DOI: 10.1002/ejoc.201000366

Absolute Configuration and Docking Study of Canescensterone, a Potent Phytoecdysteroid, with Non-Lepidopteran Ecdysteroid Receptor Selectivity

Han Sun,[a][‡] Laurence Dinan,[b] René Lafont,[c] Apichart Suksamrarn,[d] Christian Griesinger,[a] Uwe Reinscheid,*[a] and Silvia Lapenna*[e][‡][‡‡]

Keywords: Steroids / Receptors / NMR spectroscopy / Structure–activity relationships / Molecular docking

The rare phytoecdysteroid canescensterone is a potent activator of both natural and engineered ecdysteroid receptor (EcR) gene-expression systems. This steroid also shows a remarkable selectivity towards non-lepidopteran compared with lepidopteran EcRs in gene-switch assays, a property that makes it a novel candidate lead compound for use in insecticide discovery and gene-switch design. Canescensterone possesses an unusual ecdysteroid structure, including a pyrrole-2-carboxylate group attached to C-24, a moiety not yet found in any other natural ecdysteroid. We have investigated the ambiguous stereochemistry at C-24 of this steroid by 1D and 2D solution NMR experiments, which resulted in the assignment of the 24*R*-configuration and determination of the side-chain conformation. The complete stereochemical assignment allowed the generation of docking models of canescensterone in both a lepidopteran EcR and a non-lepidopteran EcR, which helps to explain its selective biological activity.

Introduction

Phytoecdysteroids are plant steroids chemically related to the insect ecdysteroid hormone 20-hydroxyecdysone (20E; **1**, Figure 1), which coordinates major events in insect development, notably embryogenesis, moulting and metamorphosis.[1] Structurally, they are characterised by a high degree of hydroxylation (typically at the 2-, 3-, 14-, 20-, 22- and 25-positions), an α,β-unsaturated ketone at the 6/7-position and a *cis*-A/B-ring junction. Ecdysteroid hormonal actions are triggered by the recognition of hormonally active molecules by ecdysteroid receptors (EcRs), nuclear transcription factors found in arthropods and probably other invertebrate phyla too.[2] 20E is the most common moulting hormone among arthropods and also the most widespread plant-derived ecdysteroid (phytoecdysteroid). Several hundred 20E related analogues have been identified from plants,[3] where they accumulate as part of defence strategies against arthropod predators through the disruption of moulting hormone activity on ingestion or by imparting anti-feedant activity.[4]

As a consequence of their ability to mimic the actions of arthropod moulting hormones while being non-toxic to vertebrates, phytoecdysteroids are interesting lead compounds for the development of environmentally friendly invertebrate pest control agents.[5] Furthermore, steroidal and non-steroidal ligands for natural and mutated EcRs have been used successfully in EcR-based inducible gene expression systems for the modulation of the expression of a gene of interest within a host cell[6] and animal models.[7] Gene-switch technology finds applications in several fields of the life sciences, including functional genomics, drug discovery and gene therapy.[6a,8] Because neither ecdysteroids nor their cognate receptors are synthesised in vertebrates, EcR-based gene switches are attractive systems for the regulation of genes of interest such as therapeutic genes in vertebrates, including humans.[8]

Over 200 ecdysteroids and related compounds have been screened for ecdysteroid activity to date, primarily by means of the fruit fly (*Drosophila melanogaster*) B_{II} cell assay.[9] Among them, canescensterone (**2**; Figure 1) is distinguished by its very high potency ($EC_{50} = 5.2 \times 10^{-10}$ M). The B_{II} cellular activity of **2** is similar to that of the most active ecdysteroid known to date, ponasterone A (PoA, 25-deoxy20E; $EC_{50} = 2.3 \times 10^{-10}$ M), and almost 15-fold higher than that of 20E itself ($EC_{50} = 7.5 \times 10^{-9}$ M). Several phy-

[a] Max Planck Institute for Biophysical Chemistry, NMR II, 37077 Göttingen, Germany
Fax: +49-551-201-2202
E-mail: urei@nmr.mpibpc.mpg.de
[b] Degree Centre Weymouth, Weymouth College, Cranford Avenue, Weymouth, Dorset, DT4 7LQ, UK
[c] BIOSIPE, ER3, Université Pierre & Marie Curie, Case 29, 75252 Paris cedex 05, France
[d] Department of Chemistry, Faculty of Science, Ramkhamhaeng University, Thailand
[e] Dipartimento Farmaco Chimico Tecnologico, University of Siena, Siena, Italy
[‡] These authors contributed equally to this work.
[‡‡] Current Address: European Commission Joint Research Centre, Institute for Health and Consumer Protection, Computational Toxicology and Biostatistics, TP 202, 21027 Ispra (Va), Italy
Fax: +39-0332-78-9725
E-mail: slapenna@gmail.com
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/ejoc.201000366.

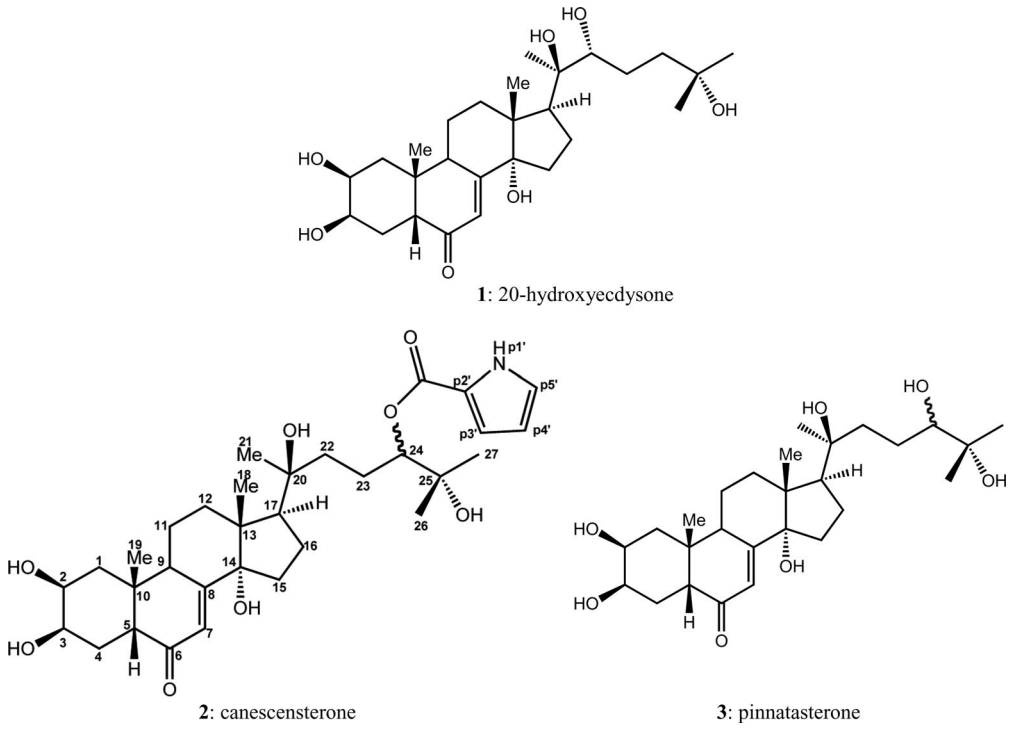


Figure 1. Structures of canescensterone and related analogues.

toecdysteroids, including PoA and muristerone A (5β,11α-dihydroxyPoA), have been previously established as effective EcR-based gene-switch actuators in mouse models using non-therapeutic reporter genes.[7b,10] However, these ecdysteroids are simply non-selective across EcRs from different species. In contrast, canescensterone shows a remarkable selectivity for non-lepidopteran EcRs, with an approximately 50-fold increase in potency and almost 5000-fold increase in efficacy over lepidopteran EcRs (moths and butterflies) when measured against a set of 10 EcRs representing nine arthropod species in a common gene-switch format.[11]

Canescensterone is a rare phytoecdysteroid isolated from the bark of *Vitex canescens*, a tree endemic to Thailand.[12] This compound possesses an unusual ecdysteroid structure insofar as it lacks the typical hydroxy group at the 22-position, which usually enhances activity, whereas it bears a pyrrole-2-carboxylate group attached to C-24, a moiety not yet found in any other natural ecdysteroid. The unexpectedly high hormonal activity of canescensterone has been associated with the presence of the pyrrole-2-carboxylate substituent at C-24.[13] However, the stereochemistry at C-24 has been ambiguous until now and consequently its structure–activity relationship has remained unclear. Because configurations can only be determined by NMR spectroscopy when the conformation is known, the main difficulty in determining the configuration was that C-24 is in a potentially flexible side-chain and thus several conformations have to be taken into account. Moreover, the very limited amount of this phytoecdysteroid from natural sources, the arduous chromatographic isolation and a challenging synthetic chemistry have further hampered its structural investigation. In recent studies, a residual dipolar coupling (RDC)-based approach has been used for the stereochemical analysis of some natural products.[14] However, in the case of canescensterone, the small amount of available sample prohibited the measurement of RDCs. Therefore we investigated the ambiguous stereochemistry at C-24 of canescensterone by robust and established 1D and 2D solution NMR experiments.

Results and Discussion

Canescensterone (**2**; Figure 1) is the 24-*O*-(pyrrole-2-carboxylate) ester of pinnatasterone (**3**; Figure 1), a minor ecdysteroid from *Vitex pinnata*.[24] Upon hydrolysis with guanidine, canescensterone can be transformed into pinnatasterone,[12] which indicates the same stereochemistry of the steroid ring system and the side-chain at C-17. In the *Musca domestica* bioassay, canescensterone showed higher activity than pinnatasterone, which indicates the importance of the substituent. Moreover, dependent on the side-chain, the stereochemistry at C-24 can influence the biological activity, as was shown for 24-*epi*-pterosterone which is seven-fold less active than pterosterone.[25] However, the phytoecdysteroids with a simple 24-OH substitution identified so far[3] provide no clear precedent for a stereochemical assignment at C-24, as exemplified by three pairs of epimers: pinnatasterone (**3**) and 24-*epi*-pinnatasterone, abutasterone and 24-*epi*-abutasterone, and pterosterone and 24-*epi*-pterosterone.[25]

An incomplete assignment of canescensterone in [D_6]pyridine has previously been published by Suksamrarn et al.[12]

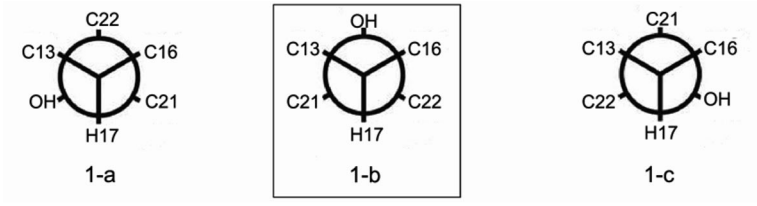
To determine the absolute configuration of C-24, it was necessary to assign the molecule completely, including the side-chain. This was achieved by a combination of 1D ^{1}H NMR, 2D ^{1}H–^{1}H P.E.COSY, ^{1}H–^{1}H NOESY, ^{1}H–^{1}H TOCSY, ^{1}H–^{13}C HSQC and ^{1}H–^{13}C HMBC spectra, which were measured in DMSO using less than 1 mg of the compound. The ^{1}H and ^{13}C NMR chemical shifts obtained for canescensterone (**2**) are listed in Table S1 of the Supporting Information).

Stereochemical NMR Analysis

There are two stereocentres in the acyclic side-chain separated by three single rotatable bonds. It is already known that the stereocentre at C-20 is *S*, but the absolute configuration of ecdysteroids with an identical pattern of substituents has not been determined. In this study we addressed the absolute configuration at C-24. To transfer the stereochemical information from C-20 to C-24, a *J*- and NOE-based conformational analysis was performed. Because the coupling constants separated by three bonds ($^3J_{HH}$ or $^3J_{CH}$) are directly related to their dihedral angles they are particularly useful for this purpose. We extracted homonuclear $^3J_{HH}$ coupling constants from a P.E. COSY spectrum. Owing to the very limited amounts of the compound, heteronuclear long-range $^3J_{CH}$ couplings at natural abundance could not be determined accurately and therefore only a qualitative interpretation of the coupling constants could be obtained. Strong signals in the ^{13}C HMBC spectrum indicate large $^3J_{CH}$ coupling constants whereas weak signals or no signals suggest small $^3J_{CH}$ coupling constants. Thus, with the help of reference coupling constants, we could decide whether the respective atoms are *anti* or *gauche* in the staggered rotamers. Additional support for stereochemical assignments came from NOESY data. In this respect, NOE integrals, usually categorised as strong, medium or weak, were considered. The predictions were then compared with model structures. The side-chain of canescensterone consists of five rotatable bonds. To determine the main conformation of the side-chain, a *J*- and NOE-based conformational analysis was extensively used for each bond. The predicted main conformers of the staggered form of each C_2 fragment along with the *J* couplings and NOESY data are depicted in Figures 2, 4 and 5.

In the C17–C20 fragment, the absolute configurations of C-17 and C-20 were already known. Within the fixed absolute configuration of a C_2 fragment there are three possible conformations (Figure 2). Because 17-H has no HMBC signal to C-21 and C-22, it was deduced that 17-H is in a *gauche* position with respect to C-21 and C-22. By this argument, only conformer **1-b** was possible, representing the *trans* orientation between C-13 and C-22. Further support for conformer **1-b** could be found in the NOESY spectrum; 21-H has a strong NOE with 17-H, the intensity of which is comparable to that of the NOE between 1-He and 18-H. Atoms 18-H and 1-He are in a *gauche* conformation with respect to each other and both atoms are a part of the rigid ring system. Furthermore, there are strong NOEs between 21-H/12-He and 21-H/12-Ha such that the possibility of conformation **1-a** could be excluded. The expected NOEs presented in Figure 3 show clearly that if conformation **1-a** were the major one, the strong NOEs should not be seen.

For the C20–C22 fragment there are three conformations possible. Owing to the prochiral protons at C-22, 22-Ha and 22-Hb, with 22-Ha as the proton at low field and 22-Hb at high field, two assignments are possible for each of the conformations. The Newman projections in the top half of Figure 4 differ from those in the bottom half by exchanging the assignment of these two prochiral protons. The HMBC spectrum shows that C-17 has a much larger coupling to 22-Hb than to 22-Ha based on the intensity of the respective cross-peaks. The intensity of the signal between C-17 and 22-Hb is comparable to that of the cross-peaks between C-17 and 15-Hα. The bond between C-17 and 15-H is part of a rigid ring system and the two atoms are nearly in an antiperiplanar conformation relative to each other. This means that C-17 and 22-Hb are also antiperiplanar in the staggered rotamers. From this observation, only conformers **2-b** and **2-d** are possible. Furthermore, C-21 shows very weak cross-peaks to both 22-Hb and 22-Ha, and 21-OH shows a larger NOE to 22-Hb than to 22-Ha.



Fragment	$^3J_{HH}$ (Hz)	Selected $^3J_{CH}$	Selected NOEs
C17–C20	–	H17–C21 (n)	H21–H17 (s)
	–	H17–C22 (n)	H21–H12a (s)
	–	–	H21–H12e (s)

Figure 2. Three possible conformers of the C17–C20 fragment. The framed conformer is the determined main conformation. Selected $^3J_{CH}$ couplings and NOE integrals in the table below are qualitatively classified: no signals (n) and strong signals (s).

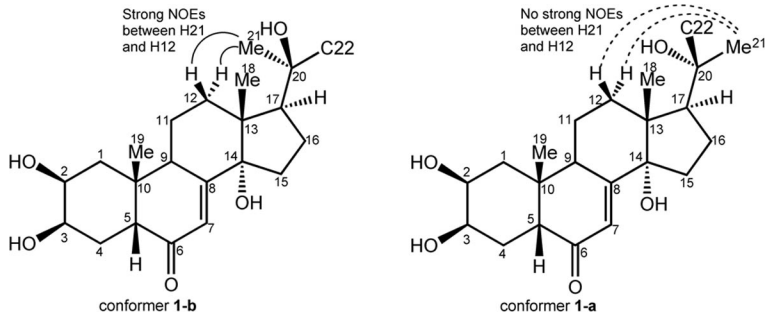
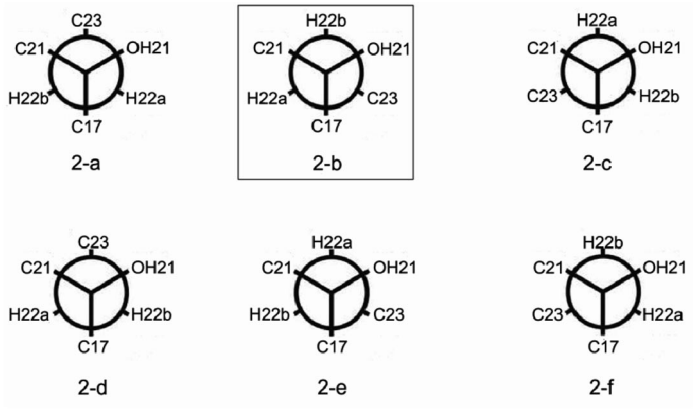


Figure 3. Expected NOE signals between 21-H and 12-H in conformers **1-b** and **1-a**. The strong NOE integrals in the NOESY spectrum additionally support our conclusion that the conformer **1-b** is the major conformer of the C17–C20 fragment.



Fragment	$^3J_{HH}$ (Hz)	Selected $^3J_{CH}$	Selected NOEs
C20–C22	–	C17–H22b (s)	OH–H22b (s)
	–	C17–H22a (w)	OH–H22a (w)
	–	C21–H22a (w)	H22a–H17 (s)
	–	C21–H22b (w)	–

Figure 4. Possible conformers of the C20–C22 fragment. The Newman projections on top differ from those at the bottom by exchange of the assignments of the prochiral protons. The framed conformer is the determined main conformation. Selected $^3J_{CH}$ couplings and NOE integrals in the table below are qualitatively classified: weak signals (w) and strong signals (s).

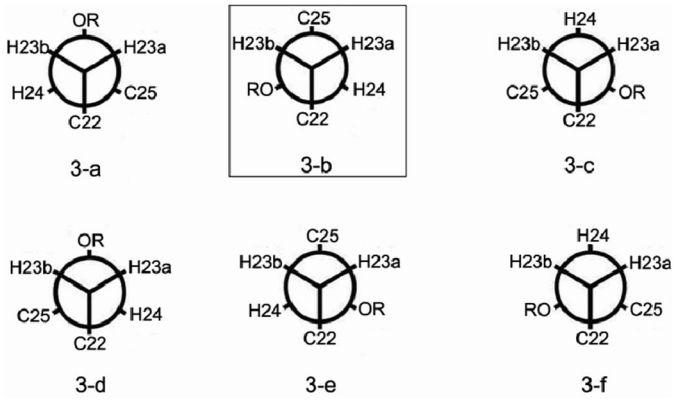
This supports the idea that **2-b** is the major conformer and therefore 22-Ha, with a chemical shift of 1.367 ppm, was assigned as pro-*R* and 22-Hb as pro-*S*. Based on the C-17/22-Ha torsional angle of +60° and the 17-H/C-22 torsional angle of –60°, a very strong NOE cross-peak between 22-Ha and 17-H was expected and observed (see below). In this conformation the methyl group at C-21 is *trans* to C-23.

With the assignment of the prochiral protons at C-22, the prochiral protons at C-23 could subsequently be assigned. As in the previous step, three conformations for each prochiral assignment are possible. The $^3J_{HH}$ values of the C-22 and C-23 atoms determined from the P.E.COSY spectrum are reported in Table 1. The $^3J_{HH}$ values observed between 22-Ha and 23-Hb, and 22-Hb and 23-Ha are in the 13.3–13.4 Hz range whereas the values between 22-Ha and 23-Ha, and 22-Hb and 23-Hb are in the range of 4.10–4.30 Hz. This clearly suggests a major conformer with a *trans* orientation of the carbon substituents at C-20 and C-24, which was further supported by two other sets of NMR data: first, C-20 has neither an HMBC cross-peak with 23-Ha nor with 23-Hb and, secondly, 23-Ha has a larger NOE with 22-Ha than with 22-Hb. In this way, 23-Ha, with a chemical shift of 1.768 ppm, was assigned as pro-*R* and 23-Hb as pro-*S*.

Table 1. NMR analysis of various fragments.[a]

Fragment	$^3J_{HH}$ (Hz)	Selected $^3J_{CH}$	Selected NOEs
C22–C23	$^3J_{H22a\text{-}H23a}$ = 4.25	C20–H23a (n)	H23a–H22a (s)
	$^3J_{H22a\text{-}H23b}$ = 13.33	C20–H23b (n)	H23a–H22b (w)
	$^3J_{H22b\text{-}H23a}$ = 13.40	–	–
	$^3J_{H22b\text{-}H23b}$ = 4.15	–	–
C24–C25	–	C26–H24 (s)	H24–H27 (s)
	–	C27–H24 (w)	H24–H26 (w)
	–	–	OH–H23a (s)
	–	–	OH–H23b (s)
	–	–	OH–H24 (s)

[a] Selected $^3J_{CH}$ couplings and NOE integrals in the table are qualitatively classified: no signals (n), weak signals (w) and strong signals (s).



Fragment	$^3J_{HH}$ (Hz)	Selected $^3J_{CH}$	Selected NOEs
C23–C24	$^3J_{H23a–H24}$ = 1.54	C25–H23a (n)	–
	$^3J_{H23b–H24}$ = 10.96	C25–H23b (n)	–
	–	C22–H24 (m)	–

Figure 5. Possible conformers of the C23–C24 fragment. The Newman projections on top differ from those below by exchange of the assignments of the prochiral protons. The framed conformer is the determined main conformation. Selected $^3J_{CH}$ couplings and NOE integrals in the table below are qualitatively classified: no signals (n), medium–strong signal (m).

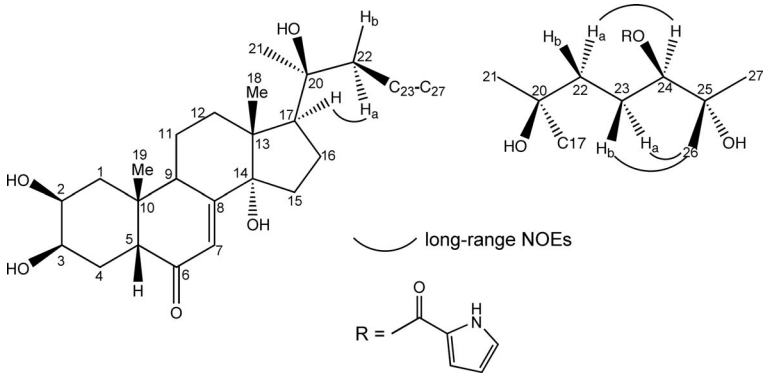


Figure 6. The main conformation of the side-chain of canescensterone, as determined by extensive *J* and short-range NOE analyses, is further corroborated by the presence of several long-range NOEs.

In the analysis of the C23–C24 fragment, six conformers again had to be compared (Figure 5). Because C-25 shows no HMBC cross-peak to 23-Ha and 23-Hb, conformers **3-b** and **3-e** are possible. Because 24-H has a small homonuclear coupling constant with 23-Ha ($^3J_{HH}$ = 1.54 Hz) and a large coupling with 23-Hb ($^3J_{HH}$ = 10.96 Hz), it was deduced that conformer **3-b** is the major one, again with a *trans* orientation of the carbon substituents at C-22 and C-25. In addition, there is a minor conformer **3-c**, which can be seen from the existence of a medium HMBC cross-peak between 24-H and C-22. From this information it was concluded that the absolute configuration at C-24 is *R*.

By using long-range NOEs, the two prochiral methyl groups at C-25 could be assigned. On the basis of the NMR results reported in Table 1, the dominating conformer of the C24–C25 fragment has a *trans* orientation between the oxygen substituents at C-24 and C-25. NOEs from the pyrrole substituent to protons of the side-chain were not detected, likewise the HMBC signal between 24-H and COO (ester) indicated no conformational preferences for the pyrrole-2-carboxylate substituent.

Based on the above conclusion, the major conformer of canescensterone with the correct configuration *R* at the stereocentre C-24 is shown in Figure 6. Long-range NOEs cross-validated this conformer: 1) two strong long-range NOEs between 17-H and 22-Ha, and 22-Ha and 24-H and 2) 23-Ha and 23-Hb have considerably stronger NOEs with 26-H than with 27-H. These two arguments give additional support to the proposed configuration of the side-chain of canescensterone.

Configuration of the Side-Chain

Ecdysone, 20E and several other members of the ecdysteroid class possess a high degree of hydroxylation which supports the formation of a water network. This was shown

in the X-ray analysis of suitable crystals.[26] The consequences are at least two-fold. 1) The biological activity is modulated, as exemplified by a thermodynamic analysis of the binding of 20E to the ecdysteroid receptor EcR–Ultraspiracle complex.[20a] Comparison with the complex formed with PoA, which only differs from 20E by the absence of the 25-OH group, shows the gain in the binding energy through the 25-OH group in 20E is counterbalanced by its larger desolvation cost.[20a] This interpretation may explain the higher affinity of PoA compared with 20E. 2) The stereochemical analysis of the side-chain might be heavily influenced by the presence or absence of substituents, in conjunction with the solvent used. In our experiments DMSO might favour a dominating conformation of the side-chain, thereby facilitating the transfer of stereoinformation. In addition, the unusual pyrrole-2-carboxylate substituent masks one of the OH groups, but at the same time favours sterically and stereoelectronically the *trans* orientation of the oxygen substituents at C-24 and C-25. Although we have identified only a small number of structural factors, the side-chain appears to be conformationally rigid enabling our type of configurational analysis. This is promising for a number of steroids with side-chains of unknown stereochemistry that might also be more rigid than anticipated.

In canescensterone there are two methylene groups between C-24 and the tetracyclic ring of the steroid. Interestingly, a literature search revealed that the side-chain configuration for a number of ecdysteroids, especially with two methylene groups in the side-chain, has only been assigned tentatively.[24,27] In the past, differences in chemical shifts were used to identify the configuration at C-24.[28] However, some ecdysteroids show negligible differences in their ^{13}C chemical shifts even though they represent epimers (e.g., 20,26-dihydroxyecdysone).[29] On the basis of our study we therefore anticipate that by a similar approach to that shown for canescensterone, the configuration at C-24 should be determined unambiguously. Another class of molecules with similar constitution and undetermined configuration are cucurbitanes, triterpenoids typically found within *Cucurbitaceae*. Spectral comparisons with literature data were used to identify the configurations of a number of cucurbitane glycosides with side-chains similar to canescensterone.[30] However, at least in the case of two compounds (Siamenoside I and 11-oxomogroside V), reference to Kasai et al.[31] does not allow an unambiguous assignment of the configuration at C-24. Again, a similar approach to that described for canescensterone should resolve the ambiguity.

Models of the Ecdysteroid Receptor–Canescensterone Complex

Canescensterone diverges from most 20E analogues, which are generally non-selective towards insect species, in that it favours non-lepidopteran EcRs. In a multi-receptor gene-switch screening employing 10 different EcR ligand-binding domains (LBDs) from nine arthropod species, this phytoecdysteroid showed an approximately constant increase in responsiveness towards the six non-lepidopteran EcRs over the four lepidopteran EcRs studied, both in terms of potency (EC_{50}) and efficacy [RMFI (maximum fold induction relative to a gene-switch activator standard)].[11] In particular, the highest responsiveness was obtained with the EcR from the homopteran *Bemisia argentifolii* (Ba; silverleaf whitefly) for which EC_{50} = 0.2 μM and RMFI = 4.881 (cf. $EC_{50} > 10$ μM and RMFI ≈ 0.001 with lepidopteran EcRs). The screened lepidopteran receptors include a variant EcR from the lepidopteran *Choristoneura fumiferana* (Cf; spruce budworm), dubbed VY-CfEcR.[18,32]

The complete stereochemical assignment of canescensterone allowed a docking study to be undertaken aimed at unravelling the molecular basis of canescensterone selectivity. BaEcR and VY-CfEcR are suitable representative proteins for this task, not only because of the availability of homogeneous activity data with both these receptors,[11] but also because of the existence of X-ray crystallography data for identical or closely related EcR LBDs complexed to an ecdysteroid analogue, that is, the EcR LBDs from the hemipteran *Bemisia tabaci* (Bt; sweet potato whitefly) complexed with PoA,[19] which is identical to the BaEcR LBD, and the lepidopteran *Heliotis virescens* (Hv; tobacco budworm) complexed with PoA[20b] and 20E,[20a] which differs for two residues from the VY-CfEcR LBD. The binding pose of the ligand is essentially identical in all the ecdysteroid-bound EcR crystal structures deciphered to date.[19,20,33] Moreover, the pattern of ligand–receptor interactions is maintained, including the hydrogen bonds involving 2-OH, 3-OH, 14-OH, 20-OH, 22-OH and, in the case of 20E, 25-OH. Previous canescensterone docking models were generated by manually superimposing both 24-epimers of the steroid (because the stereochemistry at C-24 was unknown) on the crystal structure of PoA in its EcR-bound conformation.[11] Herein, we report a computer-aided molecular docking study of canescensterone using the 24*R*-epimer as the correct formula (**2**).

Docking calculations were performed by using the GOLD genetic algorithm for the ligand **2** in both BaEcR LBD and VY-CfEcR LBD. Notably, for each of these complexes, the docked ligand resulted in substantially the same orientation as 20E and PoA in their respective EcR-bound crystal structures (Figure 7, I and II). Thus, the GOLD algorithm was able to identify the correct steroid-binding site of the EcR and detected the same binding pose for the steroid core of canescensterone as for 20E and PoA.

The binding pocket in the docked complexes is characterised by a hydrophobic region (Figure 7, orange contours) that embraces the steroid ring system of **2** and extends towards the ligand side-chain. The 24-pyrrole-2-carboxylate group overlaps with the extended C20–C27 side-chain of the PoA or 20E crystal structures with the pyrrole ring in van der Waals contact with Ba-W445 and C427 (VY-Cf-W521 and C503, respectively). Thus, the docking pose of **2** differs from the 20E and PoA crystal structures with respect to the orientation of the C25–C27 steroid tail as the relatively bulky pyrrole-2-carboxylate group of **2** is best accom-

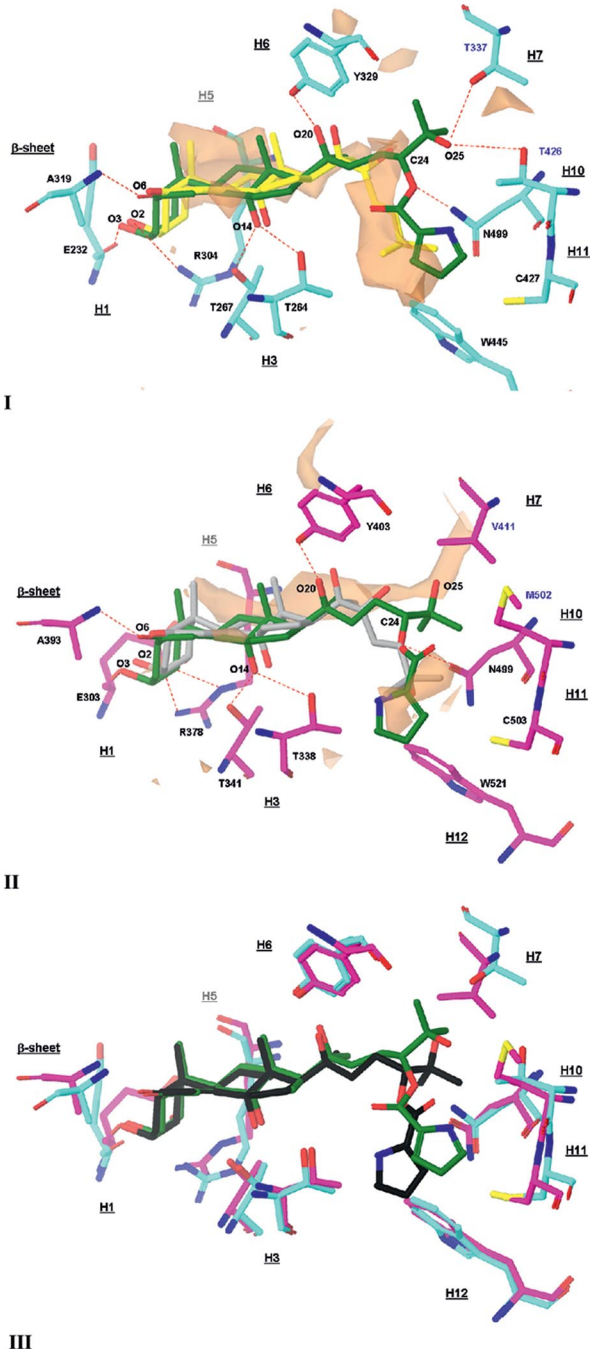


Figure 7. Docking models of (24*R*)-canescensterone (**2**; carbon atoms in green) in the binding cavity of I) the homopteran BaEcR (carbon atoms printed in cyan) superimposed on the BtEcR-bound ponasterone A crystal structure (carbon atoms printed in yellow) and II) the lepidopteran VY-CfEcR (carbon atoms printed in purple) superimposed on the HvEcR-bound 20-hydroxyecdysone crystal structure (carbon atoms printed in grey). Part III is a superimposition of part I and part II (in part III, the VY-CfEcR-docked (24*R*)-canescensterone is represented by carbon atoms printed in black). Atom colour code: red for oxygen, blue for nitrogen and grey for sulfur. Proximal secondary structure designations are in black and more distant ones are in gray. Only selected residues within 4.0 Å of **2** that are important for the binding are displayed. Dashed lines indicate hydrogen bonds between **2** and amino acid residues. Blue labels refer to residues that differ between the two EcRs. The hydrophobic surface (orange contours) of the binding cavity of each receptor is displayed (isovalue: 0.83).

modated in the ample and hydrophobic region of the pocket formed by portions of helices 10–12.

The pattern of hydrogen-bonding interactions between the steroid ring system (2-OH, 3-OH, 14-OH), the 20-OH and the C-6 carbonyl of **2** and each of the two docked receptors is similar to those observed in the EcR-bound 20E and PoA crystal structures (cf. hydrogen-bonded residues Ba-A319, E232, R304, T267, T264, Y329 and A319). Overall, the major distinction between the docked BaEcR and VY-CfEcR complexes involves the region in contact with the 25-OH of **2**. In the BaEcR complex, this group points towards an hydrophilic area formed by Ba-T337 (7-H) and T426 (11-H), with which it may form one or possibly two hydrogen bonds (3.039 and 3.420 Å between the 25-OH of **2** and the T337 and T426 hydroxy group oxygen atoms, respectively). In contrast, when the VY-CfEcR binding pocket is presented with the 25-OH of **2**, analogous hydrogen bonds are simply not possible owing to the lipophilic, non-hydrogen-bonding side-chains of the corresponding residues, Val411 and Met502 (Figure 7, II). Notably, Ba-T337 is conserved in several non-lepidopteran EcRs (Bt-Thr304, *Aedes aegypti* Thr458, *Tenebrio molitor* Thr377, *Tribolium castaneum* Thr435) or substituted by other polar amino acids (cf. *Drosophila melanogaster* Asn540, *Amblyomma americanum* Ser459), but not found in Lepidoptera in which a valine residue is strictly conserved at this position (e.g., Cf-V411, Hv-V416, *Bombyx mori* V475, *Manduca sexta* V417).[11,34]

A third hydrogen bond can be formed between the ester sp^3 oxygen of **2** and the side-chain amide nitrogen of Ba-N423 (distance O–N: 3.386 Å). Similarly, in the VY-CfEcR complex, a hydrogen bond may form between the ester sp^3 oxygen of **2** and the side-chain amide oxygen of VY-Cf-N499 (distance O–O: 3.485 Å). In the 20E-HvEcR complex crystal structure, the side-chain amide oxygen of the corresponding asparagine residue (Hv-N504) is hydrogen-bonded to the 25-OH of 20E.[20a] However, as this residue is strictly conserved across insect species, the additional hydrogen-bonding contribution exploited by the GOLD program for docking **2** into each of the two complexes does not help explain canescensterone discrimination between lepidopteran and non-lepidopteran EcRs.

In summary, the (24*R*)-canescensterone docking models in the representative non-lepidopteran BaEcR (highly responsive) and lepidopteran VY-CfEcR (poorly responsive) suggest binding poses in which the pyrrole-2-carboxylate group is propitiously located in an ample hydrophobic region of the binding site of both receptors. Furthermore, binding to the BaEcR exploits the favourable contribution from hydrogen-bonding entities pointing towards the 25-OH of **2**, whereas VY-CfEcR disadvantageously juxtaposes lipophilic residues on this group (Figure 7, III).

The previous models of (24*S*)- and (24*R*)-canescensterone[11] were generated on the principle of minimal perturbation of PoA in its HvEcR-bound conformation so that overlap of common substructures in the two steroids were maximised. This led to a complete match of the 17-carbon steroid core and 8-carbon side-chain (C20–C27). The latter

chain includes the 25-OH group in the case of canescensterone. Consequently, the 25-OH group was placed in the hydrophobic region of the binding pockets, whereas the 24-pyrrole-2-carboxylate group was placed in the region in contact with Ba-T337 and/or Ba-T426, yielding plausible poses characterised by additional hydrogen bonds between the pyrrole ring of (24*S*)- and (24*R*)-canescensterone and either Ba-T337 and/or Ba-T426. Despite the opposite orientation of the 24-pyrrole-2-carboxylate compared with 25-OH of canescensterone, both the manual and computer-aided models can explain the divergent responsiveness of lepidopteran and non-lepidopteran EcRs in terms of variation in the polarity characteristics of specific regions of their binding cavities in contact with the distinguishing chemical features of this phytoecdysteroid. Although further investigation of canescensterone–EcR complexes by X-ray crystallography and/or NMR would further assist in the elucidation of the binding mode, the full configurational assignment of canescensterone greatly advances structure–activity relationship studies aimed at developing novel compounds for agrochemical research and actuators for ligand-inducible EcR-based gene regulatory systems.

Conclusions

The information relating to the configuration of canescensterone was successfully transferred from C-17 of the steroid nucleus to C-24 of the side-chain and the absolute configuration at C-24 was determined as *R*. This was achieved exclusively by NMR spectroscopy. No additional information from chemical synthesis had to be included. The *J*- and NOE-based analysis has been proven in this case to be a powerful tool for the stereochemical determination of the acyclic side-chain.[35] Because enough information could be obtained from established and robust 2D NMR spectra, the conformation of each C_2 fragment could be defined without computational simulation. The stereochemistry at C-24 of canescensterone allowed us to construct docking models in both a lepidopteran and a non-lepidopteran EcR based on the solved crystal structures of homologous ecdysteroid ligand-receptor complexes. The models allowed canescensterone selectivity for non-lepidopteran EcRs to be rationalised in terms of different steroid-tail contact residues in lepidopteran versus non-lepidopteran receptors.

Experimental Section

Phytoecdysteroid: Canescensterone (**2**) was isolated from *Vitex canescens* by using the method described previously[12] with some modification. Briefly, the pulverised dry bark (420 g) of *V. canescens*, obtained from the same location as in the previous study, was extracted successively with *n*-hexane and methanol. The concentrated MeOH extract was extracted with 1-butanol. The butanol extract was repeatedly chromatographed on silica column using $CHCl_3$/MeOH as eluent to yield compounds **1** (310 mg) and **2** (3 mg). The ^{1}H NMR spectroscopic data of **1** and **2** were consistent with those of the previous report.[12]

NMR Analysis: All NMR spectra of canescensterone were recorded at 25 °C in [D_6]DMSO. Chemical shifts were reproducible to ±0.01 ppm for ^{13}C NMR and ±0.001 ppm for ^{1}H NMR, respectively.

1D ^{1}H NMR spectra, 2D heteronuclear-correlated ^{13}C HSQC and ^{13}C HMBC spectra were recorded with a Bruker Avance 600 spectrometer (599.90 MHz for ^{1}H) using a 5-mm cryoprobe head. A standard Bruker pulse sequence was used to record the ^{13}C HSQC spectrum. The ^{13}C HMBC spectrum was recorded with a modified ^{13}C HMBC sequence[15] that allowed accurate extraction of long-range ^{13}C and ^{1}H couplings. However, owing to the very limited amounts of the compound at natural abundance, only a qualitative interpretation of the coupling constants could be obtained. The ^{1}H NMR spectrum was acquired by using 16K data points at a spectral width of 7 kHz with 16 scans and a 2.0 s delay between transients. The data matrix was zero-filled to 32K points and processed with an exponential window function. Both heteronuclear 2D spectra were measured over 4K complex points in F_2, 320 increments in F_1 for ^{13}C HSQC and 512 for ^{13}C HMBC, collecting 80 (HSQC) or 160 (HMBC) scans per increment with a relaxation delay of 2.0 s. The spectral widths in the F_2 and F_1 dimensions were 7 and 20 kHz (HSQC), and 7 and 30 kHz (HMBC), respectively.

2D homonuclear-correlated TOCSY, P.E.COSY[16] and NOESY spectra were recorded with a Bruker Avance 700 spectrometer (700.13 MHz for ^{1}H) using an inverse-detection triple resonance 5 mm probe head (TXI). The spectral widths of these three spectra were 7.2 kHz in both F_2 and F_1 domains. 4K × 1K data points were acquired with 24 scans per increment for the P.E.COSY experiment, 4K × 952 data points with 88 scans for the NOESY experiment and 4K × 880 data points with 24 scans for the TOCSY experiment. Relaxation delays of 2.0 s were used for all experiments and the mixing time in the NOESY experiment was set to 400 ms.

Molecular Modelling and Docking: The input protein structures for (24*R*)-canescensterone docking calculations were 1) the ligand-binding domain (LBD) from the homopteran *Bemisia argentifolii* (Ba) EcR[17] and 2) the LBD from a mutant (E274V/V390I/Y410E) of the lepidopteran *Choristoneura fumiferana* (Cf) EcR, dubbed VY-CfEcR.[7a,18] The BaEcR and VY-CfEcR structures were prepared from homologous receptors whose ecdysteroid-bound LBD crystal structures have been solved, that is, the hemipteran *Bemisia tabaci* (Bt) EcR[19] and the lepidopteran *Heliotis virescens* (Hv),[20] respectively, as follows. The crystal structure of 20E-bound HvEcR LBD (PDB 2R40) (as both the 20E- and PoA-bound HvEcR LBD crystal structures have been solved and canescensterone bears the 25-OH as does 20E, we selected the 20E-bound, rather than the PoA-bound, HvEcR LBD as input protein structure for docking calculations) was superimposed on the crystal structure of PoA-bound BtEcR LBD (PDB 1Z5X) in PyMOL[21] to give a root-mean-square (rms) deviation of 0.740 Å for C_α atoms. Structures were imported into Maestro v. 8.0.[22] HvEcR LBD was mutated at two positions (Hv-V395 to VY-Cf-I390 and Hv-Y415 to VY-Cf-E410) to obtain the VY-CfEcR LBD (residues 284–524). No mutations were made to the BtEcR LBD to obtain the BaEcR LBD (residues 213–448) as the amino-acid sequences of these two LBDs are identical. The respective ligands were removed and hydrogen atoms were added. The structure of **2** was built from the available EcR-bound 20E crystal structure followed by minimisation of the 24-pyrrole-2-carboxylate group (OPLS-2005 Force Field in MacroModel[22]). The software GOLD (Genetic Optimisation for Ligand Docking) v. 3.2[23] was used to calculate plausible binding poses of **2** in the VY-CfEcR and BaEcR LBDs. The input protein structure was kept rigid, except for the torsion angles of

the Ser, Thr and Tyr hydroxy groups, which are optimised by the algorithm. To avoid bias with the known ecdysteroid-binding cavity from the solved crystal structures, we decided to keep the input ligand-binding site on each of the two docked receptors largely undefined, that is, including all atoms and associated residues within 20 Å of the 17-carbon steroid core of the reference steroid crystal structure (HvEcR-bound 20E and BtEcR-bound PoA for docking VY-CfEcR LBD and BaEcR LBD, respectively). In this way, we could test the ability of the algorithm to identify the correct steroid-binding site of the EcR. The GoldScore Fitness Function, as implemented in the GOLD program, was used to evaluate the energy of interaction between **2** and each of the two EcRs and to score the different binding poses. This function is made up of the following components: protein–ligand hydrogen-bonding energy, protein–ligand van der Waals energy and ligand internal van der Waals and torsional strain energies. The best-ranked solutions for **2** docked in the VY-CfEcR and BaEcR LBDs had fitness scores of 50.86 and 56.10, respectively, and were taken into account for visual examination of the corresponding binding modes.

Supporting Information (see also the footnote on the first page of this article): ^{1}H and ^{13}C NMR chemical shift data and $^{3}J_{HH}$ coupling constants of **2**.

Acknowledgments

C. G. thanks the Max-Planck-Gesellschaft for financial support. This work was further supported by the Deutsche Forschungsgemeinschaft (DFG) (FOR 934 and GRK 782). C. G. also acknowledges support by the Fonds der Chemischen Industrie. We thank Prof. Jean-Pierre Girault and Dr. Robert E. Hormann for their constructive comments on the manuscript.

[1] R. Lafont, C. Dauphin-Villemant, J. T. Warren, H. Rees in *Comprehenisve Molecular Insect Science* (Eds.: L. I. Gilbert, K. Iatrou, S. S. Gill), Elsevier/Pergamon, Amsterdam, **2005**, vol. 3, p. 125–195.
[2] M. R. Koelle, W. S. Talbot, W. A. Segraves, M. T. Bender, P. Cherbas, D. S. Hogness, *Cell* **1991**, *67*, 59.
[3] *Ecdybase* (a free ecdysteroid database) http://ecdybase.org.
[4] L. Dinan, *Phytochemistry* **2001**, *57*, 325.
[5] a) R. Lafont, L. Dinan, *J. Insect Sci.* **2003**, *3*, 1; b) L. Dinan, R. Lafont, *J. Endocrinol.* **2006**, *191*, 1.
[6] a) V. S. Tavva, S. R. Palli, R. D. Dinkins, G. B. Collins, *Arch. Insect. Biochem. Physiol.* **2007**, *65*, 164; b) L. D. Graham, W. M. Johnson, A. Pawlak-Skrzecz, R. E. Eaton, M. Bliese, L. Howell, G. N. Hannan, R. J. Hill, *Insect. Biochem. Mol. Biol.* **2007**, *37*, 611; c) E. R. Hormann, S. Lapenna, N. L. Dinan, *Steroidal Ligands and their Use in Gene Switch Modulation*, 20090298175, WO/2009/114201, **2009**, p. 180.
[7] a) D. Karzenowski, D. W. Potter, M. Padidam, *BioTechniques* **2005**, *39*, 191; b) D. No, T. P. Yao, R. M. Evans, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 3346.
[8] a) S. R. Palli, R. E. Hormann, U. Schlattner, M. Lezzi, *Vitamins Hormones* **2005**, *73*, 59; b) W. Weber, M. Fussenegger in *Handbook of Experimental Pharmacology* (Eds.: R. Feil, D. Metzger), Springer, Berlin, **2007**, vol. 178, pp. 73–105.
[9] L. Dinan, R. E. Hormann in *Comprehensive Molecular Insect Science* (Eds.: L. I. Gilbert, K. Iatrou, S. S. Gill), Elsevier Pergamon, Oxford, **2005**, vol. 3, p. 197.
[10] E. Saez, M. C. Nelson, B. Eshelman, E. Banayo, A. Koder, G. J. Cho, R. M. Evans, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 14512.
[11] S. Lapenna, J. Friz, A. Barlow, R. Palli, L. Dinan, R. E. Hormann, *FEBS J.* **2008**, *275*, 5785.
[12] A. Suksamrarn, C. Sommechai, P. Charulpong, B. Chitkul, *Phytochemistry* **1995**, *38*, 473.
[13] A. Suksamrarn, P. Pattanaprateep, T. Tanachtchairatana, W. Haritakun, B. Yingyongnarongkul, N. Chimnoi, *Insect. Biochem. Mol. Biol.* **2002**, *32*, 193.
[14] a) A. Schuetz, J. Junker, A. Leonov, O. F. Lange, T. F. Molinski, C. Griesinger, *J. Am. Chem. Soc.* **2007**, *129*, 15114; b) A. Schuetz, T. Murakami, T. Takada, T. Junker, M. Hashitmoto, C. Griesinger, *Angew. Chem. Int. Ed.* **2008**, *47*, 2032.
[15] L. Verdier, P. Sakhaii, M. Zweckstetter, C. Griesinger, *J. Magn. Reson.* **2003**, *163*, 353.
[16] L. Mueller, *J. Magn. Reson.* **1987**, *72*, 191.
[17] J. Zhang, D. E. Cress, T. S. Dhadialla, S. R. Palli, *Whitefly ecdysone receptor nucleic acids, polypeptides, and uses thereof*, PCT/US02/05234, WO 03/027266A2, **2003**, p. 64.
[18] S. R. Palli, M. B. Kumar, *Mutant receptors and their use in a nuclear receptor-based inducible gene expression system*, PCT/US2005/015089, WO 2005/108617A2, **2005**, p. 103.
[19] J. A. Carmichael, M. C. Lawrence, L. D. Graham, P. A. Pilling, V. C. Epa, L. Noyce, G. Lovrecz, D. A. Winkler, A. Pawlak-Skrzecz, R. E. Eaton, G. N. Hannan, R. J. Hill, *J. Biol. Chem.* **2005**, *280*, 22258.
[20] a) C. Browning, E. Martin, C. Loch, J.-M. Wurtz, D. Moras, R. H. Stote, A. P. Dejaegere, I. M. L. Billas, *J. Biol. Chem.* **2007**, *282*, 32924; b) I. M. L. Billas, T. Iwema, J. M. Garnier, A. Mitschler, N. Rochel, D. Moras, *Nature* **2003**, *426*, 91.
[21] PyMOL, DeLano Scientific, LCC, San Carlos, CA, **2006**.
[22] Maestro (v. 8.0, MacroModel v. 8.5), Schrödinger, LCC, New York.
[23] a) G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43; b) G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727; c) M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins: Struct., Funct., Genet.* **2003**, *52*, 609.
[24] A. Suksamrarn, C. Sommechai, *Phytochemistry* **1993**, *32*, 303.
[25] A. Suksamrarn, B.-E. Yingyongnarongkul, S. Charoensuk, *Tetrahedron* **1999**, *55*, 255.
[26] L. Fábián, G. Argay, A. Kálmán, M. Báthori, *Acta Crystallogr., Sect. B* **2002**, *58*, 710.
[27] J. G. Sena Filho, J. Duringer, G. L. A. Maia, J. F. Tavares, H. S. Xavier, M. S. da Silva, E. V. L. da-Cunha, J. M. Barbosa-Filho, *Chem. Biodiversity* **2008**, *65*, 707.
[28] A. Suksamrarn, S. Kumpun, B.-E. Yingyongnarongkul, *J. Nat. Prod.* **2002**, *65*, 1690.
[29] A. Suksamrarn, B.-E. Yingyongnarongkul, N. Promrangson, *Tetrahedron* **1998**, *54*, 14565.
[30] M. Ukiya, T. Akihisa, K. Yasukawa, H. Tokuda, M. Toriumi, K. Koike, Y. Kimura, T. Nikaido, W. Aoi, H. Nishino, M. Takido, *J. Nat. Prod.* **2002**, *65*, 179.
[31] R. Kasai, R.-L. Nie, K. Nashi, K. Ohtani, J. Zhou, G.-D. Tag, O. Tanaka, *Agric. Biol. Chem.* **1989**, *53*, 3347.
[32] R. Kothapalli, S. R. Palli, T. R. Ladd, S. S. Sohi, D. Cress, T. S. Dhadialla, G. Tzertzinis, A. Retnakaran, *Developmental Genetics* **1995**, *17*, 319.
[33] T. Iwema, I. M. L. Billas, Y. Beck, F. Bonneton, H. Nierengarten, A. Chaumot, G. Richards, V. Laudet, D. Moras, *EMBO J.* **2007**, *26*, 3770.
[34] I. M. L. Billas, D. Moras, *Vitamins Hormones* **2005**, *73*, 101.
[35] a) N. Matsumori, D. Kaneno, M. Murata, H. Nakamura, K. Tachibana, *J. Org. Chem.* **1999**, *64*, 866; b) G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, *Chem. Rev.* **2007**, *107*, 3744; c) C. Bassarello, G. Bifulco, A. Zampella, M. V. D'Auria, R. Riccio, L. Gomez-Paloma, *Eur. J. Org. Chem.* **2001**, 39.

Received: March 17, 2010
Published Online: September 14, 2010

