 is not dramatically more distant. The difference in distances is relatively minor, and considerable error could arise from assumptions made in homology modeling and constraints during MD simulations. Either residue could be the base or could influence deprotonation by participating in the H-bonding network. The experimentally established change in deprotonation position in Tyr410 mutants is consistent with either possibility. If Tyr410 is the base, the Tyr410Thr mutation would abolish cycloartenol biosynthesis by removing the phenolic hydroxyl that would be the proton acceptor. Alternatively, the Tyr410Thr mutation could alter the hydrogen-bonding network and induce the proton acceptor His257 to change positions. The Tyr410Thr mutant remains a competent catalyst; this suggests that the change in H-bonding does not abolish reprotonation of Asp483.

His477 is not in the active site, but is a second-sphere residue that is hydrogen bonded to Tyr410 (Figure 2) and strongly affects its location and side-chain orientation. His477 is hydrogen bonded to a backbone carbonyl through Nδ1 (donor) and to Tyr410 via Nε2 (acceptor). Changes at His477 strongly affect the location, orientation, and electronics of the Tyr410 side chain. Mutating His477 to Asn keeps a nitrogen atom in the position of Nδ1 to form the hydrogen backbone interaction. However, the Asn carbonyl oxygen is farther away, approximately in the position of the Cδ2 of the native His. As a result, the hydrogen bond with Asn477 pulls Tyr410 away from the active site (Figure 3). Mutating His477 to Gln induces a less dramatic change. The carbonyl oxygen of Gln is located between the Nε2 of His477 in wild-type *Ath*CAS1 and the carbonyl oxygen of Asn in the His477Asn mutant. Changing the orientation of Tyr410 disturbs the hydrogen-bonding pattern with His257 and Tyr532, shifting the potential proton acceptors in these three residues (Figure 3). In addition, the steric changes induced by moving Tyr410 may affect product formation. Pulling Tyr410 only slightly out of the active site with Gln477 might allow sufficient mobility for the base to access the C-11 proton to form parkeol. Further reduction of steric hindrance through Tyr410 by Asn477 would further enlarge the active site; this might allow sufficient rotation of the intermediate cation that the C-8 proton would be accessible. Further mutagenesis studies are in progress to clarify if these mutants generate diverse products because one base acquires access to several protons or if multiple proton acceptors are involved. The His477Asn and His477Gln mutations are catalytically irrelevant in the Tyr410Thr Ile481Val background because they are outside of the active site and influence catalysis only through interaction with Tyr410. Mutating Tyr410 to Thr relocates the polar group away from position 477, and interrupts H-bonding,

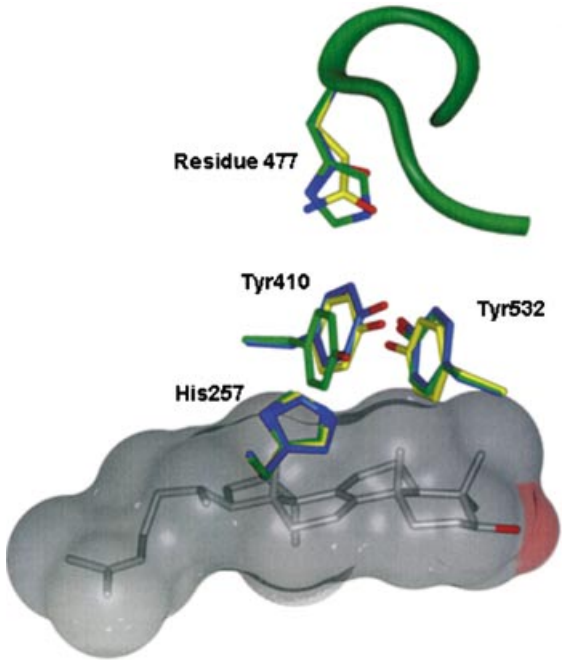


Figure 3. *Effects of mutations at position 477 on the orientation of the Tyr410 side chain (wild-type: green, His477Gln: yellow, His477Asn: blue). The lanosteryl cation is shown with a transparent Connolly surface. His477 mutants pull Tyr410 out of the active site and allow reorientation of the intermediate cation to form other products.*

so that these residues no longer influence each other's positions.

A previous effort to model *Ath*CAS1 threaded the sequence onto the *Alicyclobacillus acidocaldarius* squalene-hopene cyclase (*Aac*SHC) crystal structure. This model oriented the His477 side chain away from Tyr410 to mimic the *Aac*SHC positioning, and it consequently could not explain the observed effects of His477 mutations.[3d] Although the primary sequence alignment lacks inserts or gaps near His477, steric and electrostatic differences between residues require some reorientation of backbone atoms (this rearrangement does not further affect secondary structure). For this reason, the *Ath*CAS1 model generated with a fully automated procedure could not predict atoms in this region.

The influence of second-sphere residues on the oxidosqualene cyclase product profile might be a necessary adaptation to the unusual reactivity of the carbocationic intermediates that these enzymes handle. Their active sites must be constructed from a limited set of amino acids to preclude nucleophilic attack on an intermediate carbocation. In particular, these active sites are dominated by electron-rich but relatively unreactive[9] aromatic residues that might stabilize carbocations through π-interactions.[10] However, relatively few aromatic amino acids are available for natural protein construction, and this set may lack sufficient structural diversity to generate the subtle alterations in active-site topology needed to form the nearly 200 different cyclic products of the oxidosqualene cyclase/squalene cyclase family.[1b] The experiments and modeling describe herein show that second-sphere residues can generate diversity in active-site topology by interactions with active-site residues. The preponderance of Tyr in the active site may reflect the fact that its hydrogen-bonding potential makes it more responsive than Phe or Trp to neighboring residues; this can alter the position or electronic properties of the tyrosine by interaction with the phenolic hydroxyl group.

Experimental Section

The *Ath*CAS1 Tyr410Thr His477Asn Ile481Val and *Ath*CAS1 Tyr410Thr His477Gln Ile481Val triple mutants were constructed by oligo-directed mutagenesis by using ssDNA of the *Ath*CAS1 Tyr410Thr Ile481Val double mutant and following established protocols.[11] Mutants were expressed in the yeast strain LHY4, as described previously.[3b,6] Both recombinant yeast strains were grown on a 1 L scale in inducing medium (1% yeast extract, 2% peptone, 2% galactose, 13 mg L^{-1} heme) supplemented with ergosterol (20 mg L^{-1}), and in vitro enzymatic reactions were performed with racemic oxidosqualene as described.[12] The reactions were monitored by TLC, and the products were isolated and purified by silica gel chromatography. Products were identified by 500 MHz ^{1}H NMR analysis and by GC and GC-MS analysis of the trimethylsilyl (TMS) derivatives. Ratios were determined by GC-FID quantitation. The possibility that overlapping signals significantly distorted the numbers was precluded by quantitation using well-resolved ^{1}H NMR signals, which provided similar ratios. TMS-ethers of lanosterol and 9β-Δ7-lanosterol were previously found to comigrate on a Restek Rtx-5MS column under diverse GC conditions,[3b,4] but have now been successfully resolved by using a Rtx-35MS column (30 m, 0.25 mm id, 0.10 μm df). GC conditions: inlet and FID-detector, 280 °C; oven, 260 °C isothermal; split injection, ratio 40:1; column, helium 1 mL min^{-1} constant flow. Retention times: TMS-lanosterol, 15.9 min; TMS-9β-Δ7-lanosterol, 16.2 min.

Computational Methods

Molecular-modeling studies were performed on a Silicon Graphics Fuel R14000 Workstation. MOE 2003.04 (Chemical Computing Group Inc., Montréal, Québec, Canada) was used for sequence alignment and homology modeling. Sequence information on *Ath*CAS1 and *Aac*SHC was obtained from the Swiss-Prot database (P38605 and P33247, respectively). The *Aac*SHC crystal structure (PDB entry 2SQC, resolution 2.00 Å) obtained from the Protein Data Bank (PDB) was used as template for homology modeling. Mutations D376C and C435S were reverted to wild-type residues. All water and inhibitor molecules, as well as the B-chain were removed from the structure. Sequence alignment was performed by using the MOE sequence-alignment module and applying the BLOSUM 40 substitution matrix.[13] Conserved motifs and catalytic residues were found to align appropriately. The stereochemical quality of the final model structure was checked with the MOE protein report and manually refined. The mutations were manually introduced to the wild-type *Ath*CAS1 homology model. To relax the mutated structures, short MD simulations (100 ps, gas phase) followed by a minimization were performed with Moloc (Gerber Molecular Design). The cationic intermediate was manually docked to the active sites of the model structures by using Moloc.

Acknowledgements

We thank Dr. William K. Wilson for valuable guidance regarding GC-MS conditions. T.S.-G. and M.S. thank their colleagues from

Roche biostructure in Basel for stimulating discussions and a supportive research atmosphere. M.J.R.S. was supported by National Institutes of Health training grant T32-GM08362, held a Schlumberger Foundation Fellowship, and was a Houston Livestock Show and Rodeo Scholar. This research was funded by the National Science Foundation (MCB-0209769) and Robert A. Welch Foundation (C-1323).

Keywords: enzyme models · mutagenesis · natural products · protein design · steroids

[1] a) K. U. Wendt, G. E. Schulz, E. J. Corey, D. R. Liu, *Angew. Chemie* **2000**, *112*, 2930 – 2952; *Angew. Chem. Int. Ed.* **2000**, *39*, 2812 – 2833; b) R. Xu, G. C. Fazio, S. P. T. Matsuda, *Phytochemistry* **2004**, *65*, 261 – 290.
[2] E. J. Corey, S. P. T. Matsuda, B. Bartel, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11 628 – 11 632.
[3] a) E. A. Hart, L. Hua, L. B. Darr, W. K. Wilson, J. Pang, S. P. T. Matsuda, *J. Am. Chem. Soc.* **1999**, *121*, 9887 – 9888; b) J. B. R. Herrera, W. K. Wilson, S. P. T. Matsuda, *J. Am. Chem. Soc.* **2000**, *122*, 6765 – 6766; c) M. M. Meyer, R. Xu, S. P. T. Matsuda, *Org. Lett.* **2002**, *4*, 1395 – 1398; d) T. K. Wu, J. H. Griffin, *Biochemistry* **2002**, *41*, 8238 – 8244.
[4] M. J. R. Segura, S. Lodeiro, M. M. Meyer, A. J. Patel, S. P. T. Matsuda, *Org. Lett.* **2002**, *4*, 4459 – 4462.
[5] In previous work (ref. [3b]), the ratio of lanosterol and 9β-Δ7-lanosterol generated by the *Ath*CAS1 Tyr410Thr Ile481Val mutant was determined by ^{1}H NMR because of the difficulty of separating these compounds by GC. During the course of the present work, improved conditions were developed, and GC-FID ratios are shown in Table 1.
[6] M. M. Meyer, M. J. R. Segura, W. K. Wilson, S. P. T. Matsuda, *Angew. Chem.* **2000**, *112*, 4256 – 4258; *Angew. Chem. Int. Ed.* **2000**, *39*, 4090 – 4092.
[7] T. Schulz-Gasch, M. Stahl, *J. Comp. Chem.* **2003**, *24*, 741 – 753.
[8] E. J. Corey, H. Cheng, C. H. Baker, S. P. T. Matsuda, D. Li, X. Song, *J. Am. Chem. Soc.* **1997**, *119*, 1289 – 1296.
[9] a) K. U. Wendt, A. Lenhart, G. E. Schulz, *J. Mol. Biol.* **1999**, *286*, 175 – 187; b) K. U. Wendt, K. Poralla, G. E. Schulz, *Science* **1997**, *277*, 1811 – 1815; c) D. J. Reinert, G. Balliano, G. E. Schulz, *Chem. Biol.* **2004**, *11*, 121 – 126.
[10] D. A. Dougherty, *Science* **1996**, *271*, 163 – 168.
[11] F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, K. Struhl, *Current Protocols in Molecular Biology*, Wiley-Interscience, New York, **1999**.
[12] B. M. Joubert, L. Hua, S. P. T. Matsuda, *Org. Lett.* **2000**, *2*, 339 – 341.
[13] S. Henikoff, J. G. Henikoff, *Proteins* **1993**, *17*, 49 – 61.
[14] S. M. Godzina, M. A. Lovato, M. M. Meyer, K. A. Foster, W. K. Wilson, W. Gu, E. L. de Hostos, S. P. T. Matsuda, *Lipids* **2000**, *36*, 249 – 255.
[15] a) E. J. Corey, S. P. T. Matsuda, B. Bartel, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2211 – 2215; b) Z. Shi, C. J. Buntel, J. H. Griffin, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7370 – 7374.
[16] E. J. Corey, S. P. T. Matsuda, C. H. Baker, A. Y. Ting, H. Cheng, *Biochem. Biophys. Res. Commun.* **1996**, *219*, 327.
[17] a) C. H. Baker, S. P. T. Matsuda, D. R. Liu, E. J. Corey, *Biochem. Biophys. Res. Commun.* **1995**, *213*, 154; b) C.-K. Sung, M. Shibuya, U. Sankawa, Y. Ebizuka, *Biol. Pharm. Bull.* **1995**, *18*, 1459.
[18] a) M. Kusano, M. Shibuya, U. Sankawa, Y. Ebizuka, *Biol. Pharm. Bull.* **1995**, *18*, 195; b) I. Abe, G. D. Prestwich, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9274.

Received: March 31, 2004

