Structural studies on meiosis-activating sterols and structurally related compounds

Potential ligands of the FF-MAS receptor

Structuur-chemisch onderzoek aan meiose activerende sterolen en verbindingen met gerelateerde structuren

Potentiële liganden van de FF-MAS receptor

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de Rector Magnificus, Prof.dr. W.H. Gispen, ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op donderdag 4 oktober 2001 des middags te 12:45 uur

door

Dirk Roeland Boer

geboren op 3 april 1972, te Ambon, Indonesië

1e promotor:	Prof.dr. J. Kroon († 3 mei 2001)
2e promotor:	Prof.dr. A.L. Spek verbonden aan het Bijvoet Centrum voor Biomoleculair Onderzoek, sectie Kristal- en Structuurchemie, Faculteit Scheikunde van de Universiteit Utrecht.
co-promotoren:	Dr. H. Kooijman verbonden aan het Bijvoet Centrum voor Biomoleculair Onderzoek, sectie Kristal- en Structuurchemie, Faculteit Scheikunde van de Universiteit Utrecht.
	Dr. J. Kelder verbonden aan de N.V. Organon te Oss.

CIP-gegevens Koninklijke Bibliotheek, Den Haag
Boer, Dirk Roeland
"Structural studies on meiosis-activating sterols and structurally related compounds – potential ligands of the FF-MAS receptor"
Utrecht: Universiteit Utrecht, Faculteit Scheikunde
Thesis Universiteit Utrecht – With ref. – With summary in Dutch
ISBN 90-393-2833-1

Ter nagedachtenis aan
Jan Kroon

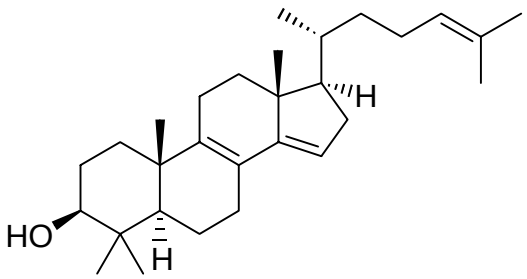
Table of Contents

Chapter 1

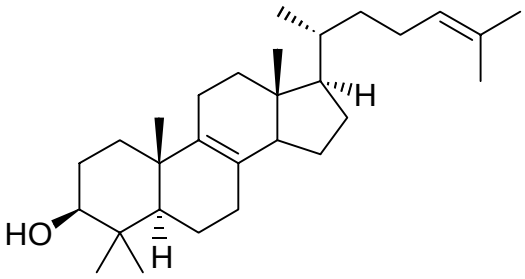
Introduction

Meiosis Activating Sterols

In 1995, Byskov and co-workers discovered two compounds that were found to be important for the development of mammalian germ cells. These steroid-like molecules are intermediates in the biosynthetic pathway of cholesterol, which are called sterols (Figure 1). They were found to exhibit a similar function: both induce immature reproductive cells to complete their development (Byskov *et al.*, 1995; Byskov *et al.*, 1998; Byskov *et al.*, 1999; Yding Andersen *et al.*, 1999). The molecules represent a class of compounds with novel pharmaceutical properties, designated Meiosis Activating Sterols (MAS) or, alternatively, Meiosis Inducing Substances (MIS). The names of the compounds are derived from their function and the location in the organisms they were isolated from. Compound **1**, 4,4-dimethyl-5α-cholest-8,14,24-triene-3β-ol, initiates the maturation of human oocytes in ovarian follicles and is called follicular fluid-MAS (FF-MAS). Compound **2**, 4,4-dimethyl-5α-cholest-8,24-diene-3β-ol, is involved in the production of spermatozoa in bull testis and is therefore called testicular-MAS (T-MAS).



4,4-dimethyl-5α-cholest-8,14,24-triene-3β-ol
(FF-MAS)



4,4-dimethyl-5α-cholest-8,24-diene-3β-ol
(T-MAS)

Figure 1. Structures of the meiosis activating compounds **1** (FF-MAS) and **2** (T-MAS).

The importance of the discovery of the two compounds lies in the possible applications of MAS-derived structures. Interference in the biological processes on which the MAS compounds operate can either promote or prevent fertility (Byskov *et al.*, 1999). Thus, MAS antagonists that block MAS action in either sex are attractive alternatives for the contraceptives used to date, since they are not likely to affect the normal hormonal balance. Furthermore, MAS agonists may prove valuable in treating problems related to reproductive cell maturation. Thus, sperm production in men could be increased and aberrant oocyte production in women could be remedied. The work described in this thesis is focused on the exploration of structural and physicochemical properties of MAS, in particular of FF-MAS and derived compounds. Implications of these

properties on the MAS activity are discussed and where possible, structure-activity relationships will be derived.

Meiosis

Meiosis is a reductive division process to produce reproductive cells with half the genetic material of the normal cell. During the meiotic process, the genetic content of a diploid precursor cell, with a single set of chromosomes and duplicate DNA, is reduced. The meiotic cell cycle, depicted in Figure 2, consists of four phases in which the cell grows, replicates DNA, segregates the chromosomes and divides. During the G1 phase the cell accumulates the enzymes needed to replicate DNA and other material that provides the necessary resources to complete the cycle. The S1 phase that follows is used to replicate the genetic material. The resulting cell, with two sets of chromosomes and therefore four DNA copies, then undergoes two successive reduction divisions called MI and MII. In mitosis, only one reduction division takes place, after which the resulting cells can enter a new division cycle starting with a second S phase. In meiosis, however, the development of the germ cells is continued and no S phase occurs between the reduction phases. In males, four haploid cells are produced, which all become functional sperm. In females, only one of the four haploids develops into a fully functional ovum, the rest of the genetic material is extruded as polar bodies (PB's). The development of the gametes can be arrested at several stages of the meiotic cycle, *i.e.* at the late G2/early MI phase, which is called prophase I and after the oocyte has developed into a fully functional egg cell.

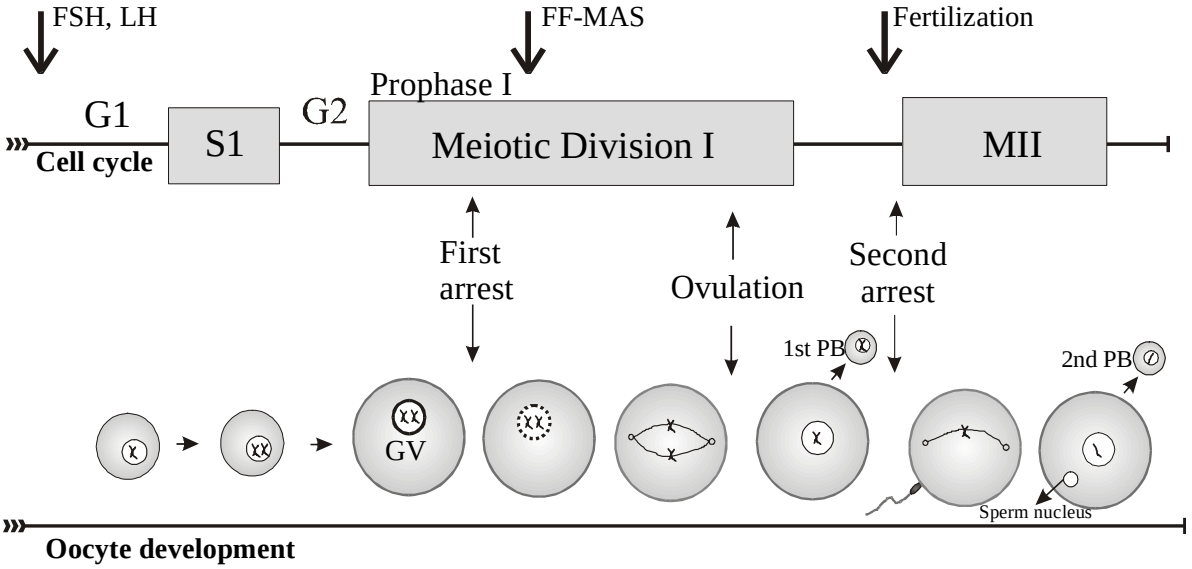


Figure 2. Schematic representation of the meiotic cell cycle.

In humans, the meiotic principle is the same for both sexes, but the timing is very different. Male germ cells are arrested in the G1 phase, and meiosis is not initiated until puberty. Thereafter, meiosis is continuous and uninterrupted. Female germ cells develop into primary oocytes during the first meiotic phase in fetal life. Their development is arrested a first time in the late prophase I stage, before or just after birth. This stage is characterized by a prominent nuclear membrane, called the germinal vesicle (GV), which can be seen through a microscope. The primary oocytes rest until puberty after which each in turn goes through the MI and MII phases. During this process, called GVBD, the germinal vesicle breaks down and the first polar body is extruded, which contains half of the DNA of the diploid mother cell. Thereafter, meiosis is arrested a second time and is resumed exclusively by entering of a sperm cell, after which the second PB is extruded.

Regulation of meiosis

The development of primary oocytes into egg cells takes place in follicles, which are the functional units of the ovary and ensure a unique environment. Primary oocytes arrested at the prophase stage are surrounded in succession by an antrum (*i.e.* fluid derived from serum), a layer of cumulus cells and a layer of mural granulosa cells. All of these cells are in communication through gap junctions, which facilitate transportation of hormones and lower molecular-weight compounds. Meiosis is initiated through a preovulatory surge of gonadotropins (Yding Andersen *et al.*, 1999), in particular follicle stimulating hormone (FSH) and luteinizing hormone (LH). The gonadotropins are also responsible for ovulation. Receptors for both hormones seem to be present on mural granulosa cells and cumulus cells exclusively, which makes them the primary target of these gonadotropins (Byskov *et al.*, 1997; Lu *et al.*, 2000). After the gonadotropin surge, the cyclic AMP level in the cumulus-oocyte complex increases (Eppig & Downs, 1988), whereas the cAMP level in the oocyte itself seems to decrease (Schultz, 1988). Meiosis does not occur when phosphodiesterase (PDE) enzymes, which are responsible for cAMP degradation, are inhibited in oocytes (Wiersma *et al.*, 1998). These findings indicate that a drop in intracellular cAMP concentration within the oocyte may be important for induction of meiosis. In fact, follicular fluid contains a PDE inhibitor, hypoxanthine, which is important in maintaining meiotic arrest (Törnell *et al.*, 1991; Hegele-Hartung *et al.*, 1999). In addition, it was found that phosphodiesterase 3 inhibitors blocked oocyte maturation in rodents, without affecting the normal cycle (Wiersma *et al.*, 1998).

Previously, it was believed that the resumption of meiosis was accomplished by withdrawing inhibitory substances from the oocytes, thus releasing it to continue meiosis spontaneously (Moor & Warnes, 1979; Dekel, 1995). However, after the gonadotropin surge, the hypoxanthine levels in mouse

follicular fluid levels drop only slightly, indicating that a stimulatory component may override the inhibitory effect of hypoxanthine. The discovery that MAS is able to induce maturation of the oocyte despite the presence of hypoxanthine (Törnell *et al.*, 1991; Hegele-Hartung *et al.*, 1999), indicates that such a mechanism is in fact present. The notion that MAS compounds are the endogenous activating stimulants of resumption of meiosis is confirmed by studies by Leonardsen *et al.* (2000a) and Lu *et al.* (2000). They showed that inhibition of different proteins involved in the synthesis of MAS prevents meiosis. In addition, it was found that 14α-demethylase, which demethylates lanosterol to give FF-MAS (see below), is important for meiosis (Yoshida *et al.*, 1996; Strömstedt *et al.*, 1998; Majdic *et al.*, 2000).

FF-MAS induces the resumption of meiosis in oocytes separated from their cumulus and granulosa cells, in contrast to FSH (Yding Andersen *et al.*, 1999). Thus, FF-MAS exerts in action on the oocyte directly. It was found that FSH is able to induce maturation only if the cumulus-enclosed oocyte complex is still intact just after the preovulatory surge of gonadotropins (Byskov *et al.*, 1997; Yding Andersen *et al.*, 1999). This suggests that FF-MAS is synthesized in the cumulus cells in response to elevated levels of FSH and transported to the oocyte, either through gap junctions, or in a paracrine way, *i.e.* by diffusion through the extracellular medium (the antrum). Communication between both cumulus cells and the oocyte appears to be necessary for the function of MAS. A receptor protein is believed to mediate the signal induced by MAS, but little is known about the receptor for MAS compounds. One possible candidate is the orphan nuclear receptor LXRα, which is activated five-fold by binding of FF-MAS (Janowski *et al.*, 1996; Ruan *et al.*, 1998). Oxycholesterol derivatives can also bind to the LXRα, but they are, in contrast to FF-MAS, not able to induce resumption of meiosis *in vitro* (Grøndahl *et al.*, 1998). Therefore, it does not seem likely that LXRα is the endogenous MAS receptor. A study on the influence of cholera toxin, a G-protein coupled receptor inhibitor, on the resumption of meiosis suggests that a G-protein-coupled receptor mechanism is responsible for the MAS mode of action (Grøndahl *et al.*, 2000). Recent studies suggest that the MAS receptor might be a plasma membrane-associated molecule (Færge *et al.*, 2001) involved in the cAMP-protein kinase A-dependent signal transduction pathway (Leonardsen *et al.*, 2000b).

Sterols: nomenclature and biosynthesis

Sterols, including the MAS compounds, are intermediates in the post-squalene biosynthesis of cholesterol. The discovery of the involvement of MAS compounds in the resumption of meiosis (Byskov *et al.*, 1995) was the first time that a specific function of a sterol compound, other than cholesterol, was found (Jackson *et al.*, 1997). Before discussion of the pathway involved in producing cholesterol, a brief description of the nomenclature and notations of sterols is

given. In Figure 3, the labeling scheme of the atoms of the androstane skeleton is given, as well as the names of the four rings of the sterol and steroid skeletons. The side above the plane of the paper of Figure 3 is called the β-side, the opposite side is called the α-side. Substituents occurring above the plane of the paper are denoted by a β, those occurring below the plane of the paper by an α following the atom number at which they occur. Double bonds are denoted by a Δ, with the atom numbers at which they occur as superscripts. If the double bond occurs between two consecutive numbered atoms, the higher atom number is not given explicitly. However, if this does not apply, the higher number is given between parentheses, directly after the lower atom number. If more than one double bond is present, the numbers that indicate their positions are separated by a comma. Thus, FF-MAS is a $\Delta^{8,14,24}$ and T-MAS a $\Delta^{8,24}$ structure.

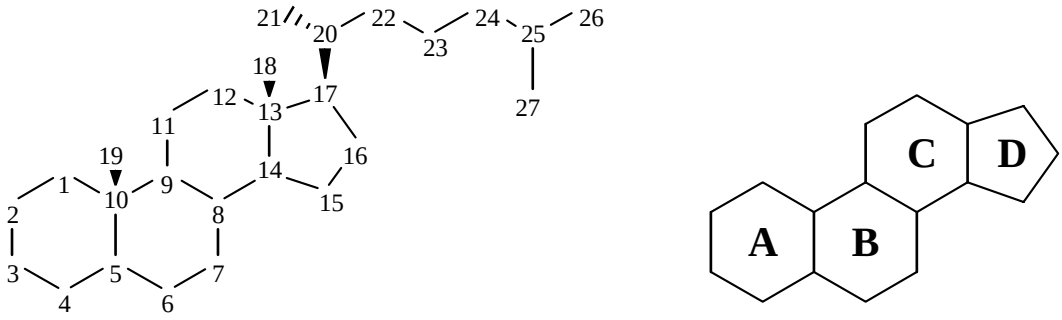


Figure 3. Numbering of the skeletal atoms in cholesterol and the MAS compounds (left) and the designation of the four skeleton rings (right). The β-side is above and the α-side is below the plane of the paper.

The post-squalene biosynthetic pathway of cholesterol is schematically represented in Figure 4 (Schroepfer Jr., 1982). Cyclization of squalene yields lanosterol, which is subsequently demethylated at C14 by cytochrome P450-14α-demethylase to give FF-MAS. Reduction of the Δ^{14} double bond by 14α-reductase yields T-MAS, which is then converted into zymosterol by elimination of two methyl groups at C4 through a number of oxidation and reduction steps. The Δ^{8} bond in zymosterol is converted to a Δ^{7} bond by a Δ^{8}-isomerase. Dehydrogenation of the single bond between C5 and C6 and reduction of the Δ^{7} double bond gives desmosterol. Reduction of the Δ^{24} double bond gives cholesterol. Alternatively, the Δ^{24} bond may be reduced before dehydrogenation of the Δ^{5} bond to give lathosterol. Dehydrogenation of the C5 to C6 single bond and subsequent reduction of the double bond at C7 gives cholesterol. The synthesis of FF-MAS and T-MAS follows the reverse direction of the biosynthetic pathway of cholesterol, starting from desmosterol ($\Delta^{5,7}$). During the reactions, several intermediate isomers can in theory be formed. It was observed that only a limited number could actually be obtained in appreciable yields. Apparently, the isomers have different stabilities.

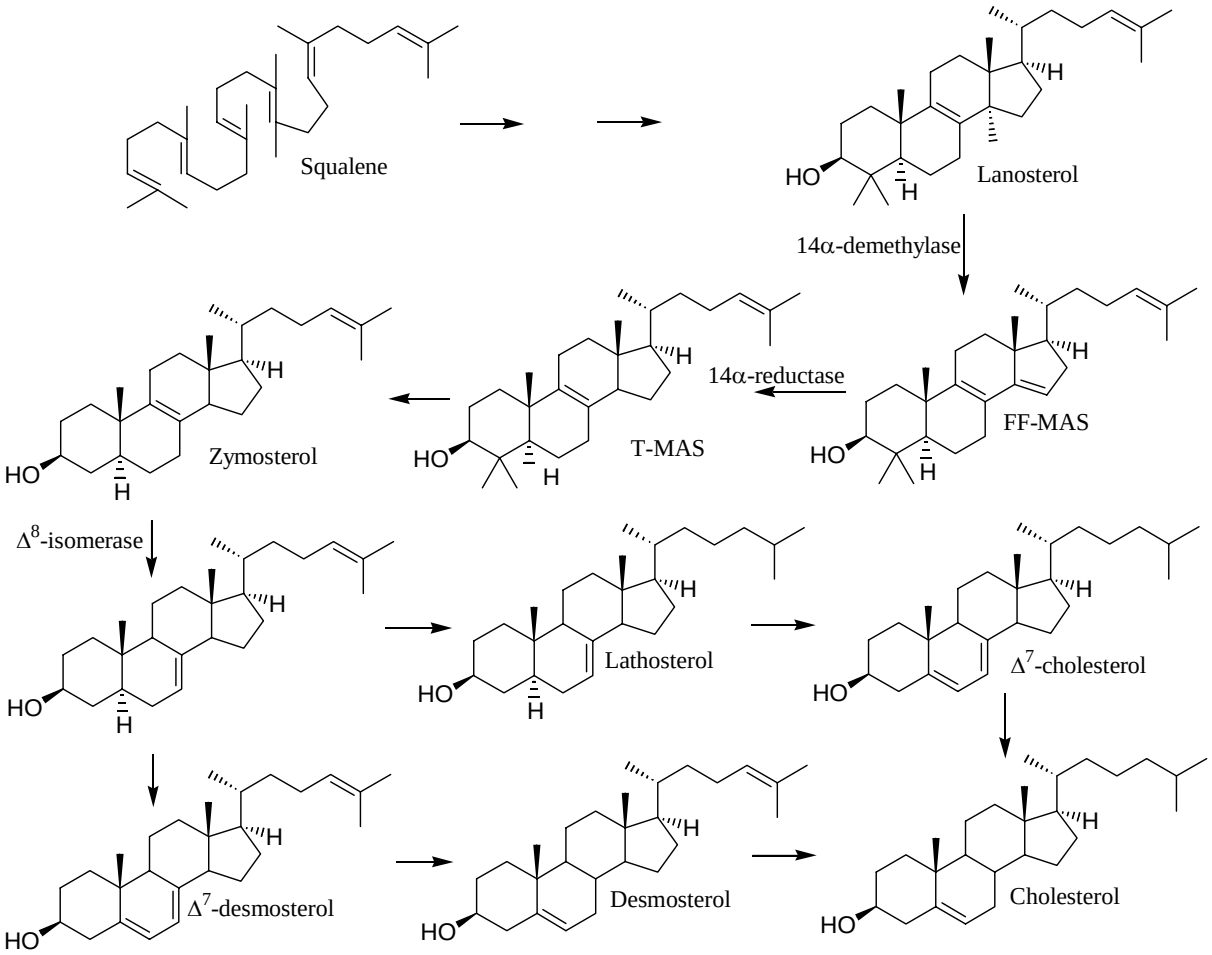


Figure 4. Schematic representation of the biosynthetic pathway of the cholesterol.

Activity of the MAS compounds

As stated above, primary oocytes contain a clearly visible germinal vesicle that breaks down when meiosis is resumed (GVBD) after the first arrest. This phenomenon is used to determine whether a compound can override the inhibitory action of hypoxanthine in *in vitro* tests. All FF-MAS related known agonists and antagonists are structurally very similar to FF-MAS (N.V. Organon, personal communication). They contain an androstane skeleton and a 3β-OH group. The 3β-OH group is mandatory, indicating that the ring A moiety plays a major role in binding MAS ligands to their receptor. All compounds contain a lipophilic aliphatic chain attached to C17. The double bond between atoms C24 and C25 enhances activity compared to compounds with a single bond at this position, but it is not mandatory (Byskov *et al.*, 1995; Strömstedt *et al.*, 1998; Wenckens *et al.*, 1998). Similarly, two methyl groups at C4 increase

the activity of the MAS ligands, but do not have to be present. Replacement of C22 by an oxygen atom leads to loss of activity for compounds with a similar configuration of C20 of FF-MAS (natural configuration). In contrast, when C22 is replaced by an oxygen atom in compounds with inverted stereochemistry of C20 (unnatural configuration), compounds can be active.

The double bond in the side chain of the MAS compounds can be mimicked by a phenyl group, as is observed in a number of structures. The same holds for the $\Delta^{8,14}$ double bond system, which can be replaced by an aromatic C ring (Grønvald *et al.*, 1997). The presence of a conjugated pi-system at ring C therefore seems mandatory. Other double bond isomers, such as Δ^{5}, $\Delta^{5,7}$, $\Delta^{7,9(11)}$ and $\Delta^{7,14}$ structures, did not result in active compounds. However, a $\Delta^{6,8(14)}$ compound was partially active.

MAS and Vitamin D3

The MAS agonists share an interesting feature in the structure-activity relationships of the stereochemistry at C20 (see above) with that of 1α,25-dihydroxy-vitamin-D_3 (calcitriol, see Figure 5) and its analogs. It was found that calcitriol analogs with an unnatural stereochemistry at C20 induced a higher transcriptional activity than those with natural configuration (Midland *et al.*, 1993; Boullion *et al.*, 1995; Yamamoto *et al.*, 1996; Liu *et al.*, 1997; Yamada *et al.*, 1998; Väisänen *et al.*, 1999; Yamamoto *et al.*, 1999). Similarly, 22-oxa compounds derived from FF-MAS with inverted C20 stereochemistry were able to induce meiosis *in vitro*.

Calcitriol is the endogenous ligand of the vitamin D receptor (VDR). The VDR is a member of the superfamily of nuclear receptors, for some of them a crystal structure is available (*e.g.* Renaud *et al.*, 1995; Wagner *et al.*, 1995; Brzozowski *et al.*, 1997; Nolte *et al.*, 1998; Tanenbaum *et al.*, 1998). It is not unreasonable to assume that the topology of the receptor active sites of the VDR and the MAS receptors are similar considering the similarity in side chain structure activity relationship (SAR). As a consequence, the C20 epimeric compounds of MAS may exhibit a binding mode similar to calcitriol and its analogs upon binding to their respective receptors. Analysis of the binding of calcitriol to its receptor may therefore increase insight into the structure-activity relationships of the C20 stereochemistry of the MAS compounds.

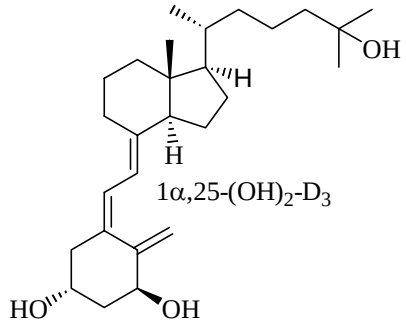


Figure 5. Chemical structure of 1α,25-dihydroxy-vitamin D_3 (calcitriol).

Methodologies

Computational modelling

Computational methods are in general inexpensive and valuable techniques for exploration of conformational behavior of molecules. Basically, two different methods can be distinguished, which are based on quantum mechanics and classical mechanics respectively. The first approach, called *ab initio* methods, use the quantum-mechanical Schrödinger equation and can in principle give an exact solution to a specific problem. Semi-empirical methods are derived from the *ab initio* method, but use parameterization to save computer time at the expense of accuracy. The second method is called molecular mechanics. It describes a molecule using classical mechanistic equations. The semi-empirical and molecular mechanics methods are discussed in more detail below.

Quantum chemical *ab initio* methods are time consuming and become increasingly expensive with increasing size of the modeled system. Most of the computer time is needed to calculate energy terms associated with electron-electron repulsion terms that occur in the Schrödinger equations for systems with more than one electron. Semi-empirical methods aim to reduce the calculation time and memory space by neglecting or approximating some of the two-electron overlap integrals that need to be calculated in the *ab initio* method. Specifically, only integrals associated with the valence electrons on the atoms in the system under consideration are evaluated. The rationale for this approach lies in the fact that most molecular properties of interest are related to the behavior of these electrons. The core electrons are considered to be part of the nuclear core of an atom. The energy contributions related to interactions between electrons in the core and the energies related to interactions between core and valence electrons are parameterized using experimental data, hence the name of the method. Several methods have been developed which differ in the degree of neglect of the two-electron integrals. The most widely used methods are MNDO, AM1 and PM3. One of the applications of the semi-empirical methods is the calculation of partial atomic charges, based on the distributions of the valence electrons on the atoms.

Molecular mechanics does not explicitly include electrons. Rather, it treats atoms as soft spheres connected to each other by a classical spring. The optimum bond distance and the magnitude of the force constant of the spring are parameters derived from experimental data. Similarly, bond and torsion angles are described by energy terms that use experiment-derived parameters, as are other non-bonded intra- and intermolecular interactions. The specific terms and parameterization used are called the force field and they determine the accuracy of the method. The energies calculated by molecular mechanics are only indirectly related to physical values, *i.e.* energy differences of different conformations indicate their relative presence in the system used in the modeling process. The fraction of a conformation present in the modeled

environment is related to the energy through Boltzmann statistics. Often, one is interested in the ensemble of conformations that a molecule can adopt. Sampling of the possible conformations can be achieved via several techniques. Generally, such an ensemble is generated by simulation of molecular movement with time, using a technique called molecular dynamics. The atomic motions (speed, acceleration) and the driving forces of the system are described by classical Newtonian equations. Another technique is stochastic sampling, in which the atom positions are modified randomly.

X-ray diffraction

X-ray crystallography is a technique to obtain experimental information on structural and conformational aspects of molecules. A successful X-ray analysis gives a structural model of the molecule(s) in the crystal in the form of the atomic positions. The method requires crystals of sufficient quality and size. A crystal can be regarded as a repetition of a building block, called the unit cell, in three dimensions. The quality of the crystal and the data obtained from it depend on the extent to which translation symmetry of this unit cell throughout the crystal is fulfilled. Obtaining good quality crystals is therefore an important part of a crystal structure analysis and has received much attention. The best crystals are usually obtained by slow growth from a slightly over-saturated solution. When the crystal is irradiated with electromagnetic waves, interaction between this incident beam and the electron cloud in the crystal cause it to scatter. As a result of the periodicity in the crystal, interference results in extinction of the scattered beam in all but a few discrete directions. These directions are indicated by indices *h*, *k* and *l* and depend on the geometry of the unit cell, which is defined by three translation vectors **a**, **b** and **c**. The intensities of the diffracted X-ray beams are related to the electron density distribution in the unit cell.

Fourier transformation can in principle be used to reconstruct the electron density of the unit cell from the diffraction intensities and the phase differences of the diffracted beams with the incident beam, called the structure factors. However, information on the phases is lost during the diffraction experiment. This is known as the ‘phase problem’. Two solutions to this problem have been developed in chemical crystallography. One approach, referred to as ‘direct methods’, uses strong reflections and certain combinations thereof to make an educated initial guess of the phases. In the second approach, called the Patterson method, a Fourier transform is calculated using only the observed intensities. The resulting Patterson map represents the self-convolution function of the electron density in which the position vectors of points with maximum intensity correspond to interatomic vectors in the crystal structure. The intensity of a Patterson peak is related to the atomic number of the contributing atoms. For small molecules, reconstruction of the electron density distribution is nowadays straightforward in most cases. The electron density obtained is an average of time and over the periodicity of the crystal. Since atoms move and the

periodicity of a crystal is not perfect, the electron density is somewhat smeared out over a limited area. This effect is taken into account by atomic displacement factors. The initial model is a rough estimation of the electron density and needs to be refined to increase the model sophistication. In the refinement stage, an anisotropic description of the atomic movement is introduced and disorder, if present, is handled. It is the most time consuming step of a crystal structure determination since it requires the personal attention of the crystallographer.

A large number of crystal structures have been solved so far and have been collected in rapidly extending structural databases. These include the Cambridge Structural Database (CSD, ~233.000 structures in April 2001) for small organic and organometallic molecules, the Brookhaven Protein Databank (PDB, ~15.000 structures in May 2001) for proteins, nucleic acids, DNA and other biologically interesting macromolecules. These databases contain valuable information on the conformation and intermolecular interactions between molecules. Therefore, these databases are valuable tools in the design and modeling of new materials and molecular structures.

Nuclear Magnetic Resonance spectroscopy

Where crystallographic structure determinations yield molecular models in the solid phase, high-resolution solution Nuclear Magnetic Resonance (NMR) experiments provide structural and conformational details of molecules in solution. The spin states of atomic nuclei are degenerate when an external magnetic field is absent. In the presence of an external magnetic field, the degeneracy between spin states is removed for nuclei with a quantum spin number of a ½ or higher. The energy differences associated with these spin states lie in the range of radio waves (RF). When an RF pulse is applied to the sample in the external magnetic field, resonance of the nuclear spins will occur at discrete frequencies. These resonance frequencies depend on the chemical environment around the nucleus, *i.e.* the electron density distribution, because the external magnetic field is shielded by the presence of the electronic cloud. During acquisition of the data, all these frequencies are measured at once, and the resulting collected data is referred as a FID (Free Induction Decay). After Fourier transform of this FID, the NMR spectrum is obtained, in which the intensity of the signal is given as function of the resonance frequency of the nucleus. The latter is often given in parts per million (ppm), also referred to as chemical shift, which is independent of the NMR spectrometer used. Magnetization on spins can be transferred through bonds, which gives rise to a phenomenon called *J*-coupling. The result of coupling in the NMR spectrum is splitting of the signal of a spin system.

NMR spectroscopy is widely used as an analytical tool for identification of small molecular compounds. For this purpose, information on the chemical shifts and coupling constants usually suffice, which can be obtained through one-dimensional NMR experiments as described above. Derivation of more

detailed structural information on molecules requires the use of more sophisticated NMR techniques. In general, multi-dimensional NMR experiments are applied for these purposes. The discussion here will focus on 2D experiments, with two frequency axes. A two-dimensional NMR experiment involves four stages. During the first stage, the preparation stage, magnetization is transferred to a certain spin. During the following evolution stage, chemical shift evolution occurs over the period of the first time variable (t_1). By incrementing t_1 in a series of measurements, modulation of the chemical shifts occurs, which enables the detection of the various nuclei in the second dimension of the 2D spectrum. The third stage is used to manipulate the spin system in some way, the details depend on the type of experiment. In the experiments described below, magnetization is transferred from one spin to another, which is called mixing. During the final detection stage, chemical shift evolution occurs over the duration of the second time variable (t_2) and magnetization on the labeled spins is detected. A two-dimensional matrix of resonance data is acquired as a result, which gives the 2D spectrum after Fourier transformation in both dimensions. Magnetization transfer during the evolution time can be achieved in various ways, leading to different types of experiments. Two of those will be discussed here, called correlated spectroscopy (COSY) and nuclear overhauser enhancement spectroscopy (NOESY).

The key concept of the COSY methods is transfer of magnetization from one spin to another via ^3J coupling, which occurs through bonds. The transfer results in cross peaks that appear above and below the diagonal line of the 2D spectrum at the intersections of the frequencies of the two spin systems involved. From the cross peaks, atom connectivity's can be deduced and used to identify the molecular structure. In NOESY techniques and its variants, the principle of magnetization transfer is comparable, but the mechanism differs: nuclei can exchange magnetization via cross-relaxation, which occurs through space. The cross peaks are associated with nuclei at a short distance r, which is usually smaller than 5 Å, and their intensities are proportional to r^{-6}. From the intensities of the cross peaks, the inter-nuclear distances can be approximated. These distances are averages over time and over the ensemble of conformations present in the solution. From these distances the conformation or conformations of a molecule can be derived.

Homology modeling

For most proteins, experimental information on the structure is not available. Some of these proteins are involved in mechanisms that are associated with diseases. These proteins constitute interesting targets for the development of new drugs and structural information is often of great assistance in the development of new pharmaceutical agents. To remedy the fact that the structure of such a protein is not known, a structural model can be obtained if structural information is available on proteins that are evolutionary related. Related

proteins often have a high percentage of identical or similar residues in their sequences. Since the fold of a protein is largely determined by the amino acid sequence, one can assume that proteins with a high sequence homology have similar folds. A sequence of the protein with known structure can then be used as a template onto which the residues of the protein with unknown 3D structure are matched. This process, called 'alignment', becomes increasingly prone to errors with decreasing homology. As a consequence, the model decreases in quality when the homology decreases. After optimal alignment, at least forty percent of the residues should be identical to those of the template structure to ensure a reliable model.

The homology model is built by first constructing the backbone of the model protein using the coordinates of the template protein, excluding the loops between α-helices and β-sheets. Thereafter, the side chains of identical residues in the sequence alignment are included with similar conformation as in the template. The side chains that do not match those of the template are constructed using statistical information on side chain conformations present in the PDB database (see above). The loops present the largest challenge in the construction of a homology model because deletions, insertions and mutations occur most often there. Several methods have been developed to model these loops, which will not be discussed here. Often, several protein models are built and one is selected after the model is complete. Finally, the selected protein model is minimized using a mild procedure, to relieve strain caused by atomic clashes. A homology model can yield valuable information on potential binding modes of molecules and on structural and conformational requirements of ligands.

SuperStar

The design of new drugs is helped greatly if the tertiary structure of the target protein has been solved. The design of new, potent ligands using the protein structure requires a good understanding of molecular interactions and an adequate model to describe them. Several approaches have been developed with this goal. The most widely used methods are based on molecular mechanics. These methods calculate energies of interaction for a given system of molecules and evaluate the binding potency of a ligand based on these energies. Other methods have been developed which rely more directly on the experimental data on molecular interactions present in structural databases (Danziger & Dean, 1989a,b; Böhm, 1992a,b; Clark *et al.*, 1995; Laskowski *et al.*, 1996). One such a method is SuperStar (Verdonk *et al.*, 1999, 2001).

SuperStar locates favorable interaction sites of a chemical functional group near a molecule. The method uses intermolecular interaction information stored in a database called IsoStar (Bruno *et al.*, 1997, 1999). The IsoStar database consists of geometrical distributions of one chemical group, called the contact group, around another, called the central group. The distributions are derived from the crystal structures of the CSD and protein-ligand complexes from the

PDB. For a SuperStar run, a set of distributions with identical contact groups is used. The central groups of the distributions are superposed on matching molecular fragments in the molecule, thereby transforming the distribution of the contact group accordingly. After matching of all structural fragments, the studied molecule is surrounded by the separate contact groups distributions, some of which overlap. In the final step, SuperStar combines the separate distributions to give a composite propensity distribution for a contact group. From these composite distributions, favorable interaction sites for a contact group in the vicinity of a molecule (*e.g.* in a protein active site) can be identified. These hot-spots can then be used to find the binding mode of a known ligand and to design new ligands.

Outline of this thesis

The main purpose of the work described in this thesis is the improvement of the understanding of MAS structure and conformation, in particular of FF-MAS and its analogs. In addition, the relative stabilities of isomers of MAS analogs are studied to gain insight into the synthesis of these compounds. Heats of formation, based on molecular mechanics methods, are calculated for this purpose. This work is reported in chapter 2. Structural and conformational aspects of the side chains in the MAS structures are studied using crystallographic structure determinations and NMR spectroscopy. In chapter 3, the crystal structures of several active and inactive MAS analogs are presented and the conformations and their implication on structure-activity relationships are discussed. In addition, the conformations of the side chains of three MAS analogs in solution are discussed in this chapter. In addition to the conformation, the molecular electrostatic potential of several double bond isomers is compared using semi-empirical calculations. The results of these calculations are reported in chapter 4.

In an attempt to exploit the analogy in the SAR of side chains of FF-MAS and calcitriol, a homology model of the VDR was built, to which calcitriol and its C20 epimer were docked. SuperStar is particularly suited for this purpose. The set of geometrical distributions used in SuperStar were initially derived from CSD structures. It can be argued that the nature of non-bonded interactions in crystal structures of small molecules may differ from the way non-bonded interactions occur in a protein environment, as has been suggested by Verdonk *et al.* (1999). Therefore, the method may benefit from the use of interactions between proteins and ligands as source data taken from the Brookhaven Protein Database (PDB) and present in the IsoStar database. The implementation and validation of SuperStar based on PDB data is described in chapter 5. The homology model of the VDR and the docking results by SuperStar are described in chapter 6.

Chapter 2

Calculated Heats of Formation of Sterol Diene Isomers Compared with Synthetic Yields of Isomerisation Reactions of $\Delta^{5,7}$ Sterols

D.R. Boer, H. Kooijman, J. van der Louw[1], M. Groen, J. Kelder and J. Kroon

D.R. Boer, H. Kooijman, J. van der Louw, M. Groen, J. Kelder & J. Kroon (2000). "Calculated Heats of Formation of Sterol Diene Isomers Compared with Synthetic Yields of Isomerisation Reactions of $\Delta^{5,7}$ Sterols", *J. Chem. Soc. Perkin Trans. 2*, 1701-1704.

[1] Department of Medicinal Chemistry, N.V. Organon, Oss, The Netherlands

Abstract

Heats of formation of five series of diene sterol isomers were calculated and compared with synthetic yields of acid-catalyzed isomerisation reactions starting from $\Delta^{5,7}$ isomers. Calculations were based on molecular mechanics, using the MM3 program package. For each of the five $\Delta^{5,7}$ starting compounds, three possible reaction paths were considered, in which heats of formation were calculated for theoretically possible intermediate double bond isomers. Similar results are found for all five series. The starting compounds are found to have unfavorable heats of formation compared to all other isomers considered within one series. In general, isomerisation reactions of diene sterols ultimately yield spiro compounds when allowed to proceed for a sufficient amount of time. These compounds are found to have lowest heats of formation in each series. However, they were not formed in the reactions considered in this chapter, because the reactions were stopped after the desired isomer was formed in excess, before the spiro compounds could occur. Most compounds identified as products in the syntheses have a favorable heat of formation compared to the isomers preceding the (most stable) spiro compounds. However, when the reactions were carried out at low temperature, isomers with less favorable heats of formation could be trapped. We show that calculated heats of formation correspond well with synthetic yields and we suggest the calculation of heats of formation can be a useful tool in planning syntheses.

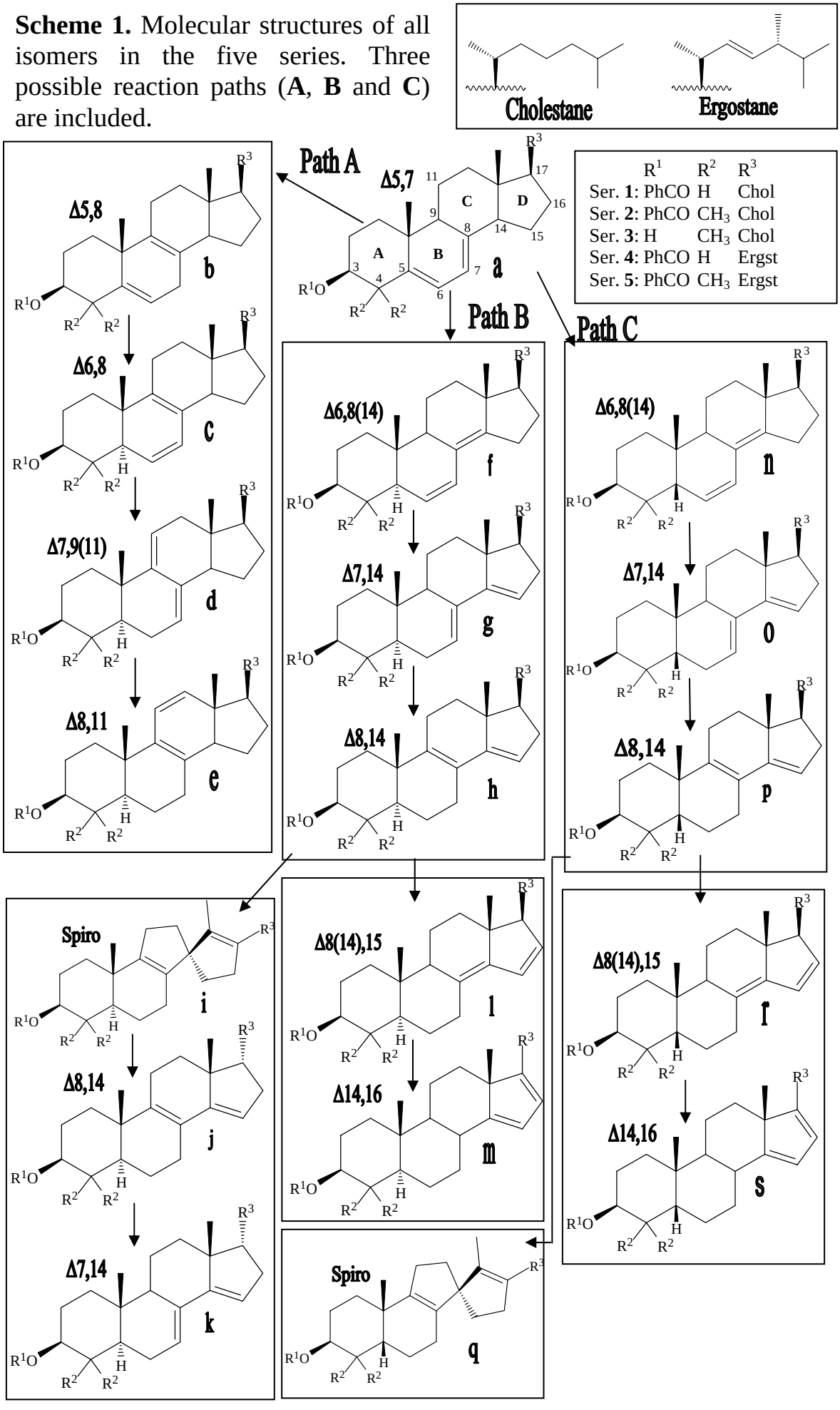
Introduction

Acid-catalyzed isomerisation of $\Delta^{5,7}$ sterols to yield, among others, $\Delta^{8,14}$ or $\Delta^{7,14}$ compounds, occurs through a pathway involving a sequence of protonation and deprotonation steps. During the reaction, other isomers are formed as well and the crude product usually is a mixture, which is undesirable. It would be convenient to be able to predict which isomers will most likely be formed and to control the relative yields of the isomers by changing reaction conditions using knowledge about the stability of the different isomers. In this chapter, yields from syntheses of sterol double bond isomers, taken from literature and obtained by us, are compared with calculated heats of formation of a range of possible isomers. Heats of formation were calculated with MM3 (Allinger *et al.*, 1989; Lii & Allinger, 1989), using the standard force field parameters. The force field is well parameterized for hydrocarbon compounds and has been shown to calculate heats of formation for a number of hydrocarbons with a standard deviation of 0.41 kcal/mole compared to experimental values (Lii & Allinger, 1989).

Since the reactions are reversible, the isomers that are formed during the reaction are assumed to be in equilibrium. As a consequence, the heat of formation differences between double bond isomers should be indicative of the isomer stability, in the same way as has been reported for global steric-energy minima of steroids and their occurrence in geological sediments as molecular fossils (Liu *et al.*, 1996). In principle, it is possible to include kinetic effects by calculating formation enthalpies for carbocations (Van Duin, 1998). These are not considered in this chapter.

Five compound series (**1**, **2**, **3**, **4** and **5**) are considered here. Scheme 1 is a key to the structural formulas of compounds of those series. The first series (**1**) consists of structures derived from 3-benzoyl-$\Delta^{5,7}$-cholesterol. The second series is analogous to series **1** with two extra methyl groups attached to C4 of ring A. Series **3** consists of 3β-hydroxy analogs of series **2**. Replacement of the cholestane side chain of the compounds from series **1** and **2** by an ergostane side chain gives the compounds for series **4** and **5**, respectively. The syntheses of compounds **2f**, **3f** and **2h**, **3h** (the 5α-$\Delta^{6,8(14)}$ and 5α-$\Delta^{8,14}$ isomers respectively) are reported here. Other authors reported synthetic data on a subset of structures from series **1** (Dolle *et al.*, 1988; Wilson & Schroepfer, 1988), **4** (Dolle *et al.*, 1988; Dolle & Kruse, 1986) and **5** (Dolle & Kruse, 1986). The latter are of significant biological and pharmaceutical interest, as they are related to FF-MAS, a signaling molecule that induces resumption of meiosis of the human oocyte, the immature egg cell (Byskov *et al.*, 1995).

Scheme 1. Molecular structures of all isomers in the five series. Three possible reaction paths (**A**, **B** and **C**) are included.

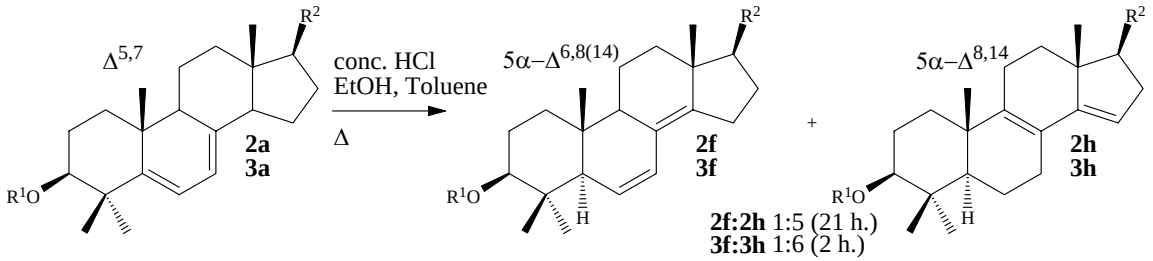


Three reaction paths were considered for each series (**A**, **B** and **C** in scheme 1). Path **A** leads to isomers containing (conjugated) double bonds in rings B and C, path **B** to 5α-isomers with conjugated double bonds in rings B, C and D and path **C** to the corresponding 5β-isomers of path **B**. Depending on the particular reaction conditions within each series, isomers **f**, **g**, **h**, and **p** were found as end products. In general, isomerisation reactions yield spiro compounds analogous to **i** and **q** when allowed to progress for a considerable amount of time (Liu *et al.*, 1996).

Methodology

Syntheses

The synthetic route used to prepare the 5α-$\Delta^{6,8(14)}$ and 5α-$\Delta^{8,14}$ isomers of the 3β-OH and benzoylated sterol cholestane series (**2f**,**h** and **3f**,**h**) is shown in scheme 2.



Scheme 2. Synthesis and relative yields of compounds **2f**, **2h** (R^1 = PhC(=O)-), **3f** and **3h** (R^1 = H). R^2 corresponds to the cholestane side chain, shown in scheme 1.

(3β,5α,20R)-4,4-Dimethyl-cholest-8,14-dien-3β-ol (3h) and (3β,5α,20R)-4,4-dimethyl-cholest-6,8(14)-dien-3β-ol (3f). A mixture of (3β,20R)-4,4-dimethyl-cholest-5,7-dien-3β-ol (**3a**, 2.32 g, 5.63 mmol), ethanol (96 %, 42 ml), toluene (6 ml), and concentrated hydrochloric acid (6 ml) was heated under reflux for 2 h. After cooling, the mixture was poured into a saturated aqueous solution of $NaHCO_3$. The product was extracted into diethyl ether. The combined organic phases were washed with brine, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product (3.69 g) was purified by column chromatography to yield a mixture of **3h** and **3f** (2.83 g, ratio 6:1). **3h**: ^{1}H NMR ($CDCl_3$): δ 5.35 (m, 1H), 3.24 (m, 1H), 2.43-0.80 (m), 1.04 (s, 3H), 1.02 (s, 3H), 0.93 (d, 6.2 Hz, 3H), 0.88 (s, 3H), 0.85 (s, 3H), 0.83 (s, 3H), 0.81 (s, 3H); **3f**: ^{1}H NMR ($CDCl_3$): δ 6.23 (dd, J = 10.0 and 3.0 Hz, 1H), 5.61 (dm, J = 10.0 Hz, 1H), 3.30 (m, 1H), 2.34 (m, 2H), 2.00 (m, 1H), 1.90 (m, 2H), 1.79-0.97 (m), 1.08 (s,

3H), 0.94 (d, J = 8.6 Hz, 3H), 0.89 (s, 3H), 0.88 (s, 3H), 0.86 (s, 3H), 0.83 (s, 3H), 0.68 (s, 3H).

(3β,5α,20R)-4,4-Dimethyl-cholest-8,14-dien-3β-ol benzoate (2h) and (3β,5α,20R)-4,4-dimethyl-cholest-6,8(14)-dien-3β-ol benzoate (2f). A mixture of (3β,20R)-4,4-dimethyl-cholest-5,7-dien-3β-ol benzoate (**2a**, 2.58 g, 5.0 mmol), ethanol (96 %, 70 ml), toluene (20 ml), and concentrated hydrochloric acid (10 ml) was heated under reflux for 1 h. The reaction mixture was worked up as described above to produce a mixture of **2h** and **2f** (2.61 g, ratio 1:5). Prolonged reaction results in a shift in ratio between **2h** and **2f** *via* 3:2 (5 h heating) to 5:1 (21 h heating). **2h**: ^{1}H NMR ($CDCl_3$): δ 8.06 (m, 2H), 7.61-7.40 (m, 3H), 5.38 (m, 1H), 4.75 (m, 1H), 2.48-0.75 (m); **2f**: ^{1}H NMR ($CDCl_3$): δ 8.06 (m, 2H), 7.61-7.40 (m, 3H), 6.27 (dd, J = 10.4 and 3.0 Hz, 1H), 5.62 (dd, J = 10.4 and 2.0 Hz, 1H), 4.82 (dd, J = 10.8 and 5.0 Hz, 1H), 2.48-0.75 (m).

Heat of formation calculations

MM3 calculates heats of formation based on a bond-energy increment method. For each bond appearing in the structure, a fixed contribution is added by the program, which depends on the atom type and substitution number of the atoms that form the bond. However, the formation enthalpy also depends on conformation and electronic configuration. These are taken into account as much as possible using formula [1].

$$\Delta H_f = \Sigma E_{bonds} + E_{st} + E_{SCF} + E_{T/R} + E_{TORS} + E_{POP} \qquad [1]$$

The ΣE_{bonds} term represents the bond and structural increments of the molecular structure. E_{st} represents the steric energy of the molecule. If a conjugated system is present in the structure, the self-consistent field energy (E_{SCF}) for that system has to be included as well. Translation and rotational energy are included with the $E_{T/R}$ term that is taken to be 4RT. The E_{TORS} term represents the corrections needed in case wide torsional motions are possible in the molecule. When this occurs, the harmonic approximation for the potential energy with respect to the internal coordinate is no longer valid. Therefore, the differences in energy levels in the potential well will be smaller than those calculated using the harmonic approximation. The MM3 program does not calculate the torsional correction terms. However, Wertz and Allinger (1979) proposed a value of 0.42 kcal/mole for the E_{TORS} term, based on theoretical considerations by Pitzer and Gwinn (1942). In the calculations described here, the heats of formation were corrected by adding the E_{TORS} term multiplied by the number of rotations for which the harmonic approximation was not valid (five times the above mentioned correction term for series **1**,**2**,**4** and **5**, three times for series **3**). E_{POP} represents the energy contribution originating from the fact that higher energy conformations are present. It is calculated using a Boltzmann weighted

summation of the heats of formation for an ensemble of conformers generated at 298 K. The ensemble was generated by randomly changing the atom positions in the molecule 200 times using the stochastic search algorithm in MM3. The conformations found were minimized and only unique ones were retained.

Results & Discussion

Calculated heats of formation for compound series **1** to **5** are given in Table 1, along with the available relative synthetic yields obtained by us and taken from the literature. Dolle *et al.* (1988) do not give explicit yields for the compounds of series **1**. However, it was stated that similar results were obtained as for the synthesis of the 17-ergostane-3β-benzoylated series (series **4**), reported in the same article.

Table 1. Heat of formation (ΔH_f in kcal/mole) and relative yields (Y_r, in % where possible) of the compounds (denoted by letter according to scheme 1) of the five isomer series.

	Series **1**			Series **2**		Series **3**		Series **4**			Series **5**	
	ΔH_f	Y_r^{+}	Y_r^{*}	ΔH_f	$Y_r^{\#}$	ΔH_f	$Y_r^{\#}$	ΔH_f	Y_r^{+}	$Y_r^{\$}$	ΔH_f	$Y_r^{\$}$
a	-113.0			-120.3		-101.7		-92.9			-100.8	
Path A												
b	-115.7			-122.4		-103.6		-96.0			-103.1	
c	-113.1			-120.3		-101.3		-93.6			-101.0	
d	-118.2			-124.2		-105.1		-98.6			-104.5	
e	-113.9			-119.8		-100.7		-94.1			-100.1	
Path B												
f	-119.7		*3*	-127.5	*20*	-107.6	*15*	-100.2			-106.6	
g	-119.2	*Trace*	*48*	-124.3		-105.7		-99.5	*4*	*75*	-104.7	*75*
h	-121.8	*Major*	*27*	-127.6	*80*	-108.7	*85*	-102.4	*72*	*25*	-107.5	*25*
i	-128.5			-135.2		-116.3		-108.0			-114.1	
j	-114.7			-120.9		-103.1		-96.6			-102.3	
k	-117.8			-123.6		-105.9		-99.5			-105.2	
l	-117.9			-124.7		-105.4		-98.2			-104.0	
m	-124.7			-133.0		-113.3		-108.0			-113.3	
Path C												
n	-120.4			-128.3		-108.8		-98.4			-109.4	
o	-117.5			-122.9		-103.9		-94.9			-102.5	
p	-122.9	*Minor*	*22*	-127.4		-107.9		-102.5	23		-107.7	
q	-130.2			-135.0		-115.4		-108.7			-113.8	
r	-117.7			-121.5		-101.3		-97.4			-101.7	
s	-125.7			-124.5		-109.5		-103.3			-109.6	

[+] Reported by Dolle *et al.* (1988)
[*] Reported by Wilson & Schroepfer (1988)
[#] Synthesis according to scheme 2
[$] Reported by Dolle & Kruse (1986)

The heat-of-formation calculations were consistent for all five series with respect to the relative stabilities of the different isomers within a series. The starting compounds (**1a**-**5a**) were found to have the most unfavorable heat of

formation compared to almost all other isomers within a series. Compounds in path **A** (scheme 1) generally have a more unfavorable heat of formation than those of paths **B** and **C** for each series. Given the finding that the isomers (**b**-**e**) in path **A** are not formed in appreciable yields in any of the syntheses, the calculation results are in agreement with the synthetic data. For each series, spiro compounds **1i**-**5i** and **1q**-**5q** were found to have the lowest heats of formation in paths **B** and **C**, respectively. It is known that isomerisation reactions, such as those described in this chapter, yield spiro compounds like **1i**-**5i** and **1q**-**5q** when the reactions are allowed to proceed for longer time spans (Liu *et al.*, 1996). However, the reactions described in this chapter and those reported in literature (Dolle *et al.*, 1988; Wilson & Schroepfer, 1988; Dolle & Kruse, 1986) were stopped after the desired isomer was formed in excess over the other isomers. In all syntheses, the isomer of interest occurred 'up-stream' from the spiro compounds in the reaction paths **B** and **C** (scheme 1). Therefore, the spiro compounds and the 5α- and 5β-$\Delta^{14,16}$ isomers **m** and **s**, which in most cases also have favorable heats of formation, are not formed.

In the synthesis of compounds in series **1** and **4**, reported by Dolle *et al.* (1988), the major products are the 5α- and 5β-$\Delta^{8,14}$ isomers (**1h**,**4h** and **1p**,**4p** respectively). These isomers are more stable than any of the other isomers within path **B** and **C** respectively, except for the 5α- and 5β-$\Delta^{14,16}$ isomers **m** and **s** and spiro compounds **i** and **q** (see discussion above). The syntheses reported in this chapter (series **2** and **3**) yield the 5α-$\Delta^{6,8(14)}$ and 5α-$\Delta^{8,14}$ isomers **f** and **h** as major products. These isomers have lowest heats of formation in path **B**, with the exception of the $\Delta^{6,8(14)}$ isomer **m** and spiro compound **i** (see above). So, for the syntheses shown in scheme 2 and those reported by Dolle *et al.* (1988), the major products are the more stable compounds (in terms of heat of formation) occurring in the first part of the formation paths **B** and **C**, which lead to the 5α- and 5β-$\Delta^{8,14}$ isomers **h** and **p**. It is interesting to note that compounds **2f** and **2h** have quasi identical calculated heats of formation, yet their yield ratio changes from 5:1 after one hour to 1:5 after 21 hours. This phenomenon is not caused by selective decomposition of compound **2f**, since no appreciable loss of total product mass was observed (data not shown). It follows that compound **2f** is converted into compound **2h** and that the latter compound actually has a lower heat of formation. The relatively slow conversion of **2f** into **2h** is probably caused by a kinetic barrier between the two compounds which has to be overcome to reach the lower heat of formation well of the latter compound.

The synthesis of compounds in series **1** described by Wilson & Schroepfer (1988) yields the 5α-$\Delta^{7,14}$ isomer (**1g**) as the major product. As was the case in the syntheses described by Dolle *et al.* (1988), the 5α-$\Delta^{8,14}$ and 5β-$\Delta^{8,14}$ isomers (**1h** and **1p**) were also formed in considerable amounts. However, the 5α-$\Delta^{7,14}$ isomer (**1g**) has an unfavorable heat of formation compared to both **1h** and **1p**, but is isolated in highest yields. Similar results are found for the syntheses of

compounds in series **4** and **5**, reported by Dolle & Kruse (1986). The 5α-$\Delta^{7,14}$ and 5α-$\Delta^{8,14}$ isomers are found in a 3:1 yield ratio (**4g**:**4h** and **5g**:**5h**). When comparing reaction temperatures of the syntheses that yield the **g** isomer (Dolle & Kruse, 1986; Wilson & Schroepfer, 1988) with those that yield the **h** isomer in highest yields (Dolle *et al.*, 1988, and the syntheses reported in this chapter), we found that the former are considerably lower (-60 to -40 ºC) than the latter (25-80 ºC). We therefore hypothesize that at low temperature, the isomers formed sequentially in the reaction path **B** are not in equilibrium during their formation. Thus, at low temperature, the reaction is slowed down considerably and the isomers are trapped. Dolle *et al.* (1988) describe the temperature-dependent formation of compounds **4g** and **4h**, and report that at -30 ºC, the $\Delta^{7,14}$ isomer 4**g** was predominant, whereas the $\Delta^{8,14}$ isomer (**4h**) was formed in excess when raising the reaction temperature to 25 ºC. In addition, Wilson & Schroepfer (1988) report increased formation of **1h** at the expense of **1g** after raising the reaction temperature from -55 ºC to 15 ºC. These findings support our hypothesis that at low temperature, equilibrium is not reached and the isomers preceding the stable 5α-$\Delta^{8,14}$ isomer (**h**) occurring in path B can be trapped.

In some cases, the 5β-isomers of path **C** are more stable in terms of calculated heats of formation than the corresponding 5α-isomers of path **B**, whereas the latter are formed in higher yields. This holds for isomer pairs **1h** and **1p** (5α-$\Delta^{8,14}$ and 5β-$\Delta^{8,14}$), **2f** and **2n** (5α-$\Delta^{6,8(14)}$ and 5β-$\Delta^{6,8(14)}$), **3f** and **3n** (5α-$\Delta^{6,8(14)}$ and 5β-$\Delta^{6,8(14)}$). It is well known that double bond sterols and steroids are more susceptible to reactions on the α-side than to the β-side (see also chapter 4). The C19 methyl group and the methyl group attached to C4 projecting towards the β-side of the molecule shield the Δ^5 double bond at this side of the molecule, while the α-side is far less sterically hindered. Therefore, protonation of C5 will occur in preference at the α-side of the molecule, even if the heats of formation of the 5β isomers are more favorable.

According to the reaction mechanism, the 17α-isomer **j** in path **B** could be formed from isomer **i**. In the article by Dolle *et al.* (1988), an unknown compound was in first instance thought to be the 5α-17α-$\Delta^{8,14}$ isomer **4j**. However, crystal structure analysis showed that this isomer was in fact the 5β-17α-$\Delta^{8,14}$ isomer (**4p**), which is more obvious when comparing the heats of formation of these isomers. Knowledge of the relative stabilities of the different isomers that can occur during synthesis may therefore help to identify products with unknown structure.

Conclusion

Calculated heats of formation of the sterol isomer compounds considered are consistent with the formation of these compounds in syntheses. Firstly, the $\Delta^{5,7}$ starting compounds (**1a-5a**) have most unfavorable heats of formation compared to all other compounds within a series. Secondly, compounds in path **A**, which were found to have unfavorable heats of formation compared to compounds in paths **B** and **C**, are not formed. Most favorable heats of formation are found for spiro compounds **i** and **q**, which are well known end products of sterol isomerisation reactions. These are not formed if the reactions are stopped at an isomer occurring up-stream in the formation pathways **B** and **C**. The same holds for the $\Delta^{14,16}$ isomers **m** and **s**, which are also found to have favorable heats of formation.

Syntheses of compounds in series **2** and **3**, described in this chapter, and syntheses of compounds in series **1** and **4,** reported by Dolle *et al.* (1988), yield isomers **f**, **h** and **f**, **h**, **p**, respectively, as major products. The heats of formation of these compounds are the most favorable in the reaction routes leading towards spiro compounds **i** and **q**. When the temperature during the reaction is lowered, it is possible to trap the more unstable 5α-$\Delta^{7,14}$ isomer (**g**), which was the case in the syntheses reported by Wilson & Schroepfer (1988), *i.e.* series **1,** and those reported by Dolle & Kruse (1986), *i.e.* series **4** and **5**.

The calculated heats of formation for isomeric compounds correlate well with synthetic yields of isomerisation reactions. When used for predictions, they may lead to new insights when planning a synthesis. Thus, it can be a useful tool for synthetic chemists designing a new experiment. In addition, the calculated stabilities of the different isomers can help to identify unknown products.

Chapter 3

Solid State and Solution Conformations of Follicular Fluid-Meiosis Activating Sterols (FF-MAS) Related Compounds

D.R. Boer, H. Kooijman, J. van der Louw[1], M. Groen[1], J. Kelder and J. Kroon

Part of this chapter is accepted for publication: D.R. Boer, H. Kooijman, J. van der Louw, M. Groen, J. Kelder & J. Kroon. “Comparison of crystal structure conformations of five follicle fluid meiosis activating sterol related active and inactive compounds”, *Acta Cryst. C.*, in the press.

[1] Department of Medicinal Chemistry, N.V. Organon, Oss, The Netherlands

Abstract

The crystal structure conformations of five analogs of Follicle Fluid-Meiosis Activating Sterols (FF-MAS) are described, as well as the solution conformations of three of these. The molecular models derived from these experiments are used to evaluate the importance of the skeleton and side-chain conformations for activity. The skeleton conformations in the structures presented in this chapter are compared to those of crystal structures retrieved from the Cambridge Structural Database of compounds with an androstane skeleton containing a varying number of double bonds at different positions. It is found that the conformation of the skeletal rings as induced by the double bond system is probably not the determinant factor for activity. The conformation of the side chains of the three compounds that were studied both in solution as well as in the solid state are very similar for each compound, respectively. One (active) compound with natural side chain stereochemistry is probably also present in a second conformation in solution, which resembles the solid state and solution conformation of an active analog with inverted C20 stereochemistry. The crystal structure conformation found for sterol analogs with unnatural configuration of C20 is therefore a likely candidate for the active conformation, as is the bound conformation of the endogenous ligand of the Vitamin D_3 Receptor (VDR), observed in the crystal structure of its complex with the VDR. It is likely that the side chain conformation responsible for activity is not the fully extended conformation found in crystal structures of sterols with natural configuration of the side chain chiral centers.

Introduction

Maturating human egg cells are arrested several times during development. FF-MAS (compound **1** in Figure 1) is able to induce resumption of meiosis during one of these arrests and is therefore an interesting candidate for the development of novel contraceptive agents. Several MAS derived compounds have been synthesized and tested on arrested mouse oocytes. Special emphasis in the design of these MAS analogs was put on the modification of the side chain and variation of the number of double bonds and their position in the four-ring skeleton. Knowledge of the side chain conformational behavior is imperative for understanding the relation between structure and the resulting activity. The study of the conformational behavior of the skeleton and side chains of these analogs is expected to aid in the rational design of more potent agonists and antagonists.

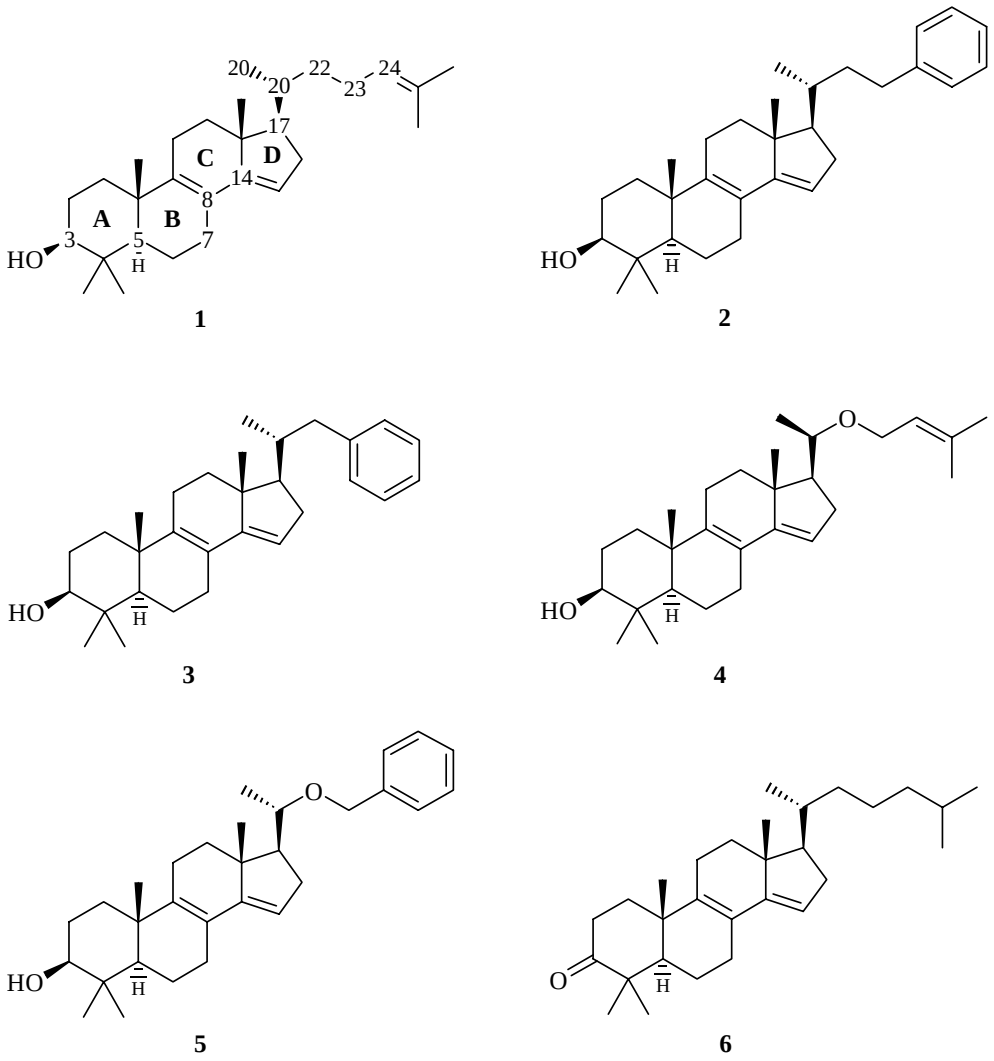


Figure 1. Molecular structures of compounds **1-6**.

In this chapter, experimentally determined conformations of five FF-MAS analogs (compounds **2**-**6** in Figure 1) are presented for this purpose. The compounds all have a $\Delta^{8,14}$ double bond configuration and all but **6** have a 3β-OH moiety. Compounds **2**-**4** are active, **5** and **6** are inactive in *in vitro* tests. The inactivity of compound **6** is most likely caused by the absence of the mandatory 3β-OH group. Very few double bond isomers of FF-MAS were found to yield active compounds. Those that are active have double bonds at positions 8 and 14 or an aromatic C ring (Grønvald *et al.* 1997). Sterols with Δ^{5}, $\Delta^{5,7}$, $\Delta^{7,9(11)}$, Δ^{8} and $\Delta^{8(14)}$ double bond configurations gave inactive compounds. Thus, the position and number of double bonds appear to be an important determinant for activity, which leads to the question whether there is a correlation between activity and the conformation of the skeleton as induced by the double bond configuration.

FF-MAS and related sterols have a similar structure-dependent activity with respect to the stereochemistry of C20 as calcitriol (1α,25-dihydroxy-vitamin D_3), the endogenous ligand of the vitamin D receptor. It was found that a derivative of calcitriol with inverted stereochemistry at C20 ('unnatural configuration') has similar affinity to the Vitamin D Receptor as calcitriol with 'natural configuration' at C20, and even has an increased transcriptional activity (Boullion *et al.* 1995, Liu *et al.* 1997, Väisänen *et al.* 1999). Similarly, analogs of FF-MAS with unnatural C20 stereochemistry are able to induce meiosis, provided that C22 is replaced by an oxygen atom. Compound **4** exemplifies this. In contrast, 22-oxa sterols with the C20 configuration as in FF-MAS are in general inactive, as is the case for compound **5**. Crystal structure determinations and Nuclear Magnetic Resonance (NMR) experiments may provide insight into the conformational behavior of the side chains of the compounds under study. We present crystal structures of five FF-MAS derived sterols, compounds **2**-**6** in Figure 1. In addition, the conformations of **3**-**5** in solution were determined using NOESY and ROESY NMR experiments.

Methodology

Crystal structure determinations

Colorless, single crystals of compounds **2**, **3** and **5** suitable for X-ray analysis were obtained by vapor diffusion of a pyridine solution against CH_3CN. Crystals formed after several days. Colorless, single crystals of compound **4** and **6** were obtained by slow evaporation of ethanol solutions.

Data was collected on Nonius KappaCCD device using COLLECT (Nonius, 1998). Cell refinement of compounds **2**, **4** and **5** was done with DiRaX (Duisenberg, 1992), data reduction with EVAL14 (Duisenberg, 1998). Cell refinement and data reduction of **3** and **6** was done with Denzo 1.11.0

(Otwinowski & Minor, 1997). All structures were solved by direct methods with SHELXS-97 (Sheldrick, 1997b) and the structures were refined on F^2 using SHELXL-97 (Sheldrick, 1997a). The material was prepared for publication using PLATON (Spek, 2001). Details on the structure solution and refinement are given in Table 1. Figure 2 shows the ORTEP drawings of compounds **2-6**. For each structure, the ORTEP drawing of one of the independent molecules is given. If disorder occurred in a structure, the major component is shown.

Table 1. Crystallographic and refinement data for compounds **2-6**.

Compound	**2**	**3**	**4**	**5**	**6**
Formula	$C_{31}H_{44}O$	$C_{30}H_{42}O$[a]	$C_{28}H_{44}O_2$	$2(C_{30}H_{42}O_2).$ ½H_2O	$C_{29}H_{46}O$
Molecular weight	432.66	418.64[a]	412.63	1756.56	410.66
Space Group	$P2_1$	$C2$	$P6_2$	$C2$	$P2_12_12_1$
Cell dimensions	a=7.271(3) Å	a=23.785(9) Å	a=24.66(1) Å	a=64.902(4) Å	a=10.574(1) Å
	b=9.733(5) Å	b=13.610(3) Å		b=7.293(1) Å	b=20.941(2) Å
	c=35.98(2) Å	c=19.080(7) Å	c=7.232(10) Å	c=11.145(1) Å	c=23.694(2) Å
	β=93.280(40)°	β=121.93(3)°		β=96.06(1)°	
Z	4	8	6	4	8
Density (calc.)	1.130 Mg m^{-3}	1.061 Mg m^{-3} [a]	1.079 Mg m^{-3}	1.112 Mg m^{-3}	1.040 Mg m^{-3}
Radiation	Mo Kα	Mo Kα	Mo Kα	Mo Kα	Mo Kα
Wavelength λ	0.71073 Å	0.71073 Å	0.71073 Å	0.71073 Å	0.71073 Å
θ_{max}	25.25°	25.25°	24.77°	25.25°	25.25°
hkl limits	*h*: -8 to 8	*h*: -28 to 28	*h*: -28 to 18	*h*: -64 to 77	*h*: -12 to 12
	k: -11 to 9	*k*: -15 to 15	*k*: -28 to 28	*k*: -8 to 7	*k*: -25 to 25
	l: -43 to 41	*l*: -22 to 22	*l*: -8 to 6	*l*: -13 to 13	*l*: -28 to 28
# measured refl.	15221	29708	13233	17152	37782
# unique refl.	4842	4917	2343	5110	5260
R_{int}	0.0477	0.0487	0.0454	0.0478	0.0990
Temperature	150 K	150 K	150 K	293 K	293 K
Abs. coefficient μ	0.066 mm^{-1}	0.062 mm^{-1}	0.062 mm^{-1}	0.068 mm^{-1}	0.060 mm^{-1}
R1	0.044	0.046	0.075	0.045	0.052
wR2	0.116	0.119	0.140	0.106	0.113
Nr. of parameters	590	559	275	596	571
Goodness of fit	1.097	1.093	1.046	1.034	1.019
Residual density (e $Å^{-3}$)	-0.15≤ρ≤ 0.17	-0.29≤ρ≤ 0.27	-0.36≤ρ≤ 0.52	-0.12≤ρ≤ 0.12	-0.12≤ρ≤ 0.16

[a] without disordered solvent contribution

Due to the absence of significant anomalous dispersion, the absolute structure cannot be determined reliably from the diffraction experiment. The configuration was chosen such that for each compound the skeleton stereochemistry corresponds to those of other androstanes. Friedel pairs were merged. H atoms were introduced at calculated positions, riding on their carrier atoms. The methyl groups were refined as rigid groups, allowing for rotation around the C-C bonds. The hydroxyl hydrogens were also treated as rigid

groups. Their initial positions were determined from a circular residual density map. H atoms were refined with a fixed isotropic displacement parameter related to the value of the equivalent isotropic displacement parameter of the carrier atom by a factor of 1.5 for the methyl and hydroxyl H atoms and 1.2 for all other H atoms. The structure of compound **3** contains a disordered solvent region (15 electrons, 170 $Å^3$), which could not be interpreted. The contribution of this electron density to the structure factors was taken into account following the Bypass procedure (Van der Sluis & Spek, 1990), implemented in PLATON as the SQUEEZE option. The two independent molecules in the crystal structure of **5** are disordered. A flip of ring A from a chair to a twist-boat conformation in one of the independent molecules causes, through intermolecular contacts, a 30° rotation over the C24 to C25 dihedral in the other independent molecule. The occupancy factors of the disorder components in both molecules were therefore linked.

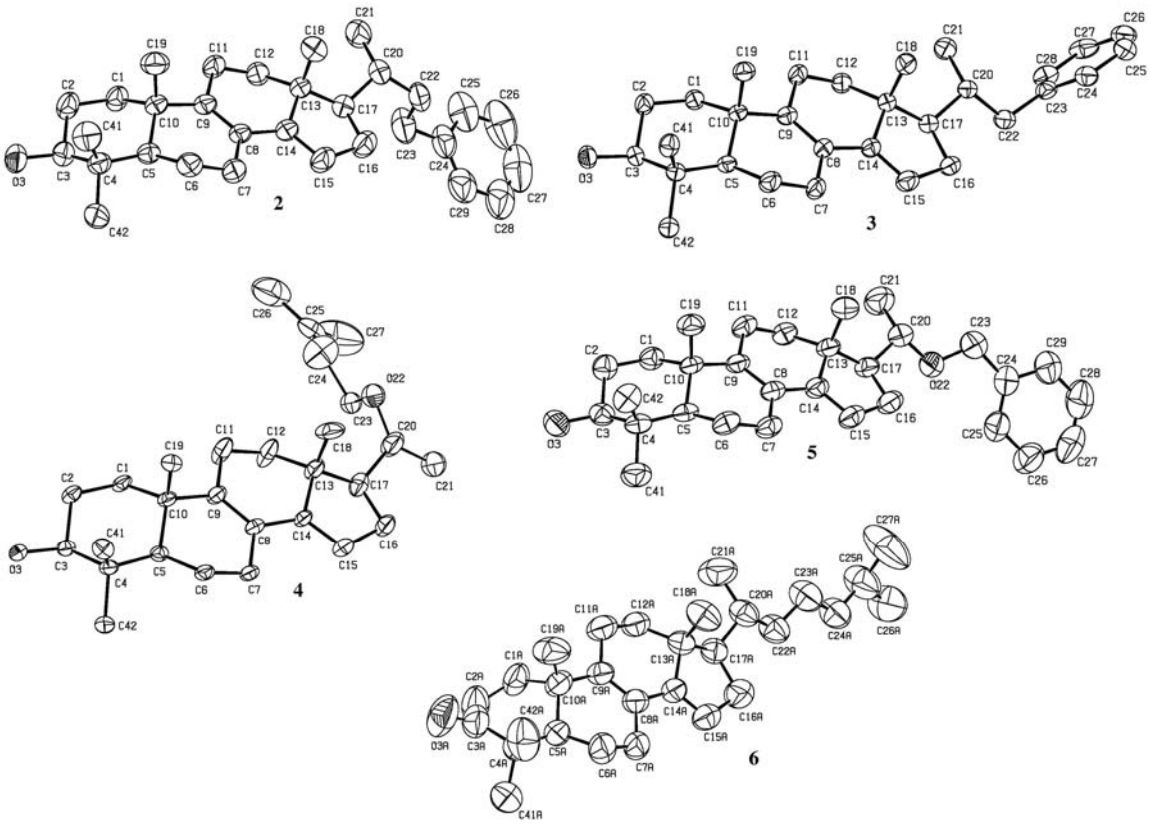


Figure 2. ORTEP drawings of one independent molecule from each of the crystal structures of **2**-**6** at probability levels of 50%. If disorder is present, the drawing of the major component is shown. The displacement parameters of the other independent and disordered molecules are similar to those shown here.

The skeletal conformations of compounds **2**-**5** and that of compound **6** are compared to those of double bond sterol isomers taken from the Cambridge

Structural Database (CSD version October 2000, 224400 entries, Allen & Kennard, 1993). The CSD was searched for error-free structures with a crystallographic R1 factor lower than 0.1, containing the four-ring sterol topology with androstane stereochemistry, with a varying number of double bonds at different positions. Through-bond connections between atoms in the fragments other than those defined were not allowed. The skeletons were divided into classes according to their double bond configurations. For each combination of two classes, the individual skeleton fragments of one class were superposed on the individual skeleton fragments of the other class using InsightII. The minimum and maximum RMS-deviations of the superpositions were recorded.

NMR experiments

One- and two-dimensional ^{1}H NMR spectra were obtained at 25° on a Bruker DRX500 (500.1 MHz for ^{1}H) spectrometer (see the appendix of this thesis). The sterols were dissolved in $CDCl_3$, purchased from Cambridge Isotope Labeling Inc. For compound **3** and **5**, two TOCSY NMR spectra (Bax & Davis, 1985) were measured with mixing times of 7 and 100 ms, respectively, in phase-sensitive mode using echo/antiecho gradient selection. For compound **3**, an off-resonance ROESY spectrum (Desvaux *et al.*, 1995) was measured using an adiabatic shaped spin lock pulse with a power corresponding to 8.4 kHz, with a mixing time of 300 ms. The spectrum was measured in phase sensitive mode using States-TPPI. The carrying frequency during spin-lock was adjusted to obtain an average spin lock angle $<\theta>=45°$. In this way, ROESY and NOESY type transfer are equally present in the spectrum. For compound **4**, a COSY spectrum with 40 ms mixing time and a NOESY spectrum with a mixing time of 300 ms were measured. A NOESY spectrum was measured for compound **5** with a mixing time of 500 ms in phase sensitive mode using States-TPPI.

Chemical shifts are expressed in ppm downfield from the signal for internal $Si(Me)_4$. NOE cross-peak intensities were translated into interproton distances using the distance between H15 and H16α in the crystal structures of compounds **3** and **5**, after normalization of the X-ray hydrogen positions. The NOE intensities of the cross peaks for compound **4** were translated into interproton distances using the H3α-H5α distance as found in the crystal structure, after normalization of the hydrogen positions. A background noise level for all NMR spectra was estimated by integration of regions devoid of cross peaks. A cut-off cross-peak intensity is defined as 1.5 times this background level, from which the cut-off distance was determined (see below).

Assignment

The signals of the skeleton protons and those attached to the C20, C21 and olefinic protons in the side chains of all three compounds could be assigned

unambiguously using the TOCSY, ROESY and NOESY NMR spectra. However, the signals of the diastereotopic protons at C23 of compounds **4** and **5** and those of C22 of compound **3** could not be assigned on the basis of the TOCSY NMR spectra alone. Assignment was based on consistency of the observed NOESY or ROESY cross peak intensities. Thus, according to the ROESY spectrum of compound **3**, one of the protons attached to C22 was closer to H17, while the other was found to be closer to H16α. The intensities of these peaks can be satisfied if the assignment is such that H22R is the proton close to H17, while H22S is the proton close to H16α. However, it could be that the cross peaks originate from different conformations present in solution. If this had been the case, one of the H22 cross peaks with H20 would be expected to have higher intensity than the other. In the measured NMR spectra, these intensities are similar, implying that the cross peaks with H17 are predominantly determined by one conformation. As a result, the initial assignment given above was considered to be correct. Similarly, the H23R and H23S assignments for compounds **4** and **5** were derived from their respective NOESY NMR spectra.

Model generation and comparison

Because the number of degrees of freedom of the side chains under consideration is relatively small, a systematic search was applied. Preferred orientations of the phenyl groups and the olefinic group were derived from the CSD. Conformations were generated using the CFF91 force field in InsightII (v. 98.0, MSI, San Diego, USA). For compounds **4** and **5**, torsion angles C13-C17-C20-O22, C17-C20-O22-C23 and C20-O22-C23-C24 were set to 180°, -60° and 60°. In addition, torsion angle O22-C23-C24-C25 of compound **4** was set to -100° and 100°, and torsion angle O22-C23-C24-C25 of compound **5** was set to 90° and -90°. For compound **3**, torsions C13-C17-C20-C22 and C17-C20-C22-C23 were set to 180°, -60° and 60°. The torsion angle C20-C22-C23-C24 was set to 90° and -90°. The resulting conformations were minimized while constraining the side-chain torsions. For compounds **4** and **5**, 54 conformers were generated. For compound **3**, 18 conformers were generated.

The protons attached to carbons on the side chains as well as those attached to C12, C16, C17 and C18 were included in the derivation of the side chain conformations. In cases where overlap occurred of chemical shifts (*e.g.* in methyl groups), the protons responsible for the signals were grouped into a single set. The cross-peak intensities of all possible combinations of each of these proton sets were determined from the NOESY and ROESY NMR spectra. Calculated intensities of interactions between proton sets are derived from the modeled conformations using formula [1].

$$A_{ij}^{calc}(n) = S * \left(\sum_k \sum_l r_{kl}(n)^{-6} \right) \qquad [1]$$

$A_{ij}^{calc}(n)$ is the calculated intensity between the proton sets i and j for conformer n, S is the scale factor determined by calibration (see above), k and l represent the protons in group i and j respectively, and $r_{kl}(n)$ is the distance between two protons k and l in conformer n.

Conformers were mixed using formula [2], up to combinations of four non-identical conformations.

$$A_{ij}^{calc} = \sum_n f_n * A_{ij}^{calc}(n) \qquad [2]$$

A_{ij}^{calc} is the calculated intensity of interaction of proton sets i and j for a combination of conformers and f_n is the fraction of conformer n. For each combination of conformers, the fractions of the individual conformers were incremented with 0.1 in a stepwise fashion. The combinations of fractions that gave an optimum value for R_r (see below) was retained.

For each of the conformers or conformer combinations, an R-factor based on distance criteria was calculated (Gonzalez *et al.*, 1991), similar to that used as quality measure of crystal structures, as shown in [3].

$$R_r = \left(\frac{\sum_i \sum_j \left| \left(A_{ij}^{calc}\right)^{-1/6} - \left(A_{ij}^{exp}\right)^{-1/6} \right|^2}{\sum_i \sum_j \left(A_{ij}^{exp}\right)^{-1/3}} \right)^{1/2} \qquad [3]$$

R_r is quality measure for a certain combination of conformers and A_{ij}^{exp} is the experimentally determined NOE intensity for proton sets i and j. If a NOE cross peak was not observed for a particular interaction between two sets of protons, A_{ij}^{exp} was set to the cut-off intensity. If the distance between two proton sets i and j was larger than the cut-off distance, $\left|\left(A_{ij}^{calc}\right)^{-1/6} - \left(A_{ij}^{exp}\right)^{-1/6}\right|$ was set to zero. In case the calculated distance between two proton sets was smaller than the cut-off distance, $\left|\left(A_{ij}^{calc}\right)^{-1/6} - \left(A_{ij}^{exp}\right)^{-1/6}\right|$ was determined with A_{ij}^{exp} set to the cut-off intensity. This way, the final quality measure includes distances that are expected to give a cross peak in the NOESY and ROESY NMR spectra but that are not observed (referred to as lacking NOE's).

Results & Discussion

Crystal structure conformations

Examination of the skeleton structure of the different sterol double bond isomers is focused on comparison of double bond configurations that can yield active compounds with those that did not result in active compounds. Thus, skeleton classes are compared instead of individual molecules. The RMSd's of superposition of skeletal fragments of $\Delta^{8,14}$ structures (7 fragments) with those of Δ^{5} (120 fragments), $\Delta^{5,7}$ (2 fragments) and fully saturated sterols (60 fragments) range from 0.23-0.49, 0.23-0.39 and 0.15-0.47, respectively. These values lie in the same range as the RMSd of the superposition of fragments taken from the same class of double bond isomers, which are 0.14-0.47, 0.19-0.42 and 0.0-0.44 for the Δ^{5}, the $\Delta^{5,7}$ and the fully saturated structures respectively. The RMSd of superposition of skeletal fragments of $\Delta^{8,14}$ (7 fragments) and an aromatic C ring compound (Ferguson *et al.* 1982) ranges from 0.21-0.26. It can be concluded that differences in conformation between the $\Delta^{8,14}$ structures and the aromatic C ring structure are of similar magnitude as those between the $\Delta^{8,14}$ structures and skeletons that have so far not yielded active compounds, *i.e.* the Δ^{5}, $\Delta^{5,7}$ and fully saturated fragments. This finding indicates that, although the positions of the double bonds in skeletons are important, the skeleton conformation induced by the double bond system is not the discriminating factor for activity of double bond FF-MAS isomers.

Table 2. Side chain torsion angles of the independent and disordered molecules in the crystal structures of **2**-**6**. *T1* corresponds to C13-C17-C20-X22, *T2* to C17-C20-X22-C23, *T3* to C20-X22-C23-C24 and *T4* to X22-C23-C24-C25. Primes indicate a second independent molecule if present, an asterisk indicates the minor component of a disordered side chain.

	T1	*T2*	*T3*	*T4*
2	179.2	59.0	168.7	-80.7
2'	172.7	63.7	-179.9	-91.4
3	179.4	-177.7	-100.9	-
3'	174.3	-174.0	-100.3	-
4	-32.7	-79.4	169.7	147.2
4*	-72.2	163.0	-154.1	-132.3
5	175.4	-161.7	156.3	127.5
5'	179.8	-159.5	168.3	173.0
6'	178.5	-179.5	179.1	171.0
6	-178.8	-174.7	179.2	169.2
6*	-178.8	-174.7	179.2	-160.1

Although the skeleton structure and conformation of **5** satisfies the conditions required for activity discussed above, the compound is not active. Its lack of activity is therefore probably related to the chemical structure of the side chain. In order to compare the side chain conformations of the molecules in all five crystal structures, the skeletons of all independent molecules were superposed. A stereo view of the superposition of the skeleton atoms of crystal structures **2-5** is given in Figure 3. Rings A, B and C were

included in the fit but are omitted from the Figure for clarity. Table 2 gives the successive torsion angles in the side chains of the five structures for each of the independent molecules and for all disorder components. The orientations of the side chains correspond to those found for other sterols with both an R and S C20 stereochemistry (Nes *et al.*, 1984). Compounds **2**, **3** and **5**, with natural configuration at C20 (see above), crystallize with the side chain in an extended conformation pointing away from ring D (the so-called 'north-east' orientation). The orientation of the side chain in the structure of **4**, which has unnatural C20 stereochemistry, has been called the 'north-west' orientation and points upwards in the plane of the paper of Figure 1. It can be seen from Figure 3 that the phenyl rings of the side chains of compound **2** occupy the same region in space as that of the inactive compound **5**. The side chains of the active compounds **2** and **3** occupy different regions in space. The finding that the side chains of three active compounds **2**-**4** are pointing in different directions and that the side chain of the inactive compound **5** overlaps with that of the active compound **2**, indicates that in this case the active conformation(s) cannot be distilled from the crystal structure conformations. Either the active conformation of the side chain of compound **2** is a conformation present in solution but not in the crystal structure, or the side chain of compound **5** is mainly present in another, inactive, conformation in solution.

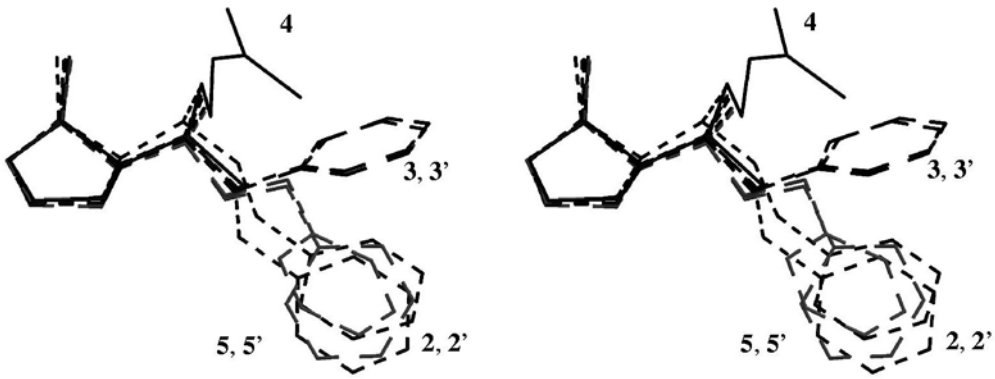


Figure 3. Stereo picture of the side chains in the crystal structures of compounds **2**-**5** after superpositioning of the skeletons. Compound **5** is shown in light gray, compounds **2**-**4** are shown in dark gray. The primes indicate the independent molecules in the structures.

An interesting aspect of the aromatic C ring structure mentioned above is the orientation of the side chain with respect to the skeleton. The C17-C20 bond of the aromatic C ring structure points upwards with respect to the plane through the skeleton rings (Figure 4). This is also found for calcitriol bound to its receptor (Rochel *et al.*, 2000). By rotation of the C17-C20 torsion angle, the

conformation of the side chains of compounds **1**-**6** can change such that the direction is similar to the side chain of the aromatic C ring structure.

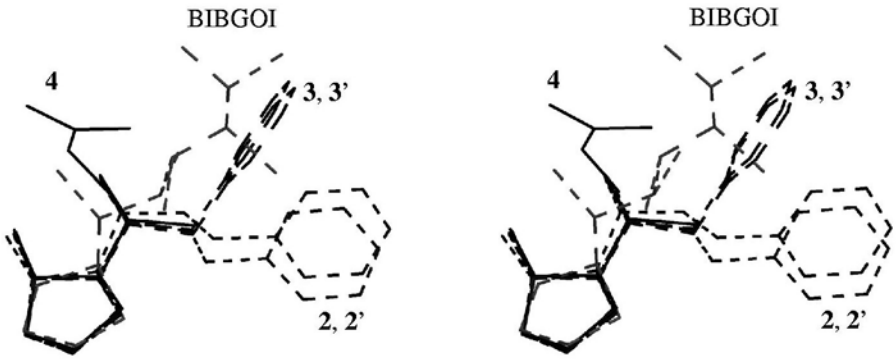


Figure 4. Stereo picture of the side chains of in the crystal structure of compounds **2**-**4** (dark gray) and the side chain of the aromatic C ring structure (22α,23α-dibromo-18-nor-12-methyl-5α,17β-ergosta-8,11,13(14)-triene-3β-t-butyldimethylsilyl ether, Ferguson *et al.* 1982, CSD reference code: BIBGOI), shown in light gray.

Solution conformations

Table 3 gives the chemical shifts of the protons in the ^{1}H NMR spectra of compounds **3** to **5**. Figures 5A-5C give the potential energy of each of the generated conformers with respect to the R_r-values. The R_r-values of the crystal structure conformations, for which no energies were calculated, are given as well. The independent molecules in the crystal structure of **3** and those in the structure of **5** have similar conformations, respectively (see Figure 3), which is reflected in the similarity in R_r-values. The conformations found in the crystal structures of **3** and **5** give the lowest R_r-values. For both compounds, two of the model conformers give substantially lower R_r-values compared to the other modeled conformations. For each of these pairs of conformations, one corresponds to that found in the crystal structure. The second conformer in each pair differs in the torsion angle at the end of the side chain and therefore differs only in the orientation of the phenyl ring. The number of experimentally observed cross peaks that involve the phenyl protons is small. As a consequence, the two orientations of the rings of compounds **3** and **5** cannot be distinguished. For both compounds, the modeled conformations that resemble those found in the crystal structures have lowest R_r-values. The most populated component of the disordered molecule in the crystal structure **4** is very similar to the model conformer that gives the lowest R_r-value. The less populated crystal structure component gives a higher R_r-value. Probably, the crystal packing imposes a conformational change in the side chain of the minor component upon

crystallization. It appears that the crystal structure conformations of **3** and **5** and the most populated component of **4** are also present in solution.

Table 3. Chemical shifts of ^{1}H signals of compounds **3-5** in ppm, relative to an internal standard, $Si(CH_3)_4$.*

Atom	**3**	**4**	**5**	Atom	**3**	**4**	**5**
H1α	1.85	1.37	1.84	H15	5.41	5.33	5.40
H1β	1.33	1.25	1.32	H16α	2.54	2.03	2.58
H2α	1.73	1.76	1.72	H16β	2.23	2.20	2.30
H2β	1.63	1.65	1.60	H17	1.60	1.89	1.88
H3α	3.25	3.27	3.24	Me-18	0.84	0.87	0.79
Me-4α	1.02	1.04	1.02	Me-19	1.04	1.06	1.03
Me-4β	0.82	0.86	0.83	H20	1.91	3.57	3.60
H5α	1.23	1.25	1.22	Me-21	0.84	1.17	1.27
H6α	1.78	1.80	1.77	H22R	2.93		
H6β	1.57	1.58	1.56	H22S	2.11		
H7α	2.33	2.34	2.33	H23R		3.85	4.64
H7β	2.21	2.24	2.20	H23S		4.15	4.41
H11α	2.17	2.21^{+}	2.21^{+}	H24		5.40	
H11β	2.17	2.18^{+}	2.14^{+}	Me-26		1.77	
H12α	1.40	1.43	1.40	Me-27		1.70	
H12β	2.02	2.18^{+}	1.94	Phenyl	7.27		7.35

* Chemical shifts given to two decimals are generally accurate to ±0.01 ppm, values that are marked by $^{+}$ are accurate to about ±0.02 ppm.

The conformations derived from the NOE cross peaks are corroborated by the ^{3}J coupling constants determined from the 1D NMR spectra. The coupling constants of side chain protons attached to consecutively bonded carbon atoms are given in Table 4, as well as the expected torsion angles values derived from the coupling constants using the Karplus relation (Friebolin, 1993) and the torsion angle values of the conformer with lowest R_r-value. For ^{3}J-coupling constants larger than 6 Hz, the model torsion angles are larger than those expected. Apparently, the molecules in solution do not adopt this conformation during the full duration of measurement. A second conformation may exist in which the molecules spend a sufficient amount of time for detection using through-space chemical exchange as measured in the ROESY and NOESY NMR spectra.

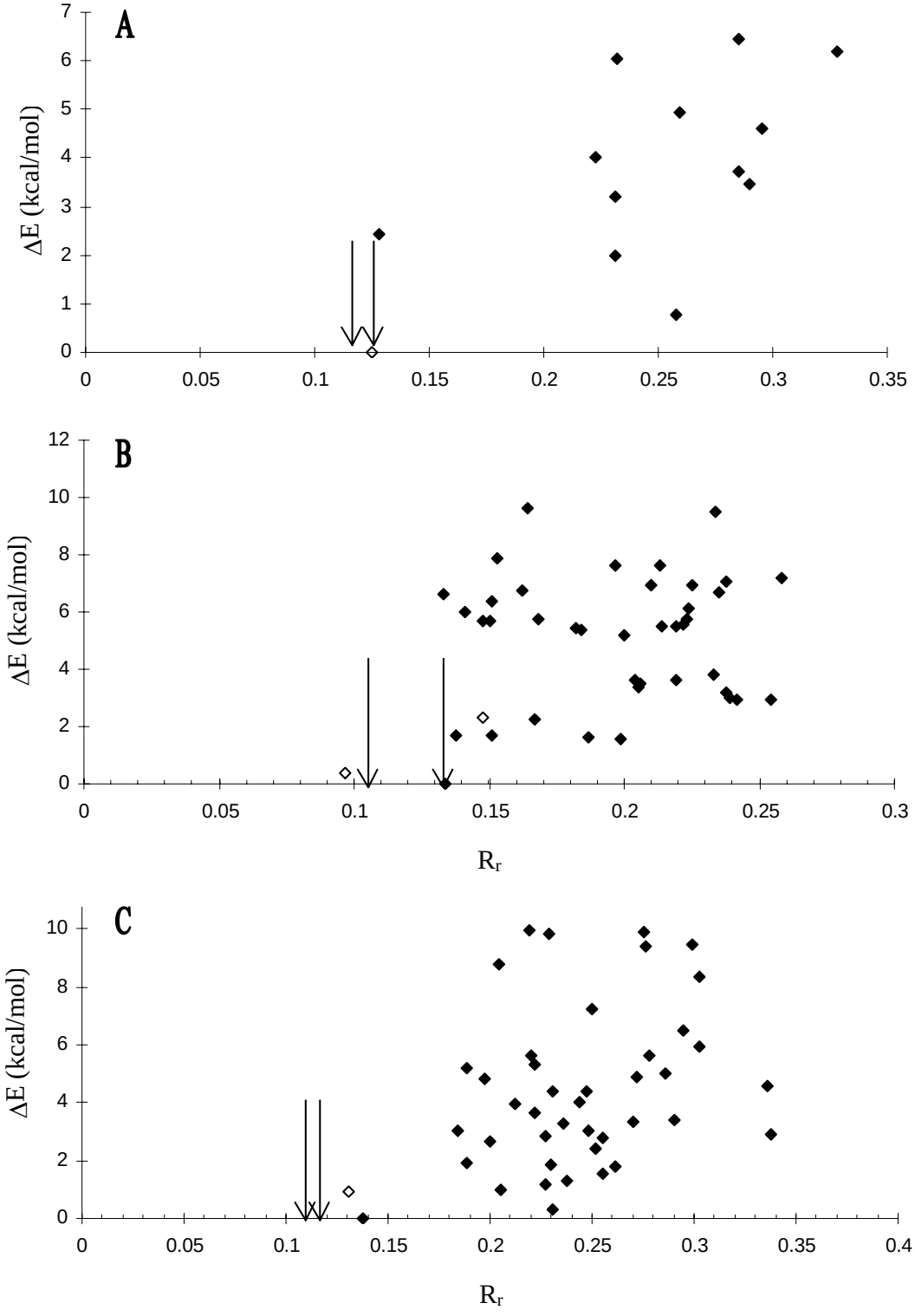


Figure 5A-5C. Scatter plots of the relative potential energies of the single conformers as function of the R_r-values for compounds **3-5**, respectively. The diamonds represent modeled conformers. Open diamonds represent conformers with similar conformation as the crystal structures. The arrows indicate the R_r-values of the crystal structure conformations.

Table 4. Selected ^3J coupling constants in Hz of protons bound to consecutive carbon atoms in the side chains of compounds **3-5**. Torsions estimated using the generalized Karplus relation (T_{lit}, Friebolin, 1993) are given as well as the corresponding torsion angles of the modeled conformer with best R_r-value (T_{exp}).

	3			**4**			**5**		
	^3J	T_{lit}	T_{exp}	^3J	T_{lit}	T_{exp}	^3J	T_{lit}	T_{exp}
H17:H20	nd*	-	-178°	8.6	130°-160°	-179°	9.3	135°-165°	-176°
H20:H22R	nd*	-	178°	-	-	-	-	-	-
H20:H22S	3.1	45°-75°	63°	-	-	-	-	-	-
H23R:H24	-	-	-	5.5	30°-60°	-43°	-	-	-
H23S:H24	-	-	-	6.1	120°-145°	-159°	-	-	-

* Coupling constants were not determined.

The R_r-value as a function of the number of mixed conformers for compounds **3-5** is given in Figure 6. In the two-conformer combinations for all three compounds, the single conformer with the lowest R_r-value was present with the highest fraction. For compounds **4** and **5**, the drop in R_r-value after adding a second conformer to the single conformer is small and expected to fall within the error of the intensities of the cross peaks. For compound **3**, the drop in R_r going from 1 to 2 conformers is more pronounced. The drop in R_r-value is caused by addition (fraction 0.2) of a conformer with torsion angles of -60° for torsion C13-C17-C20-C22, 180° for torsion C17-C20-C22-C23 and 180° for torsion C20-C22-C23-C24, with a ΔE of 3.21 kcal/mol compared to the conformer with lowest potential energy. This conformation corresponds to the conformation of the active compound **4**.

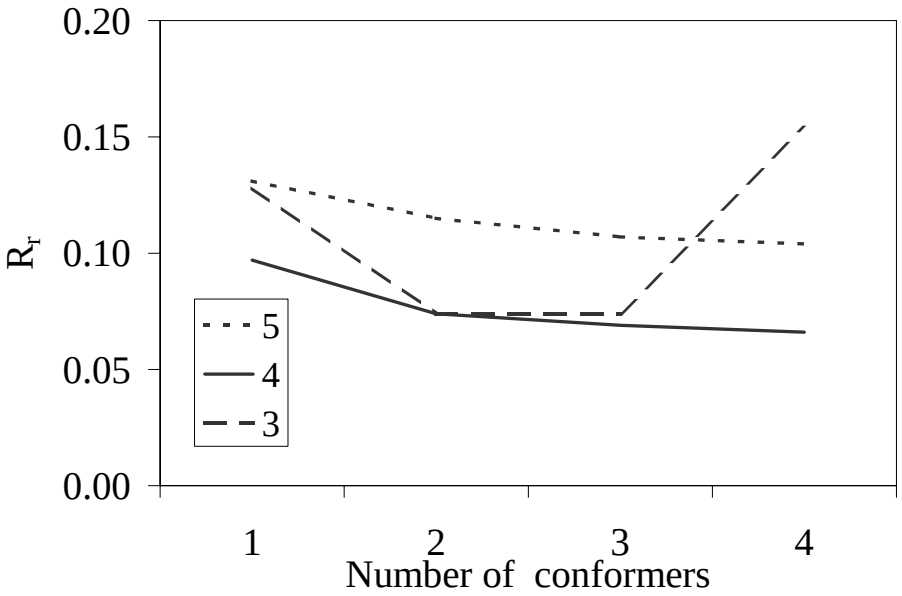


Figure 6. R_r-values given as a function of the number of conformer combinations of compounds **3-5**.

It appears that *active* compounds **3** and **4** are both present in a similar conformation in solution that is not found for the inactive compound **5**. This conformation might therefore be the active conformation. However, it is possible that interaction of compounds **3**-**5** with their endogenous receptor stabilizes other conformations than those present in solid and solution phases, as is the case in the crystal structure of the complex of the VDR and its natural ligand, calcitriol (see also chapter 6). This conformation is also exhibited by the aromatic C ring structure and is a candidate for the active conformation as well. It is also possible that the putative receptor of MAS can accommodate both conformations. In any case, it seems likely that the active conformation is not the extended conformation of compounds with natural configuration at C20 as found in known crystal structures of this class of compounds.

Conclusions

The skeleton conformations in crystal structures of double bond isomers that can yield active compounds are very similar to those that do not yield active compounds. Therefore, the conformation of the skeleton induced by the double bond system does not seem to determine activity. The side chain conformations in the crystal structures are not discriminative with regard to activity. It was found that the crystal structure conformations of sterols **3**-**5** are most abundant in solution. Probably, the active compound **3** exists in solution in a second conformation that is very similar to that of the active compound **4**. Therefore, the conformation of compound **4** is a candidate active conformation for the compounds described here, as is the conformation exhibited by the endogenous ligand of the vitamin D receptor in the bound state, which resembles the conformation of the aromatic C ring structure. It seems likely that the extended, north east conformation observed in the crystal structures of compounds **2**-**3** and **5**-**6** is not the active conformation.

Acknowledgement

Paul Erbel and Bas Leeflang are gratefully acknowledged for their help in the measurement and the interpretation of the NMR spectra. Alexandre Bonvin is acknowledged for the useful discussions in deriving the solution conformations from the spectra.

Chapter

4

Relation between the Molecular Electrostatic Potential and Activity of some FF-MAS Related Sterols

D.R. Boer, H. Kooijman, J. van der Louw[1], M. Groen[1], J. Kelder and J. Kroon

D.R. Boer, H. Kooijman, J. van der Louw, M. Groen, J. Kelder & J. Kroon. "Relation between the molecular electrostatic potential and activity of some FF-MAS related sterols", *Bioorg. & Med. Chem.* In the press.

[1] Department of Medicinal Chemistry, N.V. Organon, Oss, The Netherlands

Abstract

Follicular Fluid-Meiosis Activating Sterol (FF-MAS) is a compound important for maturation of gametes in mammals. Therefore, it may serve as a lead compound for a novel method of contraception. We studied the Molecular Electrostatic Potential of a series of active and inactive analogs of FF-MAS. It is found that double bond configurations required for activity result in a local negative electrostatic potential which is larger as well as denser compared to those of inactive molecules. We therefore hypothesize that the interaction energy of the double bond system of the MAS compounds with its receptor substantially contributes to the overall interaction energy. This notion is supported by interaction studies of the electrostatic potential originating from the double bonds in crystal structures of cholesterol and four MAS-derived $\Delta^{8,14}$ structures synthesized and crystallized by us. In addition, we were able to derive a pharmacophore model that relates the local average ESP and its distance to the 3β-OH oxygen atom to the activity of the molecules.

Introduction

Recently, two 4,4-dimethyl-$\Delta^{8,24}$-sterols were found to have a regulatory function in meiosis (Byskov *et al.*, 1995) and were called follicular fluid-meiosis activating sterol (FF-MAS) and testicular-MAS (T-MAS), respectively (compounds **1** and **2** in Figure 1). Because of their involvement in the production of reproductive cells, their putative receptors are promising targets for the development of more tissue specific contraceptives. As of yet, the molecular target(s) of the MAS compounds have not been identified, although some evidence exists that the putative FF-MAS receptor is a membrane protein, possibly G-protein coupled (Grøndahl *et al.*, 2000). Because of lack of information on the MAS-receptors, any drug design work on this subject has to be based on activity data of FF-MAS or T-MAS derived compounds.

The derivation of structure-activity relationships of FF-MAS related sterol compounds has been the focus of our research. Preliminary structure-activity studies showed that the 3β-OH group, the type of side chain and the double bond configuration within the skeleton determine activity (see also chapters 1 and 3). In addition, two methyl groups attached to C4 enhance activity, but are not mandatory. The aim of the work described in this chapter is to link activity with the double bond configuration within the skeleton of the sterols. For this purpose, a series of compounds is required in which the positions of the double bonds are changed, at the same time allowing only minor changes in the rest of structure. Activity data on such a series is presented in this chapter. The compounds are derivatives of cholesterol, 4,4-dimethyl-cholesterol and desmosterol and contain one or two double bonds in the skeleton (compounds **3**-**11** in Figure 2).

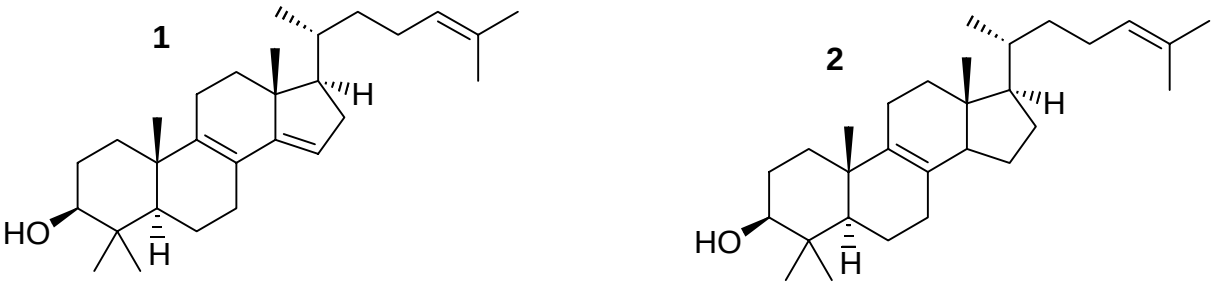


Figure 1. Molecular structures of FF-MAS (**1**) and T-MAS (**2**).

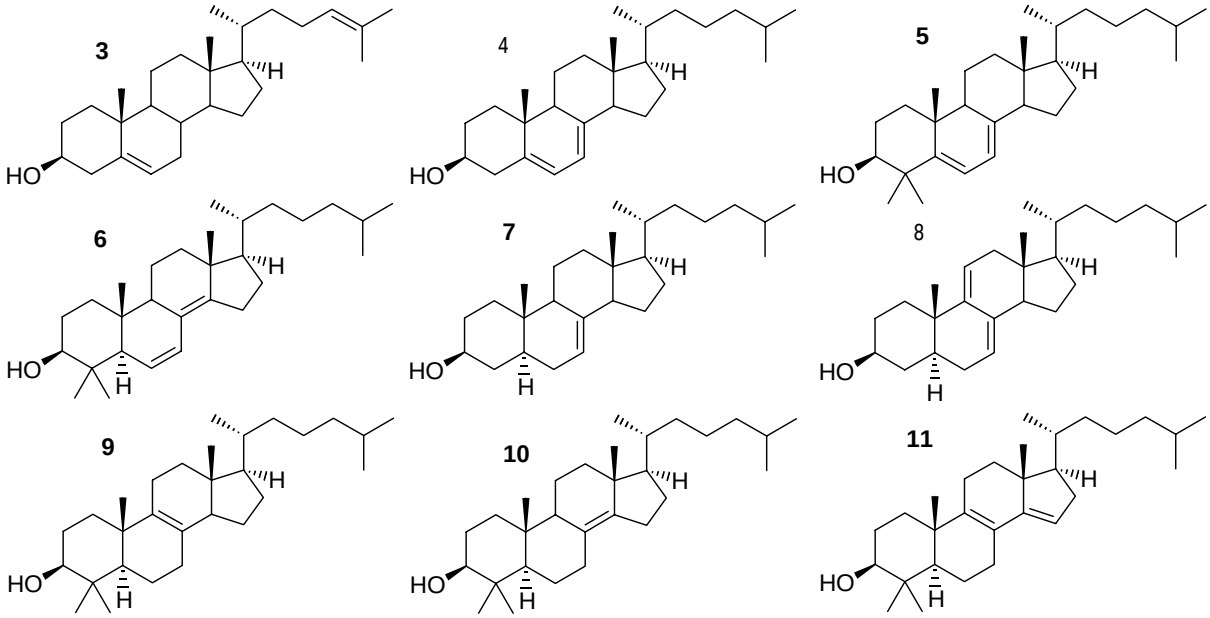


Figure 2. Molecular Structures of double bond sterol isomers **3-11**.

The observed diversity of the skeletal structure of known active sterol compounds is relatively small compared to that of the side chain. All active structures presented in this chapter are $\Delta^{8,14}$ sterols. The skeletal conformations of these compounds in crystal structures are very similar to those of sterol compounds with double bond configurations that are known to impair activity (see chapter 3). Therefore, differences in activity of double bond sterol isomers are most probably caused by differences in the molecular electrostatic potential in vicinity of the double bonds. Subsequently, we compared semi-empirically derived molecular electrostatic potentials (MEP's) caused by the double bond systems in order to derive a concept that distinguishes between active and inactive double bond configurations.

To validate this concept, the calculated MEP's are compared to experimental data in two ways. Firstly, it is known that sterols containing double bonds in the skeleton are more susceptible to protonation on the α-side of the molecule than on the β-side: 5α isomer products are found in much higher yields than the 5β isomers as described in chapter 2 (Dolle *et al.*, 1988; Dolle & Kruse, 1986; Wilson & Schroepfer, 1988; Boer *et al.*, 2000). This is probably the result of an electrostatic effect, causing a larger attraction on the proton at the α-side than at the β-side of the sterol isomers. The MEP's should be consistent with these findings. Secondly, crystal structure contacts made in the vicinity of the double bond(s) are examined to answer the question whether differences in activity can be explained by differences in the electrostatic potential (ESP) dependent interaction behavior of the molecules. We determined crystal structures of five $\Delta^{8,14}$ sterols (compounds **13-17** in Figure 3), as described in chapter 3. Two

crystal structures of cholesterol (**12**), a Δ^5 isomer, were retrieved from the CSD (reference codes CHOLES20 and CHOLEU01). To explore intermolecular interactions, a MEP was calculated for each independent molecule (reference molecules) in the crystal structures of the Δ^5 and $\Delta^{8,14}$ isomers. For each reference molecule, a contact molecule was identified that interacts with the part of the surface of the reference molecule in the vicinity of the double bond system. The part of the surface of the contact molecule (contact surface) close to the MEP surface in vicinity of the double bonds of the reference molecule was determined. The sign and size of the ESP of the relevant part of the MEP surface of both reference and contact molecule is then evaluated and compared for the Δ^5 and $\Delta^{8,14}$ isomers.

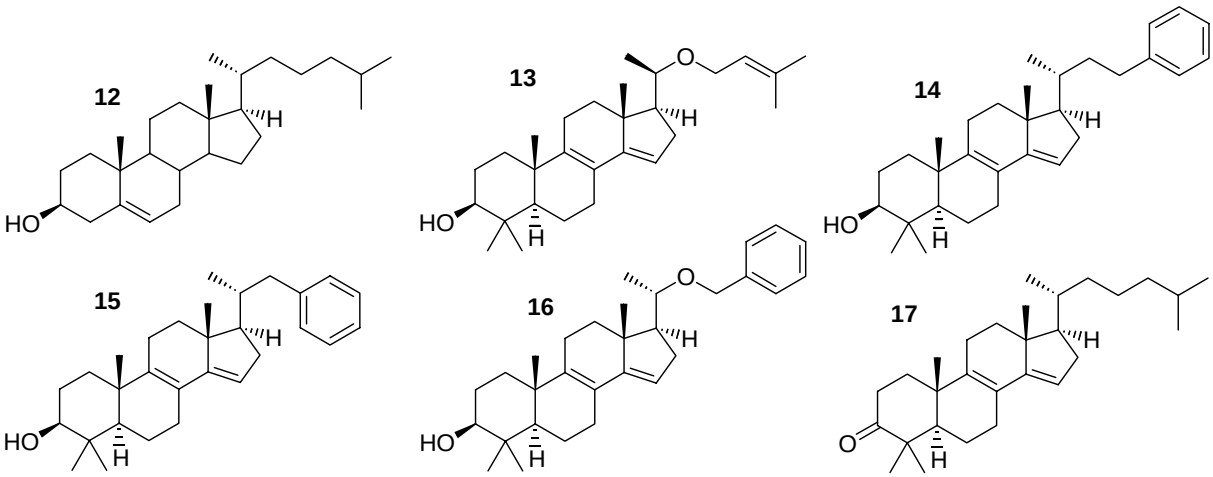


Figure 3. Molecular structures of compounds **12-17**.

Methodology

ESP calculations

ESP charges were calculated *in vacuo* using the semi-empirical method implemented in MOPAC v6.00 (Stewart, 1990). For compounds **1** and **3-11**, ESP charges were calculated using minimized conformations with the side chain extended. For compounds **12-17**, ESP charges were calculated for all independent molecules in the crystal structures after minimization, keeping all torsion angles fixed. Since the ESP charges were calculated on isolated molecules *in vacuo*, charge-transfer and induction effects are neglected. DelPhi (Honig & Nicholls, 1995), as implemented in GRASP (Nicholls & Honig, 1991), was used to calculate the MEP at the solvent-accessible grid surface using the MOPAC ESP charges. The inner and outer dielectric constants were set to unity. To study interactions in crystal structures, the ESP was calculated at a Connolly surface nearer to the molecule, at a distance of approximately 1.4 Å. ESP values are expressed as energies in kJ/mol.

The ESP's near the double bonds were negative and approximately circular and flat. Therefore, they will be referred to as a 'patch'. The most negative grid point of a patch was defined as its center point. d_{OH} is defined as the distance of this center point to the 3β-OH oxygen atom. To determine the size of the patch, the fraction of negative ESP points with respect to the total number of grid points in (planar) circular shells around the center point was calculated. This fraction of negative ESP points was plotted against the outer radius of each shell. The grid points in the shells closest to the center point were all negative in all cases, that is, the fraction of negative ESP points in those shells was unity. At a certain shell radius, the fraction dropped to a value of approximately 0.3. The radius of a patch, R_p, was defined as the distance at which 50% of the grid points were negative. The number of grid points on the molecular surface within R_p distance of the patch center point was counted. The total ESP of all points in a patch divided by the number of points gives an averaged ESP, $<ESP>_p$, which is indicative of the electrostatic potential near the double bond system of the molecules. The area of the patch, A_p, was calculated by multiplying the number of points in a patch with the area of the total molecular surface divided by the total number of points thereof. Superscripts are used to indicate on which side of a molecule the patch occurs.

In the crystal structures of molecules **12**-**14** and **15**-**17**, the α-side patches (reference patch) originating from the double bond system made contact with other molecules in the crystal structure. Using the patch size, A_p, determined as described above, a contact patch on the surface of the contact molecule was constructed. The contact patch is defined as all grid points of the contact molecular surface within 0.7 Å of the surface points of the reference patch. An average ESP on this contact patch, $<ESP>_{cp}$, was determined for all independent molecules in each of the crystal structures.

Activity determinations

Induction of maturation is histologically visualized by disappearance of the oocyte nuclear envelope or germinal vesicle (GV) after MAS treatment (Downs *et al.*, 1985; Guoliang *et al.*, 1993). Oocytes from immature, humegon® treated mice are used to study this germinal vesicle breakdown (GVBD). The oocytes were harvested from the antral follicles in the ovaries and freed from cumulus cells. The oocytes are cultured, in the presence of the studied compounds, in hypoxanthine containing medium to exclude spontaneous GVBD. At the end of a 22 h culture period the percentage of GVBD, which is a measure of MAS activity, is calculated. The activities of compounds **1**, **3**-**4** and **6**-**17** are determined in this way (see Table 1). A compound is considered to be active when it induces a 100% GVBD at 10 μM or lower concentrations. Compound **5** was inactive (N.V. Organon, personal communication).

Table 1. Activities for compounds **1**, **3**-**4** and **6**-**17** expressed as the percentage of germinal vesicle breakdown (GVBD) with respect to the concentration in the culture medium (c).

Compound	GVBD (%, ± s.d.)	c (μM)
1: FF-MAS	100	5
3: Desmosterol	14 (5)	10
4: 4,4-H-$\Delta^{5,7}$	18 (5)	10
6: $\Delta^{6,8(14)}$	68 (3)	10
7: 4,4-H-Δ^{7}	27 (8)	10
8: 4,4-H-$\Delta^{7,9(11)}$	35 (24)	10
9: Δ^{8}	24 (9)	10
10: $\Delta^{8(14)}$	15 (3)	10
11: $\Delta^{8,14}$	84 (9)	10
12: cholesterol	35 (10)	10
13: ORG 38799	100	10
14: ORG 39823	100	10
15: ORG 39097	100	10
16: ORG 38580	1.5 (2)	10
17: ORG 38899	16 (7)	10

Results & Discussion

As expected, calculations show that two negative ESP regions (patches) occurred on the α- and β-side of the solvent accessible surfaces in vicinity of the skeletal double bond(s), originating from the π-electron density. The size of the α- and β-side patches and their precise location with respect to the skeleton atoms depend on the number of double bonds and their position in the structure. For compounds **1**, **3**-**11** and **13**-**16** the area, A_p, and average ESP, $<ESP>_p$, of the patches on both the α- and β-sides are given in Table 2, as well as the distance of their centers to the 3β-OH group, d_{OH}. As expected, d_{OH}^{α} and d_{OH}^{β} increase with migration of the double bond system towards the D-ring. In all compounds except one (compound **3**), A_p^{β} was considerably smaller than A_p^{α}. Also, $< ESP >_p^{\beta}$ is in general equal to or smaller than $< ESP >_p^{\alpha}$, except for compounds **3** and **5**. For compound **3**, this can be explained by the shielding effect caused by the C19 and the C4β methyl groups projecting towards the β-side. The presence of the methyl groups causes the molecular surface to occur at a larger distance from the double bonds and leads to charge relocalization over the methyl groups. Both effects lead to a decrease in $< ESP >_p^{\beta}$ with respect to $< ESP >_p^{\alpha}$. In the structure of compound **3**, the Δ^5 double bond is less shielded on the β-side of the molecule because the C4 methyl groups are not present. For compound **5**, $< ESP >_p^{\beta}$ is more negative, but A_p^{β} is smaller than A_p^{α}. Therefore,

the integrated ESP will be smaller on the β-side of the molecule than on the α-side.

Table 2. Area and average ESP (in kJ/mol) of the α- and β-side ESP patches of sterol isomers **1**, **3-11** and **13-16**. The distances between the patch centers and the 3β-OH oxygen atom is also given.

Molecule	A_p^{α} (Å^2)	$< ESP >_p^{\alpha}$	d_{OH}^{α} (Å)	A_p^{β} (Å^2)	$< ESP >_p^{\beta}$	d_{OH}^{β} (Å)
1: FF-MAS	64	-7.5	7.4	24	-5.5	9.1
3:4,4-H-$\Delta^{5,24}$	5.2	-3.5	5.14	36	-8.2	5.0
4:4,4-H-$\Delta^{5,7}$	39	-7.7	5.6	23	-7.2	5.8
5:$\Delta^{5,7}$	54	-6.9	5.3	16	-7.2	6.9
6:$\Delta^{6,8(14)}$	43	-8.9	8.0	23	-6.9	7.2
7:4,4-H-Δ^{7}	21	-5.2	8.1	2.4	-0.7	6.8
8:4,4-H-$\Delta^{7,9(11)}$	48	-6.4	8.1	9.8	-3.7	6.2
9: Δ^{8}	17	-3.7	5.2	0.93	-0.5	8.2
10: $\Delta^{8(14)}$	29	-5.0	8.2	1.7	-0.2	8.9
11: $\Delta^{8,14}$	49	-9.4	8.7	25	-5.5	8.3
13:ORG 38799	29	-6.1	8.1	7	-2.6	8.6
14:ORG 39823	42	-8.1	7.3	28	-5.5	9.1
	40	-8.1	7.4	20	-4.5	9.1
15:ORG 39097	38	-10.0	8.1	25	-4.9	8.5
	38	-9.3	7.8	27	-5.9	8.8
16:ORG 38580	47	-10.7	7.6	17	-5.4	8.6
	74	-16.6	7.1	14	-3.0	8.8

In Figure 4, the ESP patches at the solvent accessible surface on the α- and β-sides of the $\Delta^{5,7}$ isomer **5** are shown. The small spheres represent the grid points within the patch. The ESP patch on the α-side of the molecule is about three times as large as that on the β-side (54 Å^2 and 16 Å^2 respectively), whereas the average ESP is similar for both sides. This finding is consistent with synthetic results found for isomerisation reactions starting from $\Delta^{5,7}$ isomers: 5α isomer products are found in higher yields than the 5β isomers (see chapter 2 and Dolle *et al.*, 1988), because the α-side is more susceptible to protonic attack than the β-side. Thus, it appears that the calculated ESP's are transferable to molecules in solution.

The size and average ESP of the α-side patches of the molecules depend roughly on the number of double bonds in the skeleton. In general, structures containing one double bond gave smaller patches compared to those with two double bonds (see Table 2). *E.g.*, A_p^{α} of a Δ^{8} structure (compound **9**) is approximately half the size compared to that of the $\Delta^{8,14}$ structure (compound

11). In addition, $< ESP >_p^{\alpha}$ is more negative for the latter compound. We expect that this property will influence the contacts that the patch makes with other molecules. Patches of compounds containing two double bonds will preferably interact with ESP contact regions that are more positive compared to contact regions for structures containing one double bond. Table 3 gives the $< ESP >_p^{\alpha}$ and the A_p of the reference patch and the $< ESP >_{cp}$ of the contact patch for all independent molecules in the crystal structures of compounds **12-14** and **16-17**. The double bond system in the structure of compound **15** interacts with a disordered solvent region that could not be interpreted and is therefore excluded from Table 3.

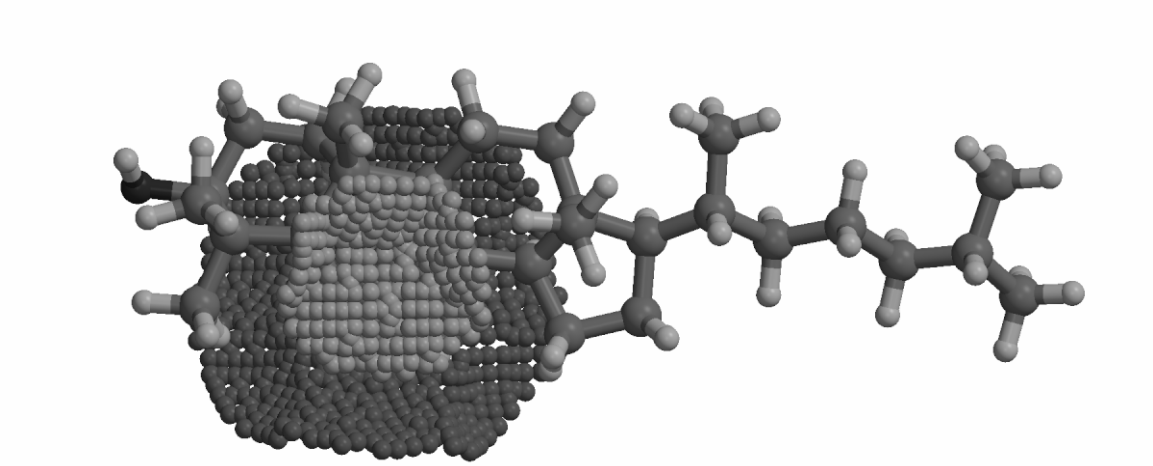


Figure 4. ESP patches on the solvent accessible surface at the α- and β-sides of compound **5**. The view is on the β-side of the molecule, showing the α-side patch in dark gray and the β-side patch in light gray.

The ESP reference patches of the molecules in the crystal structures of compounds **13-14** and **15-17** are larger than those in the crystal structure of **12** and make contacts with positive or neutral ESP regions on the molecular surfaces of other molecules. For the two cholesterol (compound **12**) crystal structures from the CSD, which all have smaller reference patches, a negative ESP in the contact patch was found in five out of sixteen interactions. Figure 5 shows the crystal packing of the cholesterol molecules in both structures, as well as the α-side ESP patches of the molecules. Some of the patches overlap considerably.

It is reasonable to assume that close vicinity of the Δ^5 reference patches and negative contact patches results in a slightly repulsive interaction. However, it is conceivable that the overall interaction energy of the cholesterol molecules in the crystal structures could compensate for this repulsion because the patch sizes and the resulting repulsive energy are relatively small. In contrast, interactions between the larger ESP reference patches originating from the $\Delta^{8,14}$ double bonds with contact patches with negative ESP probably result in more

unfavorable interaction energies and therefore a larger repulsion. The latter unfavorable interactions cannot as easily be compensated for by interactions occurring elsewhere in the crystal structure, and in fact they do not occur. In summary, the ESP patches of $\Delta^{8,14}$ compounds, which in general are larger and have a more negative $< ESP >_p^{\alpha}$ than those of cholesterol, do not interact with negative contact ESP regions, in contrast to what is observed in the crystal structures of cholesterol. In addition, the ESP patches of $\Delta^{8,14}$ compounds occur near rings C and D, whereas the patches of the Δ^5 compounds occur in the vicinity of ring B. The fact that the $< ESP >_p^{\alpha}$ is in general more negative and d_{OH}^{α} is larger for potentially active molecules suggests that these parameters (partly) determine activity.

Table 3. Average ESP of the α-side patches, $< ESP >_p^{\alpha}$ in kJ/mol, of all independent molecules in the crystal structures of compounds **12-14** and **16-17**, their radius A_p in Å^2 and of the corresponding contact patch ESP's, $< ESP >_{cp}$ in kJ/mol, at the contact surface of the interacting molecule.

Molecule	$< ESP >_p^{\alpha}$	A_p	$< ESP >_{cp}$
12: CHOLES20	-11.4	12.2	1.1
	-13.4	18.2	-1.0
	-9.4	12.2	0.9
	-14.9	18.9	1.0
	-11.9	19.4	0.3
	-12.4	15.4	-2.2
	-10.7	16.0	2.1
	-13.9	14.9	-0.2
12: CHOLEU01	-14.1	17.7	0.7
	-7.4	14.5	-11.4
	-13.1	16.4	-5.5
	-12.4	15.9	-13.9
	-8.7	16.0	9.7
	-9.9	9.2	5.5
	-11.4	21.6	4.2
	-9.4	10.7	0.2
13: ORG 38799	-13.6	27.5	5.0
14: ORG 39823	-16.4	32.2	1.0
	-16.6	30.6	2.5
16: ORG 38580	-21.6	40.0	6.2
	-18.8	43.0	1.5
17: ORG 38899	-18.6	24.0	21.8
	-19.1	23.7	12.6

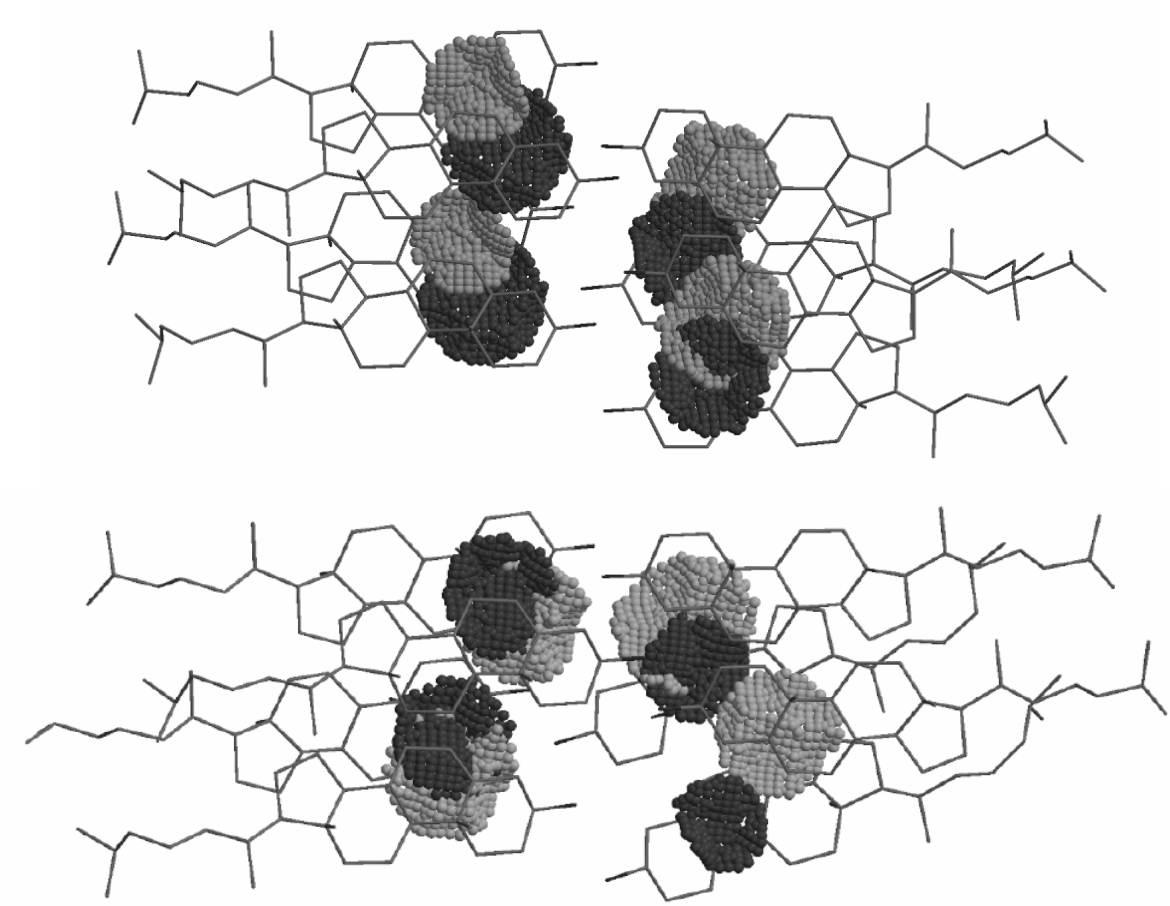


Figure 5. Interactions of the ESP patches of cholesterol molecules (compound **12** in Figure 3) in two crystal structures present in the CSD. The top Figure is from CHOLES20, the lower from CHOLEU01. In six of the sixteen possible interactions, the spheres overlap considerably, as do the patches.

In Figure 6, $< ESP >_p^{\alpha}$ is plotted against d_{OH}^{α} for compounds **1**, **3-11** and **13-16**. Compound **17** was excluded from the plot, because the lack of activity of this compound is related to alterations in the 3β-OH group that was found to be mandatory for activity. In the plot, clusters emerge which contain inactive and active compounds only. As can be seen, the active compounds (represented by a square) are clustered together and have an average ESP of about –9 kJ/mol at approximately 8 Å from the hydroxyl oxygen. The outcome for compound **6**, which is partially active, is close to those of the active compounds. One of the independent molecules in the crystal structure of **16** has an ESP patch with similar d_{OH}^{α} and $< ESP >_p^{\alpha}$ compared to that of active compounds, but is not active. This lack of activity is most probably due to the chemical structure of its side chain: other 22-oxa analogs with a natural configuration of C20 were shown to be inactive as well (see chapter 1). For the other independent molecules in the crystal structure of the inactive compound **16**, $< ESP >_p^{\alpha}$ was about twice as large as those of the active compounds. This is due to overlap between the ESP

around the side chain oxygen atom and that near the double bonds. As a result, the ESP in the vicinity of the double bonds becomes more negative.

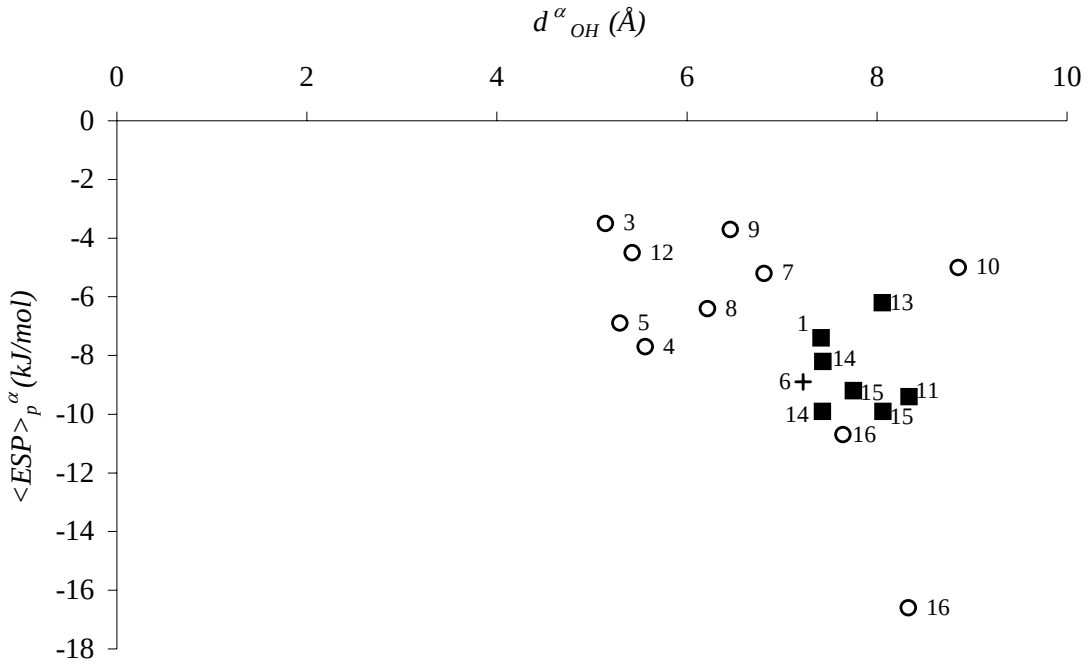


Figure 6. Plot of the average ESP at the α-side patch of the double bonds against the hydroxyl oxygen to patch center distance of compounds **1**, **3-11** and **13-16**. Active compounds are represented by squares, inactive ones by circles and the partly active compound **6** by a cross. Compound numbers are placed close to corresponding points.

From Figure 6, it appears that a correlation between activity and patch position and the magnitude of the ESP exists. This finding, together with the interaction studies in the crystal structures of compounds **12-14** and **16-17**, suggests that a specific interaction occurs between the putative receptor of FF-MAS related compounds and the double bond system in rings C and D. Since the double bonds form a conjugated π-system, an interaction is feasible that is comparable with the edge-face or face-face stacking found for phenyl ring contacts. Such an interaction in fact occurs in the crystal structure of (3β,5α,17β,22E)-ergosta-8,14,22-trien-3β-ol benzoate (Dolle *et al.*, 1988), with CSD reference code GAKFON. This compound also contains a $\Delta^{8,14}$ double bond system, which interacts with the edge of a phenyl ring in the side chain (see Figure 7). It is an example of how the patch may interact with an aromatic amino acid when MAS compounds bind to their receptor. The molecules in the crystal structures of compounds **14-16** do not give similar interactions, although the side-chains contain a phenyl ring for both structures. Crystal packing probably prevents such interactions.

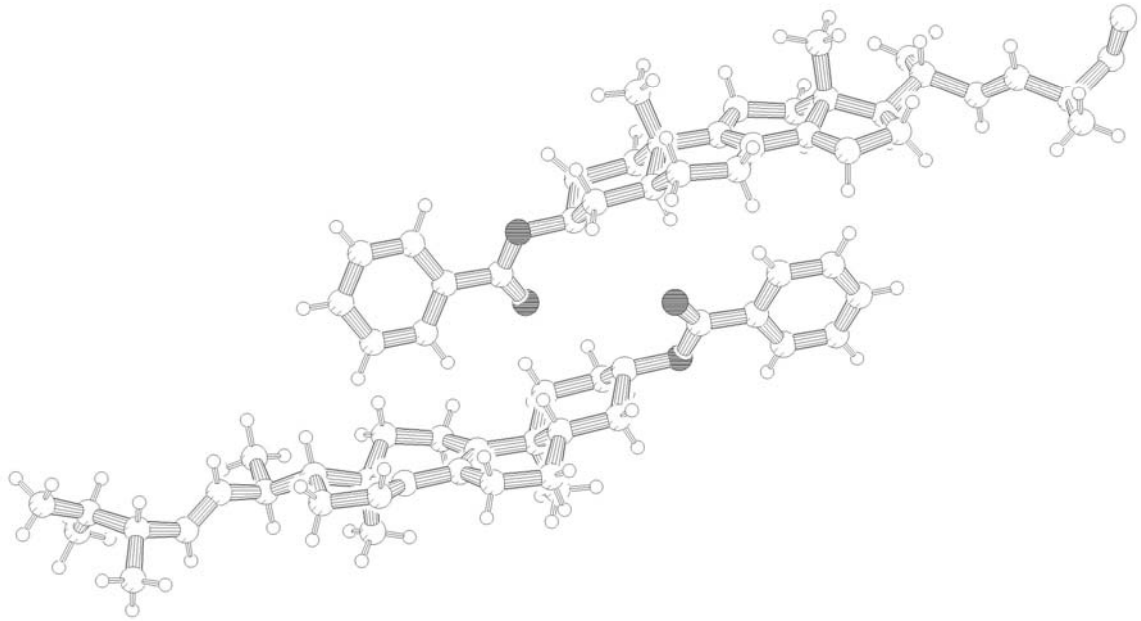


Figure 7. Interaction between two molecules in the crystal structure of ergosta-8,14,22-trien-3β-ol benzoate (Dolle *et al.*, 1988). The edge of the phenyl group interacts with the negative ESP patch of the double bond system.

A pharmacophore model emerges from the results shown in Figure 6. It appears that a hydroxyl group is needed, attached to one end of a rod-shaped molecular scaffold, at approximately 8 Å from a negative ESP region. In addition, activity requires an unpolar side chain attached to the other end of the scaffold, at approximately 11 Å from the OH group and 5 Å from the ESP patch. The three-way interaction topology provided by the negative ESP patch and the two molecular moieties, *i.e.* the OH group and the side chain, offers a pharmacophore model that can be used to evaluate in advance whether a compound is possibly active. Furthermore, molecular databases can be searched for novel lead compounds using the pharmacophore model presented here, provided that the electrostatic field information can be generated for each entry.

Conclusions

The calculations of the MEP's of sterols give results that fit experimental data in two ways. Firstly, calculated ESP patches above and below the double bonds of the double bond system of sterol compound **5** are consistent with respect to its reactivity in acid-catalyzed protonation reactions. The patch at the α-side is about twice as large as the patch at the β-side of the molecule and is the preferred side of proton-attack in isomerisation reactions starting from $\Delta^{5,7}$ isomers. The fact that the calculations *in vacuo* provide an explanation for a

phenomenon in solution is encouraging, since it implies that we can use them for studying interactions in crystal structures. Furthermore, patches in the crystal structures of $\Delta^{8,14}$ compounds **13-14** and **16-17**, with a large size and relatively low negative average ESP, interact with positive or neutral ESP regions on other molecules. In contrast, in crystal structures of cholesterol, we find interactions between the (smaller) α-side patches with contact regions of negative potential. This relation suggests that a specific interaction of the double bond system is related to activity. The observed correlation between the activity on the one hand and the position and average ESP of the α-side patches of the sterol compounds on the other hand supports this assumption. In active compounds, the α-side patch is located at 8 Å from the hydroxyl group, with an average ESP of approximately –9 kJ/mol. The crystal structure of (3β,5α,17β,22E)-ergosta-8,14,22-trien-3β-ol benzoate, with unknown activity, represents a clue as to how an interaction between the double bond system and a protein possibly may occur. The pharmacophore model presented here can help to identify potentially active molecules in further lead development.

Chapter

5

SuperStar: Comparison of CSD and PDB-based Interaction Fields as a Basis for the Prediction of Protein-Ligand Interactions

D. R. Boer, J. Kroon, J.C. Cole[1], B. Smith[1] and M.L. Verdonk[2]

D. R. Boer, J. Kroon, J.C. Cole, B. Smith & M.L. Verdonk. "SuperStar: Comparison of CSD and PDB-based Interaction Fields as a Basis for the Prediction of Protein-Ligand Interactions", *J. Mol. Biol.* In the press.

[1] Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK

[2] Astex Technology Ltd., 250 Cambridge Science Park, Milton Road, Cambridge CB4 0WE, UK

Abstract

SuperStar is an empirical method for identifying interaction sites in proteins, based entirely on the experimental information about non-bonded interactions. The original version of SuperStar was based entirely on scatterplots from the CSD. Here, scatterplots based on protein-ligand interactions are implemented in SuperStar, and validated on a test set of 122 X-ray structures of protein-ligand complexes. In this validation, propensity maps are compared with the experimentally observed positions of ligand atoms of comparable types. Although non-bonded interaction geometries in small molecule structures are similar to those found in protein-ligand complexes, their relative frequencies of occurrence are different. Polar interactions are more common in the first class of structures, while interactions between hydrophobic groups are more common in protein crystals. In general, PDB and CSD-based SuperStar maps appear equally successful in the prediction of protein-ligand interactions. PDB-based maps are more suitable to identify hot-spots in hydrophobic pockets, and inherently take into account the experimental uncertainties of protein atomic positions. If the protonation state of a histidine, aspartate or glutamate protein side chain is known, specific CSD-based maps for that protonation state are preferred over PDB-based maps that represent an ensemble of protonation states.

Introduction

The identification and prediction of favorable protein-ligand interactions is an important step in any drug design project. A number of approaches have been introduced and developed to aid in the design of more promising leads or more potent drugs. Some methods focus on predicting the binding mode of a ligand to a protein (Miller *et al.*, 1994; Jones *et al.*, 1995, 1997; Rarey *et al.*, 1996; Liu *et al.*, 1999). Other methods aim to identify favorable interaction sites for smaller chemical fragments, often referred to as probes, within a binding site (*e.g.* Goodford, 1985; Kellogg *et al.*, 1991; Wireko *et al.*, 1991; Miranker & Karplus, 1991; Caflisch *et al.*, 1993). All these approaches need some sort of evaluation function that ranks interactions between probe or ligand and the protein binding site residues according to how favorable they are. Some use force field based potential energies as a measure of interaction propensity. Alternatively, with the exponential growth of structural databases such as the Cambridge Structural Database (CSD, Allen & Kennard, 1993) and the Protein Data Bank (PDB, Berman *et al.*, 2000), methods were developed that extract the interaction information from these databases and use them to predict favorable protein-ligand binding interactions. Several programs are now available that predict ligand binding or dock ligand groups using a knowledge-based approach (*e.g.* HSITE: Danziger & Dean, 1989a,b; LUDI: Böhm, 1992a,b; PRO_LIGAND: Clark *et al.*, 1995; X-SITE: Laskowski *et al.*, 1996).

Recently, we compiled a knowledge base of non-bonded interactions, based on crystallographic data: IsoStar (Bruno *et al.*, 1997). In this database, non-bonded interactions are presented in the form of scatterplots, which show the distribution of one functional group (the contact group) around another (the central group). These distributions can be transformed into propensity maps, which can then be viewed as contoured surfaces. IsoStar 1.2 (Bruno *et al.*, 1999) contains information on 301 central groups, combined with up to 43 contact groups. The library contains 17,051 scatterplots based on non-bonded contacts from the CSD, and 3,608 based on protein-ligand interactions observed in the PDB.

Recently, we introduced a program called SuperStar (Verdonk *et al.* 1999, 2001), which makes use of the non-bonded contact data contained in the scatterplots in IsoStar to generate composite propensity maps for protein binding sites. SuperStar maps estimate the propensity of a given functional group (a 'probe') to bind at different positions around a template molecule (*e.g.* a binding site). The original version of SuperStar only uses CSD-based scatterplots, which has several important advantages: *e.g.* there is more data in the CSD, it contains hydrogen atoms and the data is more precise. But the use of CSD data also has an important disadvantage. When a ligand binds to a protein binding site, both have to be stripped of the water molecules around them. This causes a driving

force towards forming hydrophobic interactions between protein and ligand. In contrast, in most cases, this effect will be absent during the formation of a small-molecule crystal (*e.g.* because the crystal is formed from an apolar solvent). We did correct for this effect by applying a hydrophobic correction on the CSD-based SuperStar maps for hydrophobic probes. However, there may be more differences between non-bonded interactions in CSD and PDB, which make CSD-based scatterplots less appropriate than PDB-based scatterplots to describe protein ligand interactions.

Here, we describe a systematic comparison of CSD and PDB-based scatterplots from the IsoStar database. The analysis was done for both the geometries of the interactions and for the relative interaction frequencies. Also, an attempt was made to assign protonation states to aspartate and glutamate carboxyl groups and histidine ring nitrogens in protein binding sites, by comparing CSD-based propensity maps with PDB-based ones. We implemented PDB-based propensity maps in SuperStar, and validated them in the same way as we validated the CSD-based SuperStar maps, *i.e.* using a test set of 122 protein-ligand complexes from the PDB. Validation results are compared with those obtained for the original, CSD-based, version of SuperStar.

Methodology

Scatterplots and propensity maps

IsoStar scatterplots

A full description of the methodologies used to build the IsoStar library was given by Bruno *et al.* (1997). The QUEST package was used to search the CSD for contacts between two functional groups, A and B. Most atoms in moieties A and B were defined to be *target* atoms, which are important in the definition of a contact (see below). The remaining atoms are the *backbone* atoms, which are used only to define the chemical connectivities of the functional groups more precisely. A 'contact' in IsoStar is defined as an intermolecular contact between any pair of *target* atoms in groups A and B, shorter than $d_{vdw} + 0.5\text{Å}$, where d_{vdw} is the sum of the Van der Waals radii of the atoms involved. The search results were transformed into an easily visualized form by overlaying the A moieties. This results in a 3D distribution (scatterplot) showing the experimental distribution of B (the contact group) around A (the central group). The positions of the overlaid A moieties are averaged and displayed as a single group in the scatterplots.

All PDB-based scatterplots were generated especially for this study. For 20 central groups (the protein groups, see Table 1), mimicking the most common functional groups that occur in proteins, scatterplots were generated for 6 contact groups (the ligand groups): NH, alcohol OH, water, C=O, aliphatic CH and aromatic CH. For the CSD-based SuperStar calculations presented in this

study, all scatterplots were taken directly from IsoStar 1.2 (Bruno *et al.*, 1999). Some of the CSD-based scatterplots that were used for the CSD vs. PDB comparison, however, were newly generated to allow for a better comparison with the corresponding PDB-based plot.

Converting scatterplots into density maps

From each scatterplot, a density map can be calculated for a selected probe. A probe is an atom in the contact group, *e.g.* methyl carbon. For a scatterplot *j*, each probe *k* contributes to the densities at its surrounding grid points. The probe density *d(i,j)* at each grid point *i* is calculated from:

$$d(i,j) = \frac{1}{\Delta^3} \sum_{k=1}^{N_p(j)} w(i,j,k)$$

where $N_p(j)$ is the number of probes in scatterplot *j*, *w(i,j,k)* is the contribution of probe *k* to the density at grid point *i*, and Δ is the grid-spacing (see Verdonk *et al.*, 1999).

Normalizing the maps

To scale the densities of different maps, we calculate an *average density*, *i.e.* the density that is expected by chance. At each grid point the probe density is divided by the average density, yielding the propensity:

$$p(i,j) = \frac{d(i,j)}{d_{av}(j)}$$

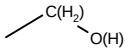
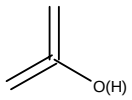
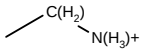
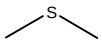
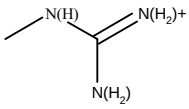
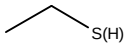
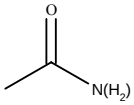
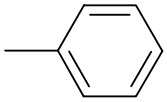
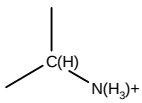
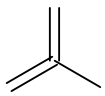
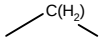
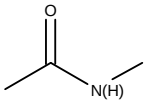
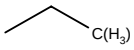
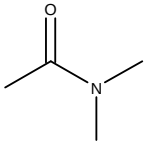
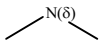
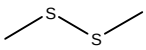
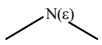
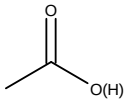
where $d_{av}(j)$ is the average density for scatterplot *j*. The propensity *p(i,j)* indicates whether the density at a grid point is higher or lower than is expected by chance. For example, a propensity of 2.0 implies that contacts at that grid point are twice as frequent as is expected. Propensities of grid points outside the $d_{vdw} + 0.5\text{Å}$ range of the target atoms in the central group are set to 1.0.

The average densities for CSD-based plots are defined as follows:

$$d_{av}(j) = \sum_{c=1}^{N_c(j)} \frac{N_{central}(c,j) N_{contact}(c,j)}{V_{cell}(c)}$$

where $N_{central}(c,j)$ and $N_{contact}(c,j)$ are the number of unique central groups and the total number of contact groups, respectively, in the unit cell of crystal structure *c*. $V_{cell}(c)$ is the volume of the unit cell of crystal structure *c* and $N_c(j)$ is the number of the crystal structures contributing to scatterplot *j*.

Table 1. List of the 20 central groups that were selected to represent the protein moieties. The fragment names, structures and the protein moieties in which the fragments occur are listed.

Name	Structure	Amino acids	Name	Structure	Amino acids
Aliphatic OH		Ser, Thr	Secondary aromatic NH		Trp
Aromatic OH		Tyr	Aromatic secondary CH		Trp, His
Aliphatic NH_3^+		Lys	Secondary S		Met,
Guanidino		Arg	primary SH		Cys
Carbamoyl		Asn, Gln	Phenyl ring		Tyr, Phe, Trp
Aliphatic tertiary CH		C_α	Aromatic tertiary carbon		C_γ in His and Trp
methylene		side chains	Peptide link		All amide bonds
methyl		Ala, Leu, Ile, Val	Proline peptide link		amide bonds in proline
δ-ring nitrogen		His	Disulphide link		S-bridges
ε-ring nitrogen		His	Carboxyl group		Asp, Glu

For PDB-based plots, we feel that this approach is not appropriate because the ligand (and therefore the contact group) atoms are not distributed throughout the entire unit cell of the protein structure. Instead, they are confined to the protein

binding site. Therefore we decided to use an alternative method to scale the PDB-based plots, also used by Laskowski *et al.* (1996) in the X-site program. In this approach, an *expected number of contacts*, $n_e(x, y)$, is calculated for each combination of central group *x* and contact group *y*, based on the total numbers of contacts these two groups form:

$$n_e(x,y) = \frac{\sum_{x'=1}^{20} n_o(x',y) \sum_{y'=1}^{6} n_o(x,y')}{\sum_{x'=1}^{20}\sum_{y'=1}^{6} n_o(x',y')}$$

where $n_o(x, y)$ is the *observed number of contacts* in the scatterplot of contact groups *y* around central group *x*. The summation over x' is over all 20 central groups used to describe protein functional groups (see Table 1); the summation over y' is over the six contact groups defined above. We can now define the contact propensity, $P_c(x, y)$, as the propensity of central group *x* to form contacts with contact group *y*:

$$P_c(x,y) = \frac{n_o(x,y)}{n_e(x,y)}$$

The expected number of contacts in scatterplot *j*, $n_e(j)$, is equal to $n_e(x, y)$, *x* and *y* representing the central and contact group, respectively, in scatterplot *j*. $n_e(j)$ can be translated into the average density for scatterplot *j*:

$$d_{av}(j) = \frac{n_e(j)}{V_{acc}(j)}$$

where $V_{acc}(j)$ is the volume in scatterplot *j* accessible to the contact group. It is calculated using stochastic sampling techniques, based on the Van der Waals radii of the target atoms in the central group and that of the probe atom in the contact group. Using this average density, PDB-based density maps can be scaled in the same manner as the CSD-based plots (see above).

Unfortunately, the above methodology cannot be used to scale CSD-based density maps. The method requires that there is no overlap between central groups, or between central and contact groups. In the PDB-based scatterplots, overlap does not occur, because the central and contact groups originate from different molecules, *i.e.* a protein and its ligand. For the CSD-based scatterplots, such a distinction cannot be made, and as a consequence, central and contact groups may originate from the same molecule.

Similarity calculations

Each PDB scatterplot was compared with the corresponding scatterplot from the CSD, *i.e.* containing identical central and contact groups. PDB based plots for which the central group protonation state is ambiguous, *i.e.* the histidine pro- and tele-nitrogens and carboxyl groups, were paired with CSD central groups with different numbers of hydrogens. As the PDB based plots do not contain hydrogen atoms, all hydrogens were removed from the central and contact groups of the CSD plots. Contact groups for which the probe atom is outside $d_{vdw} + 0.5\text{Å}$ of a target atom of the central group were removed from the CSD scatterplots.

Next, for each pair of CSD and PDB scatterplots, the central groups were aligned, and propensity maps were constructed using methodology described above. For each scatterplot pair, the optimal grid spacing was derived from the average number of probe atoms per grid cube, $< n_p(j) >$, of the less dense scatterplot:

$$< n_p(j) >= \frac{\Delta^3 n_o(j)}{V_{acc}(j)}$$

where $n_o(j)$ is the number of contacts in scatterplot *j* (see above). $< n_p(j) >$ is calculated for grid-spacing (Δ) values ranging from 0.5 up to and including 1.5 Å. The minimum grid spacing for which $< n_p(j) >$ is greater than or equal to 4.0 is used. If $< n_p(j) >$ is less than 4.0 for a grid spacing of 1.5 Å, no similarity calculation was performed.

To quantify the similarity of two propensity maps *A* and *B*, we used two similarity indices: the Carbó index and the Hodgkin index:

$$S_{carbó}(A,B) = \frac{\sum_{i=1}^{N_g} p(i,A)\,p(i,B)}{\sqrt{\sum_{i=1}^{N_g} p^2(i,A) \sum_{i=1}^{N_g} p^2(i,B)}}$$

$$S_{Hodgkin}(A,B) = \frac{2 \cdot \sum_{i=1}^{N_g} p(i,A) \cdot p(i,B)}{\sum_{i=1}^{N_g} p(i,A)^2 + \sum_{i=1}^{N_g} p(i,B)^2}$$

$p(i,A)$ and $p(i,B)$ represent the propensities at grid point *i* in propensity maps *A* and *B*, respectively. The summations are over all N_g grid points in the maps. The Carbó index is a scaling independent measure of the geometrical similarity

of the maps. The Hodgkin index also takes into account the overall intensities of the propensity maps.

Finally, we defined a weighted average similarity index for two central groups as:

$$S_w = \sum_{j=1}^{6} w(j) \cdot S_{Carbó}(j)$$

where the summation is over all six contact groups (probes) defined in the previous section. $S_{Carbó}(j)$ is the Carbó index for contact group j, and $w(j)$ is a normalized weight, which is proportional to the number of contacts in the less dense of the two scatterplots for contact group j.

SuperStar methodology

The overall methodology used to calculate SuperStar maps based on protein-ligand interactions from the PDB is virtually the same as described earlier for CSD-based maps (Verdonk *et al.*, 1999; Verdonk *et al*., submitted). The template molecule is prepared such that all the atom types and protonation states are set correctly. This also involves adding hydrogen atoms with the correct geometries. For PDB-based propensity maps the hydrogen atoms are only used for the 2D fragment matching (see below). Next, a probe is selected by the user. A probe is usually an atom in an IsoStar contact group, *e.g.* carbonyl oxygen. The rest of the procedure is fully automated and is described below.

Selection of fragments

The template molecule is built up automatically from 2D structure fragments from a library, using the methodology described by Verdonk *et al.* (1999). To select the optimal set of fragments to describe the template molecule, it is matched against all fragments in the SuperStar fragment library. For each atom in the template, the fragment that matches it best is used and the corresponding atom in that fragment is then defined to be a *target* atom. *Target* atoms are important when scatterplots are superimposed on the template (see the above section on IsoStar scatterplots, and that on superimposing scatterplots, below).

Selection and superimposing of scatterplots

First, the entire template molecule is placed on a 3D grid. For each fragment match in turn, the best 3D match of the corresponding IsoStar central group is found, using the methodology described by Verdonk *et al.* (1999). Then, applying the resulting transformation matrices, each scatterplot in turn is superimposed onto the corresponding part of the template molecule.

Calculation of individual propensity maps

After a scatterplot has been superimposed on the relevant part of the template molecule, a propensity map is calculated for a selected probe, using the methodology described above. In all SuperStar calculations presented in this chapter, we used a grid spacing of $\Delta = 0.7$Å. To prevent imprecise propensities having a detrimental effect on the final map, these propensities are set to 1.0 (see Verdonk *et al.*, 1999).

Combining different maps

For every grid point *i* in the 3D grid, the propensity *p(i,j)* in each individual scatterplot *j* is calculated using the methodology described above. Where maps from different central groups overlap, we multiply their propensities, to obtain the overall propensity, *i.e.*:

$$P(i) = P_O(i)\prod_{j=1}^{N_S} p(i,j)$$

where N_s is the number of scatterplots contributing to the composite SuperStar map. $P_o(i)$ originates from the propensity map that accounts for template atoms, other than those selected for the SuperStar calculation. These atoms are treated as hard spheres. For all grid points *i* where the probe is within $d_{vdw} - d_{tol}$ of a hard-sphere atom, $P_o(i) = 0$. For all other grid points, $P_o(i) = 1$. The maximum penetration distance, d_{tol}, is set at 0.4Å.

Hydrophobic correction

To correct for the lower frequency in the CSD of interactions between hydrophobic groups, relative to that in the PDB, we used the methodology described by Verdonk *et al.* (1999): for hydrophobic probes, the propensities at all grid points closest to a hydrophobic group in the template molecule were multiplied by a factor *P(hydro)*. In all the SuperStar calculations presented in this chapter, we used $P(hydro) = 10$.

Results & Discussion

Comparison of CSD and PDB based propensity maps

20 central groups were selected to represent the protein moieties (see Table 1). Six chemically different contact groups were used to represent ligand moieties, *i.e.* OH, C=O, NH, water, aliphatic CH and aromatic CH. Thus, 120 scatterplots, based on protein-ligand interactions observed in the PDB, were generated using only protein groups as central groups and ligand groups as contact groups. Each PDB-based scatterplot was paired with a CSD-based scatterplot with similar

central and contact group. Then each CSD/PDB scatterplot pair was converted into a pair of propensity maps, and the similarity of these two maps was assessed. For histidine ring nitrogens and glutamate and aspartate carboxyl groups, the protonation state is ambiguous. Therefore, these groups were excluded from the similarity assessments described in the two sections below. Also, pairs for which (one of) the scatterplots contained insufficient data were excluded.

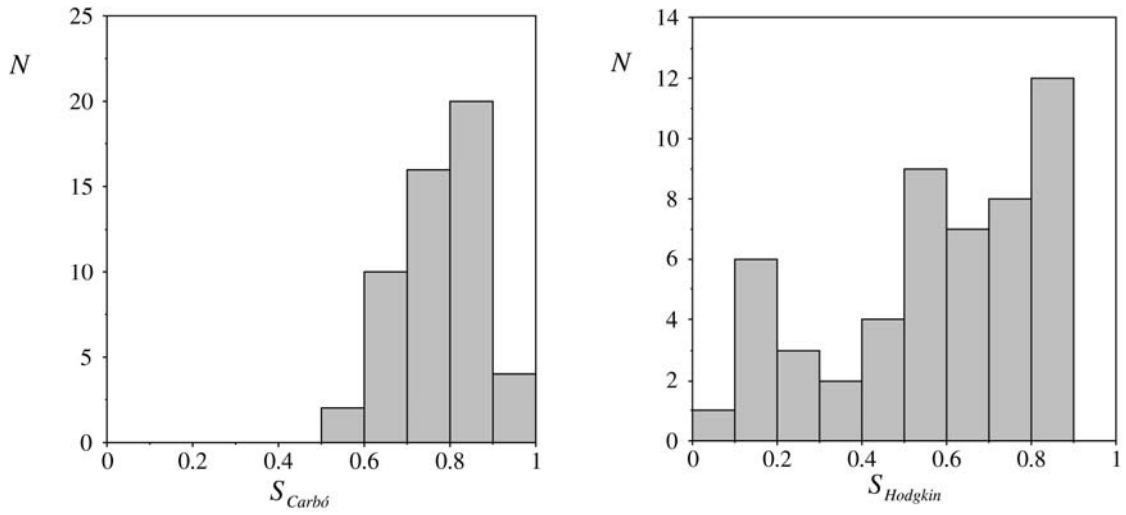


Figure 1. Histogram of Carbó indices (left) and Hodgkin indices (right), for CSD-PDB propensity map comparisons. Pairs for with insufficient data were excluded from the analysis.

Non-bonded interaction geometries

The Carbó index is used as an indicator of the geometrical similarity of a pair of propensity maps. The distribution of the Carbó indices of the pairs of CSD and PDB propensity maps is given in Figure 1 (left panel). The bulk of the indices are higher than 0.6, with an average of 0.76 (0.11). From these data we may conclude that the geometry of the interactions in the CSD and PDB scatterplots are similar for most scatterplot pairs. In some cases, however, the geometries of the CSD and PBD contact distributions differ. For most of these non-similar pairs, at least one of the scatterplots contains very little data (only slightly more than the threshold).

For some scatterplot pairs, however, the differences appear to be significant. One such case is the carbamoyl to carbonyl contact (see Figure 2), with a Carbó index of only 0.54. In the CSD plot (left panel in Figure 2), there are two favorable regions for a carbonyl group in the vicinity of the carbamoyl nitrogen, corresponding to the two hydrogen atoms bonded to that nitrogen. In the corresponding PBD plot (right panel in Figure 2), the ‘side-on’ interaction is hardly observed at all, and the plot is dominated by the ‘head-on’ interaction. This difference can be attributed to the fact that the asparagine and glutamine

side chains are surrounded by other protein moieties. At protein binding sites the ‘side-on’ interaction side is often difficult to access for ligand functional groups and, as a result, is frequently occupied by water molecules or by other protein hydrogen bond acceptors. In small-molecule crystal structures, this ‘side-on’ interaction side often *is* available for intermolecular interactions with carbonyl groups. This particular discrepancy between the CSD and PDB-based plot is unlikely to have a negative effect on the use of CSD-based plots in SuperStar: in a specific binding site, a glutamine residue may be exposed so that the ‘side-on’ interaction side is readily accessible to a ligand carbonyl. In that case the CSD-based plot would provide an even better representation of the hydrogen-bond donating properties of that glutamine side chain.

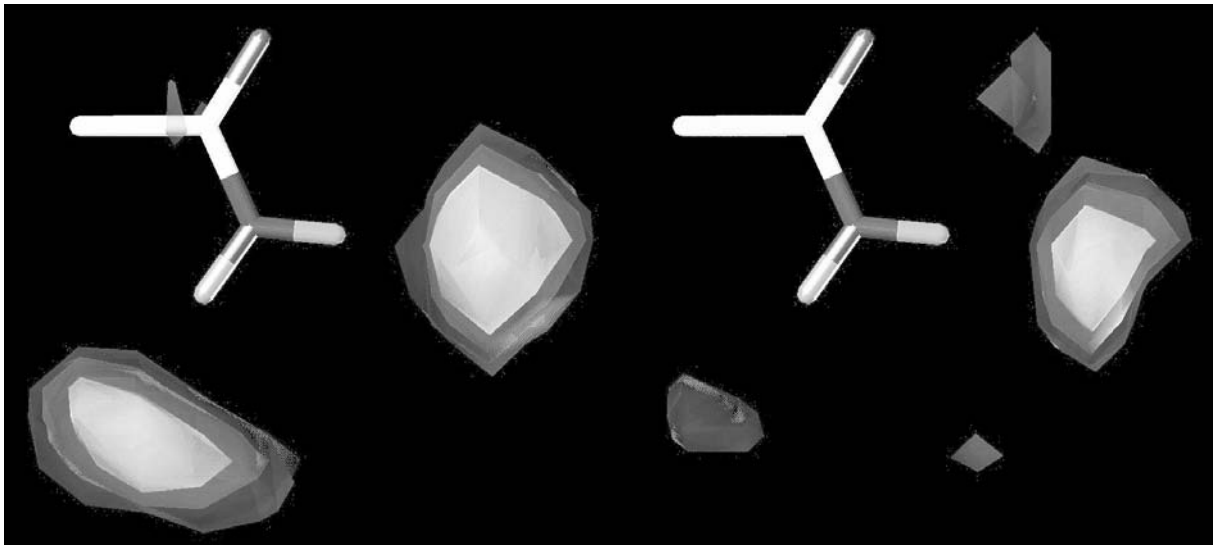


Figure 2. Density maps for interactions between carbamoyl groups (central group) and NH groups (contact group), derived from the CSD (left) and the PDB (right).

Non-bonded interaction frequencies

It may not be surprising that interaction *geometries* in the CSD and PDB are similar, since the underlying chemical principles of the interactions are the same. However, in a previous study (Verdonk *et al.*, 1999), we found that the relative interaction *frequencies* of certain interactions can be different for small-molecule crystal structures and protein-ligand interfaces. More specifically, we found that contacts between two methyl groups are more common in protein-ligand interfaces than in the CSD, when compared with carbonyl-methyl contacts. This difference was attributed to the hydrophobic effect caused by the aqueous environment around protein and ligand *in vivo*, which drives hydrophobic groups in ligands to form interactions with hydrophobic protein groups. Having normalized the PDB-based density maps to give propensity maps, we can now compare the relative interaction frequencies for all scatterplot pairs.

First, we used the Hodgkin similarity index to assess the similarity of each scatterplot pair. Apart from the geometries of the two propensity maps, the Hodgkin index also takes into account their overall intensities, *i.e.* the magnitudes of the propensities. The distribution of the Hodgkin indices of the pairs of CSD and PDB propensity maps is given in Figure 1 (right panel). The Hodgkin similarity indices are clearly lower than the Carbó indices, indicating that the intensities of the propensity maps differ considerably. Next, we calculated the contact propensities, P_c, for each PDB and CSD-based scatterplot, and compared the values for each CSD/PDB scatterplot pair. P_c reflects the relative propensity of interaction for a particular central and contact group combination. In Figure 3, the contact propensity of each CSD-based scatterplot is plotted against that of the corresponding PDB-based scatterplot. It is clear from this Figure that hydrophobic interactions (the filled circles) have higher contact propensities (interaction frequencies) in the protein-ligand interfaces than in the CSD. This is a confirmation of the hydrophobic effect we observed in a previous study (Verdonk *et al.*, 1999). But the reverse is also true: hydrophilic contacts (the open circles) are more common in the CSD than in protein-ligand interfaces. This 'reverse hydrophobic effect' probably reflects the fact that hydrogen-bonding groups in a protein can form hydrogen bonds to the water environment or to ligand hydrogen-bonding groups. Hence the driving force to form hydrogen bonds between proteins and ligands is not as strong as, *e.g.*, in a situation where a polar (small) molecule crystallizes from an apolar solvent. On average, contacts between hydrophilic and hydrophobic groups (the crosses in Figure 3), seem to be as common, or uncommon, in the CSD as they are in protein-ligand interfaces.

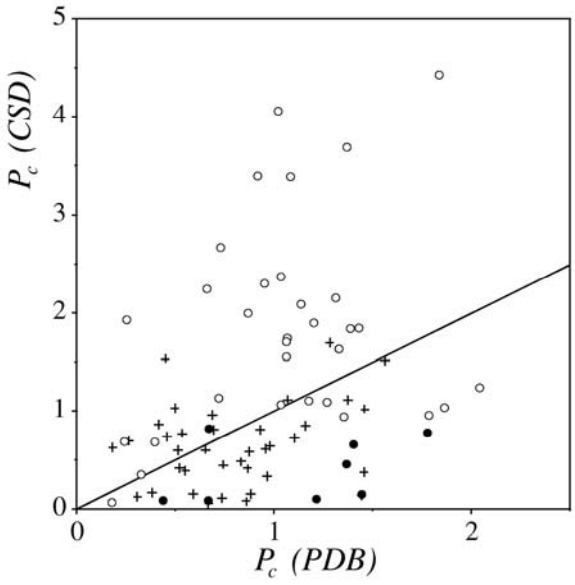


Figure 3. Contact propensities (P_c) of CSD propensity maps plotted against those of corresponding PDB propensity maps. Hydrophobic interactions are represented by filled circles, hydrophilic interactions by open circles and mixed contacts by crosses.

Protein groups with ambiguous protonation states

We concluded that interaction geometries in protein-ligand interfaces are generally quite similar to those in small molecule crystal structures. Hence, we

attempted to use Carbó similarity indices to derive the 'average' protonation state of carboxyl groups and histidine ring nitrogens in protein binding sites. For aspartate and glutamate carboxyl groups, we paired the PDB-based scatterplots for carboxyl groups with the CSD-based plots for (charged) carboxylate, and (uncharged) carboxylic acid. For each of the six contact groups, two Carbó similarity indices were calculated: one comparing the PDB-based scatterplot with the CSD-based plot for (charged) carboxylate, the other comparing the PDB-based plot with the CSD-based plot for (uncharged) carboxylic acid. The Carbó indices are averaged over all six probes to give *weighted average similarity indices*, S_w. $S_w = 0.81$ for the PDB-carboxyl *vs*. CSD-carboxylate, and $S_w = 0.63$ for the PDB-carboxyl *vs*. CSD-carboxylic-acid comparisons. S_w is clearly higher for the comparison with the CSD-carboxylate plots than it is for the comparison with the CSD-carboxylic-acid plots, which indicates that, in protein-ligand interfaces, most aspartate and glutamate side chains are charged. This is in agreement with the results of Hooft *et al*. (1996), who estimated that only 0.2% of all aspartate and glutamate side chains in protein structures in the PDB are protonated.

The PDB-based scatterplots for the histidine pro- and tele- ring nitrogens were compared to CSD-based plots for three different types of ring nitrogens: unprotonated, protonated (uncharged) and protonated (charged). S_w values for both pro- and tele- ring nitrogens were calculated for each of the three nitrogen types (see Table 2). Compared to the assignment of the protonation state of carboxyl groups, it is less clear what the predominant protonation state is for either of the two histidine ring nitrogens. If we use the S_w values as an indicator for the fraction of nitrogens in each of the three protonation states, the relative order of these fractions would be unprotonated > protonated (uncharged) > protonated (charged), for both the pro- and tele- nitrogens. This roughly agrees with the results of Hooft *et al*. (1996), who estimated that the fractions are about 40:40:20% for both ring nitrogens. The S_w value for the comparison of the PDB-based plots for the pro- and tele- ring nitrogens is 0.85, which is higher than any of the S_w values for comparisons with CSD-based plots. Again, this suggests that both nitrogens occur in more than one protonation state, and that the distributions over the three protonation states are similar for both ring nitrogens.

Table 2. S_w values for the PDB histidine pro- and tele- nitrogen comparisons with CSD nitrogens with different numbers of protons (*i.e.* uncharged, uncharged protonated and charged protonated).

	C-N-C (CSD)	C-NH-C (CSD)	$C=N^+H-C$ (CSD)
Pro-N (PDB)	0.77	0.73	0.67
Tele-N (PDB)	0.81	0.74	0.71

PDB-based SuperStar maps

For all SuperStar maps presented in this section, the ligand and all the water molecules were removed from the binding site, except where indicated differently. The protein binding sites (and ligands) were set up as described by Jones *et al.* (1995), *i.e.* hydrogen atoms were added with appropriate geometries, taking account of protonation states. For the PDB-based SuperStar maps, the hydrogen atoms were only used for the 2D fragment matching.

Example: Glutathione transferase

Figure 4 (right panel) shows the PDB-based propensity map of a carbonyl oxygen probe for the glutathione transferase binding site (PDB entry 1GLP). The ligand, glutathione sulfonic acid, is also shown, although that was removed during the calculation of the propensity map. High-propensity areas are clearly observed in the regions where the two carboxylate groups of the ligand bind. At the position of the amide oxygen that interacts deep in the pocket of the binding site, a strong peak is observed in the propensity map. For the second amide oxygen, however, no peak is observed. This oxygen atom is in fact solvent accessible, which makes it impossible for SuperStar to make accurate predictions in this region. One of the terminal sulfonate oxygen atoms in the ligand (although not a true carbonyl) overlaps with a high-propensity region in the SuperStar map. The remaining sulfonate oxygens, again, are solvent exposed. In fact, one of them forms a hydrogen bond to a water molecule in the binding site (not shown in Figure 4), which was not included in the SuperStar calculation.

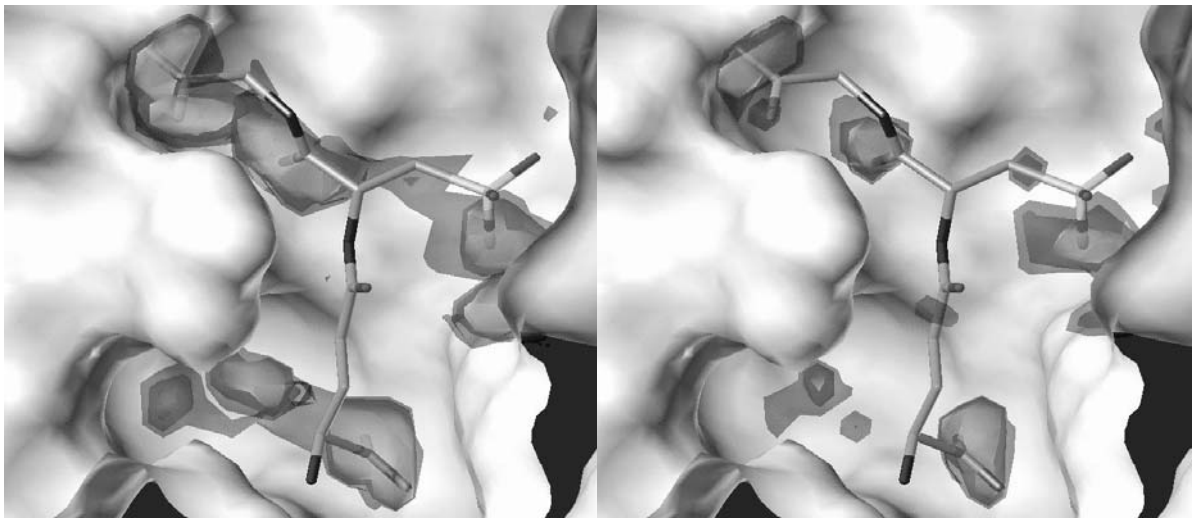


Figure 4. Superstar propensity maps of a carbonyl oxygen probe in the active site of a glutathione transferase complex (PDB entry 1GLP), based on CSD data (left) and on PDB data (right). The protein surface is shown in white, the ligand is represented by sticks. Contours are shown at propensity levels 2.0, 4.0 and 8.0.

Validation

The PDB-based maps were implemented in SuperStar (see Methodology section) and the performance of these PDB-based interaction fields in the prediction of ligand atoms was evaluated. To enable an accurate comparison between PDB and CSD-based SuperStar maps, we used the same test set that was used in the original validation of SuperStar (Verdonk *et al.*, 1999), *i.e.* 122 protein-ligand complexes from the PDB, and used the same success criteria. First, we check whether the ligand contains one of four types of atoms: a NH_3^+ nitrogen (a pure hydrogen bond donor), a carbonyl or carboxylate oxygen (a pure hydrogen-bond acceptor), a hydroxyl oxygen (a donor/acceptor) or a methyl carbon (a hydrophobic group). These atoms in a ligand will be referred to as "matching ligand atoms". In previous work (Verdonk *et al.*, 1999) we found that the validation results are strongly dependent on the solvent accessibility of the ligand atom involved. Therefore, all validation results are presented for two sets of ligand atoms, one set that contains all 431 matching ligand atoms and another set that contains the 127 matching ligand atoms with solvent accessibilities less than 2%.

For each matching ligand atom, we calculated SuperStar maps for four probes: NH nitrogen, carbonyl oxygen, hydroxyl oxygen and aliphatic CH carbon. In the validation of the CSD-based SuperStar maps, we could use probes that exactly match the four ligand atom types, *i.e.* NH_3^+ nitrogen, carbonyl oxygen, hydroxyl oxygen and methyl carbon. There is insufficient data in the PDB-based plots for the NH_3^+ and methyl probes, hence the more general NH nitrogen and aliphatic CH carbon probes were used instead. We used three success criteria: (i) $f_{P>1}$, the percentage of matching ligand atoms for which the propensity of the matching probe (*e.g.* if the matching ligand atom were a carbonyl oxygen, the propensity of the carbonyl oxygen probe would be used) is greater than its random expectation value of 1.0 (see Methodology section); (ii) $f_{P'<P}$, the average percentage of grid points, accessible to the probe, that have lower propensities than that at the matching ligand atom position; (iii) f_{corr}, the percentage of matching ligand atoms for which the propensity of the matching probe is higher than that of the other three probes.

Table 3 lists the averaged propensities, $<P>$, for each of the probes. Like we observed for the CSD-based SuperStar maps, there is a clear correlation between the $<P>$ values and the strength of the interactions these groups generally form (Verdonk *et al.*, 1999). Table 3 also lists $f_{P>1}$ and $f_{P'<P}$. The results are clearly better for solvent-inaccessible ligand atoms than for solvent-accessible ones, just like we observed in the CSD-based validation. The reason for this is obvious: SuperStar has no knowledge about water molecules or bulk water, so it cannot

be expected to make reliable predictions for ligand atoms that are solvent exposed. For solvent-inaccessible ligand atoms, $f_{P>1} = 81\%$, slightly lower than for the CSD-based validation, indicating that a smaller number of matching ligand atoms are found in regions of propensities above 1.0. However, $f_{P'<P} = 79\%$, which means that the matching ligand atoms are still found in the densest 20% of the relevant SuperStar map, which is comparable to CSD-based validation results.

Table 3. Propensities at the positions of matching ligand using PDB-based SuperStar maps

		All ligand atoms			Solvent-inaccessible atoms		
Atom type	<*P*>	*N*	$f_{P>1}$	$f_{P'<P}$	*N*	$f_{P>1}$	$f_{P'<P}$
C=O	10.2	81	72	63	28	89	87
C=O (anionic)	42.1	107	78	69	30	87	84
OH	11.7	98	72	61	29	93	84
NH_3^+	19.0	20	85	82	6	100	100
CH_3	3.5	125	57	57	34	56	58
Overall		431	70	63	127	81	79

It is interesting to note that for methyl groups, $f_{P>1} \approx 57\%$ is considerably lower than we observed for CSD-based maps (80%). This indicates that we may have overcorrected the CSD-based maps for the hydrophobic effect, and that the optimum hydrophobic correction factor should be lower than 10.0 (see Methodology section). On the other hand, $f_{P'<P} \approx 58\%$, compared to 50% for the CSD-based validation. This means that PDB-based plots seem more selective in the identification of hydrophobic areas in protein binding sites (see below). This is not surprising, considering the rather crude way in which the hydrophobic correction is applied on the CSD-based maps. Still, $f_{P'<P}$ is lower for ligand methyl groups, than for all the other groups, which reflects the less directional nature of hydrophobic interactions, compared to hydrogen-bonding ones.

Table 4 lists the results for the atom type predictions for both solvent-accessible and inaccessible ligand atoms. Overall success rates, f_{corr}, are marginally lower than those obtained in the CSD-based validation, 79% vs. 82% for solvent-inaccessible atoms. We note that the f_{corr} values for ligand methyl groups are considerably lower then what was found in the CSD-based validation, again indicating that we may have overcorrected the CSD-based maps for the hydrophobic effect (see above). On the other hand, the success rates for hydroxyl groups are much higher for the PDB-based plots than we observed in the CSD-based validation.

Table 4. Atom type predictions for matching ligand atoms using PDB-based SuperStar maps

	Predicted probe				
Observed probe	C=O	OH	NH	CH	f_{corr} (%)
A. Ligand atom type predictions for all atoms					
C=O	114	36	8	30	61
OH	17	57	12	12	58
NH_3^+	0	3	16	1	80
CH_3	3	14	18	90	72
Overall					64
B. Ligand atom type predictions for solvent-inaccessible atoms					
C=O	47	8	1	2	81
OH	4	23	1	1	79
NH_3^+	0	0	6	0	100
CH_3	0	4	6	24	71
Overall					79

CSD-based vs. PDB-based SuperStar maps

Generally, PDB and CSD-based SuperStar maps for protein binding sites are quite similar. An example of a direct comparison of a PDB-based SuperStar map with a CSD-based one is given in Figure 4. It shows the binding site of glutathione transferase, complexed with glutathione sulfonic acid (PDB entry 1GLP, see above), and the propensity maps of a carbonyl oxygen probe, based on CSD data (left panel) and based on PDB data (right panel). The propensities in the CSD-based map are higher, compared to the PDB-based map, presumably because of the ‘reverse hydrophobic effect’ discussed above. But all the same interaction sites are identified in both maps, reflecting the similarities of interaction geometries we observed in the systematic comparison of CSD and PDB-based scatterplots (see above).

The validation results demonstrate that, on average, CSD and PDB-based SuperStar maps are equally successful in predicting how ligand functional groups bind to proteins. This result is remarkable, particularly because there is much less data in most PDB-based scatterplots than there is in the corresponding CSD-based plots. The use of PDB data makes the calculations considerably faster, apparently without a loss of predictivity.

However, it is not in all cases preferable to use PDB-derived scatterplots. There are situations where PDB-based SuperStar maps are more valuable, but in other cases it is better to use CSD-based maps. The advantages and pitfalls of both types of maps balance out to give similar overall success rates in our validation exercise. In protein X-ray structures, the standard deviations in the

atomic positions are generally considerably larger than in small-molecule crystal structures, which causes a higher level of noise in the PDB-based plots, compared to the CSD-based ones. This means that CSD-based plots generally provide more accurate distributions, but at the same time are more sensitive to (small) experimental errors in the positions of protein atoms. The PDB-based maps inherently contain these positional uncertainties, and hence may identify interaction sites in pockets that are simply considered too tight, based on CSD-based maps.

In certain protein binding sites, the protonation state of an aspartate or glutamate carboxyl group, or that of a histidine ring, may be known. In these cases, PDB-based SuperStar maps are not very useful, as they have no knowledge of protonation states. A PDB-based map will still give a peak for an amino probe close to a histidine nitrogen, even if that nitrogen is known to be protonated. Separate CSD-based propensity maps are available for ring nitrogens of different protonation states, allowing different protonation states of histidine rings to be distinguished.

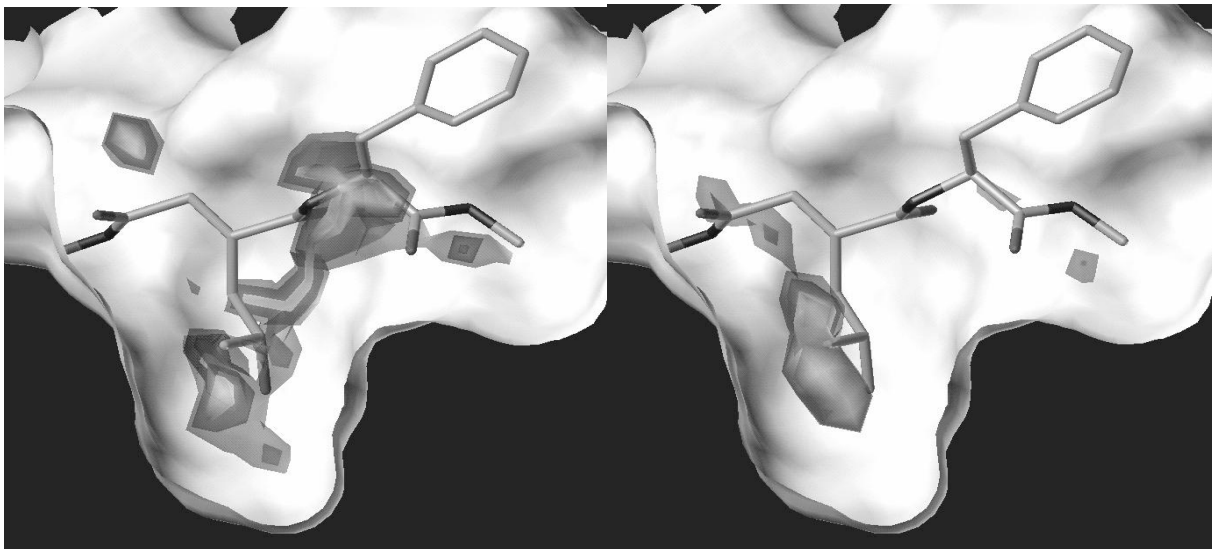


Figure 5. Superstar propensity maps of a methyl carbon probe in the active site of a fibroblast collagenase complex (PDB entry 1HFC), based on CSD data (left) and on PDB data (right). The protein surface is shown in white, the ligand is represented by sticks. Contours are shown at propensity levels 2.0, 4.0 and 8.0.

We concluded that PDB-maps appear to be more selective in the identification of hydrophobic interaction sites (see above). This is illustrated in Figure 5, which shows the binding site of a fibroblast collagenase complex (PDB entry 1HFC). This binding site contains a pocket that is quite hydrophobic, although it can contain a water molecule at the bottom. In this particular complex, an isopropyl group of the ligand occupies this pocket. The CSD-based propensity

map for a methyl carbon probe (left panel in Figure 5) clearly shows elevated propensities in this region, but also identifies various other regions as hydrophobic. The PDB-based map for an aliphatic CH carbon probe (right panel in Figure 5) is much more selective and clearly identifies the hydrophobic pocket. This finding illustrates that, unsurprisingly, the hydrophobic effect is accounted for more effectively in the PDB-based maps than in the CSD-based maps. Finally, CSD-based maps can be calculated for a wider range of probes than PDB-based ones. We only have sufficient data for 6 contact groups to calculate statistically significant PDB-based maps. CSD-based SuperStar maps are available for 23 contact groups, including groups like nitro and chlorine.

Conclusions

The systematic comparison of PDB and CSD-based scatterplots reveals that non-bonded interaction geometries in small-molecule X-ray structures are generally similar to those observed in protein-ligand interfaces. The relative frequencies of different types of interactions, however, are quite different. Interactions between hydrophobic groups are more common in protein-ligand interfaces, polar interactions are more common in small-molecule crystal structures. These differences are probably caused by the fact that a ligand and a protein binding site are stripped of a shell of water molecules before they interact, a scenario that does not normally apply to the formation of small-molecule crystals. In general, PDB and CSD-based SuperStar maps are quite similar, and are equally successful in the prediction of protein-ligand interactions. PDB-based maps are more suitable to identify hydrophobic pockets, *i.e.* using aliphatic or aromatic CH probes, and inherently take into account the experimental uncertainties of protein atomic positions. If the protonation state of a histidine, aspartate or glutamate side chain is known, specific CSD-based maps for that protonation state are preferred over PDB-based maps that represent an ensemble of protonation states.

Chapter 6

A Study on Ligand Binding to the Vitamin D Receptor in Relation to the Active Conformation of the Follicle Fluid-Meiosis Activating Sterol Analogs

Homology building and validation of the ligand-binding domain of the Vitamin D Receptor

D.R. Boer, M.L.C.E. Kouwijzer[1], J. Kelder, H. Kooijman and J. Kroon

[1] Department of Molecular Design and Informatics, N.V. Organon, Oss, The Netherlands.

Abstract

The follicular fluid-meiosis activating sterol (FF-MAS) has similar structure-activity relationships concerning the side chain stereochemistry as calcitriol (1α,25-dihydroxyvitamin D_3). To gain insight into the binding modes of calcitriol and its analogs and therefore possibly into the binding modes of FF-MAS and analogs, a homology model of the ligand binding domain of the human vitamin D receptor (VDR-LBD) is built based on the crystal structure of the human peroxisome proliferase activating receptor gamma. The results of this modeling study are evaluated using three crystal structures of a VDR-LDB construct complexed with calcitriol and two analogs. The sequence alignment is supported by secondary structure prediction techniques, which are found to give improvement compared to alignment based on sequence homology alone. The tertiary structure of the homology model compares well with that of the crystal structure, except for the loop region before helix 3, at the N-terminus. The binding modes of calcitriol and its analogs in the crystal structures are well reproduced in the binding pocket of the homology model by calculations with the program SuperStar (see chapter 5). In agreement with the crystal structures of the VDR-LBD complexes with calcitriol and its 20-epimer analogs, the SuperStar results indicate that only one binding site is available in the VDR-LBD pocket. For the same reason, the binding mode of gemini could not be derived. It may bind at a similar position in the VDR-LBD as one of the disordered components of eicosapentaenoic acid to the PPAR-δ structure through an induced-fit mechanism. The conformations of calcitriol and its 20-epimers are accessible to FF-MAS and its 22-oxo-20-epimer, respectively. It is therefore possible that MAS and its analogs bind to their putative receptor in a similar fashion as calcitriol and its analogs to the VDR. The side chains of the MAS analogs are in that case directed to the β-side of the steroid skeleton perpendicular to the hypothetical plane through the skeletal ring, which is different from the conformation found for MAS analogs in crystal structures and solution (see chapter 3).

Introduction

The follicular fluid-meiosis activating sterol (FF-MAS, compound **1** in Figure 1) shares an interesting structure-activity relationship with the endogenous ligand of the human Vitamin D receptor (VDR), calcitriol (compound **3** in Figure 2). For both classes of structures, compounds with inverted stereochemistry of the first atom of the side chain can be active, exemplified by the 22-oxo-20-epimer analog of FF-MAS (**2**) and the 20-epimer analog of calcitriol (**4**). Calcitriol is the hormonally active form of Vitamin D and effects biological processes via regulation of gene transcription and via non-genomic pathways (Bouillon *et al.*, 1995). Many of the genomic actions of calcitriol are mediated through the Vitamin D Receptor (VDR), a member of the nuclear receptor (NR) family. The similarity between the effects of the side chain structure of MAS compounds and calcitriol derivatives suggests that the binding pockets of the VDR and the putative receptor of MAS are topologically similar. Therefore, it may be possible to extrapolate the results of a study on the active conformations of calcitriol and its derivatives to the MAS compounds.

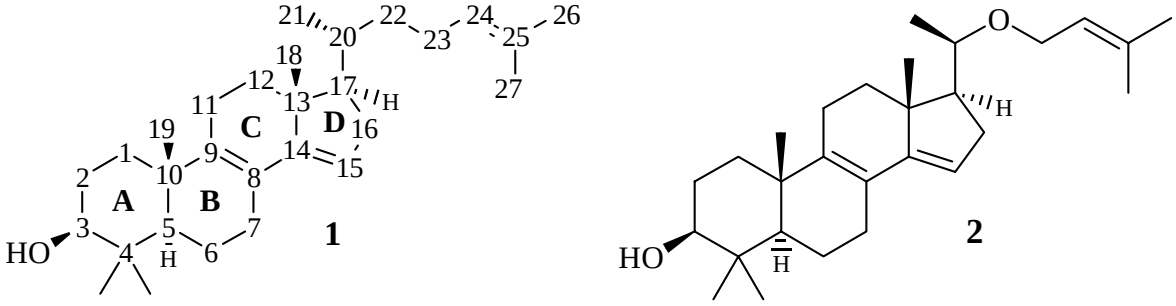


Figure 1. Chemical structures of FF-MAS (**1**) and its 22-oxo-20-epimer (**2**).

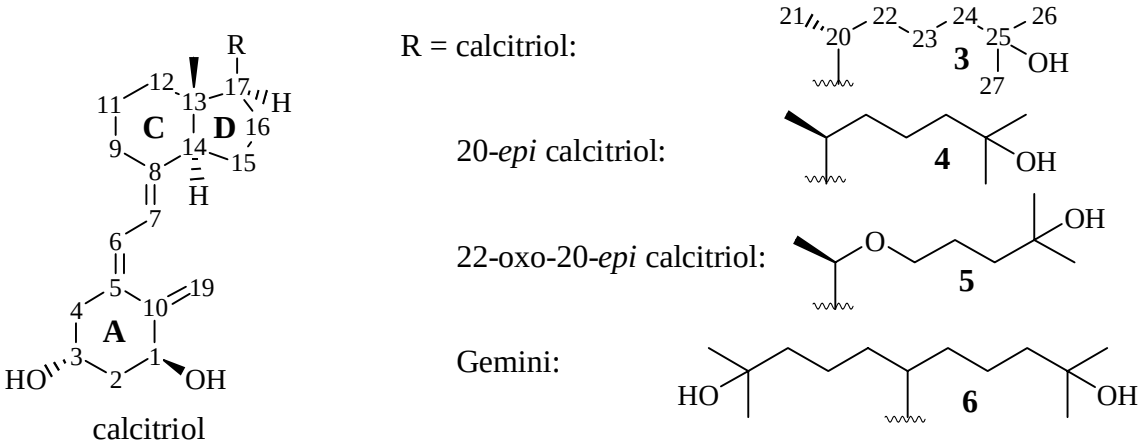


Figure 2. Molecular structures of calcitriol (**3**), its 20-epimer (**4**), a 20-oxo-20-epimer (**5**) and an analog with two side chains (gemini, **6**).

At the start of the work described in this chapter, no structural model of the ligand-binding domain (LBD) of the VDR was available. Therefore, a homology model was built in order to study possible binding modes of the ligands **3**-**6**. The homology model of the VDR-LBD was based on the crystal structure of the human peroxisome proliferator activating receptor (PPAR-γ). Recently, two other models were published (Norman *et al.*, 1999a; Yamamoto *et al.*, 2000). We used different approaches in the derivation of our homology model and the determination of the binding modes of the ligands compared to the two homology models published earlier. Firstly, secondary structure-prediction techniques were applied for the alignment of the sequences of the ligand binding domains of the VDR and PPAR-γ. Secondly, the program SuperStar was used to find the binding modes of the ligands shown in Figure 2. This chapter discusses the results obtained with these two approaches. The implications of the binding modes of calcitriol and its analogs on the active conformations of the meiosis activating sterol compounds will be discussed.

Our homology model and the derivation of the binding modes of calcitriol and its analogs are evaluated using three recently published crystal structures of a VDR-LBD construct, a deletion mutant missing residues S165-P215, complexed with calcitriol (Rochel *et al.*, 2000), its 20-epimer **4** and the 22-oxo-24,25,26-tri-*homo*-20-epimer analog **5** (Tocchini-Valentini *et al.*, 2001). The crystal structure of the ligand-binding domain of the VDR is another example of an LBD structure of a nuclear receptor. Those of the peroxisome proliferase activating receptor γ (Nolte *et al.*, 1998), the estrogen receptor (Brzozowski *et al.*, 1997; Tanenbaum *et al.*, 1998), the progesterone receptor (Williams & Sigler, 1998; Matias *et al.*, 2000), the androgen receptor (Matias *et al.*, 2000; Weinmann *et al.*, 2001), the retinoic acid receptor γ (Renaud *et al.*, 1995) and the thyroid hormone receptor (Wagner *et al.*, 1995) have, among others, been solved. The N-terminal sequence of the NR-LBD's is highly variably, both in sequence and structure. The VDR is not different in this respect: in contains a long loop between the well defined DNA binding domain and helix 3 of the LBD with no sequence homology to any of the known nuclear receptors.

Calcitriol induces a conformational change in the VDR, which leads to a cascade of processes that result in the genomic response (Peleg *et al.*, 1995). The conformational change occurs in the so-called activating factor 2 (AF-2), which consists of helix 12 and the preceding loop, and induces heterodimerization of VDR with the retinoid X receptor at a dimerization surface near helices 10 and 11. The dimerization enhances the complexation of the DNA binding domain of the VDR to DNA and thereby affects gene activation or repression (Bouillon *et al.*, 1995). As mentioned above, an interesting aspect of the structure-dependent activity of calcitriol concerns the stereochemistry at C20. The 20-epimers **4** and **5**, with inverted stereochemistry at C20 compared to calcitriol, increase transcriptional activity of the VDR (Liu *et al.*, 1997) and are therefore called

superagonists. Compound **6**, containing both side chains (called gemini), was found to have a similar affinity as calcitriol and has similar superagonist properties as the 20-epimers (Norman *et al.*, 2000). *In computro* studies designed to determine the active conformation of the ligands have been performed in an attempts to explain the behavior of these compounds. Calcitriol is a more flexible ligand than other NR ligands, (*e.g.* progesterone, androsterol and estradiol) with respect to the aliphatic side chain and the conjugated π-system between ring A and C. Several studies have been published that attempted to find the active conformation of the side chains of calcitriol and its 20-epimer analogs (Midland *et al.*, 1993; Yamamoto *et al.*, 1996; Liu *et al.*, 1997; Yamada *et al.*, 1998; Yamamoto *et al.*, 1999). The conformational behavior of ring A and the conjugated π-system between rings A and C have also been studied (Mosquera *et al.*, 1988; Norman *et al.*, 1999b).

Because of the high flexibility of the natural ligand and the large size of the VDR binding pocket, the number of possible conformations and positions in which a ligand can be docked in the active site is large. It would take too much time to consider all of the possible binding modes separately. Rather, one would like to find regions in the binding pocket in which the hydroxyl and aliphatic groups of compounds **3**-**6** preferably bind. For this end, we employed the program SuperStar described in chapter 5 (see also Verdonk *et al.*, 1999, 2001). This program calculates propensity grid maps for a variety of chemical functional groups in the binding pocket of a protein. From the SuperStar maps, high-propensity regions (called hot-spots) for the functional groups can be derived, which may provide clues on the binding modes of the ligands. The SuperStar maps are calculated for the homology model as well as the crystal structures. Since the homology model is a crude approximation of the actual structure of the VDR-LBD, it will be interesting to compare the maps calculated for the homology model to those obtained for the crystal structures.

Methods

The sequence of the human VDR-LBD (residues S119-S427) was aligned to that of the LBD of the PPAR-γ using CLUSTALW, v. 1.74 (Thompson *et al.*, 1994). The secondary prediction server JPRED (Cuff *et al.*, 1998) was used to find secondary structural elements. The atomic coordinates of the PPAR-γ-LBD template were retrieved from the Protein Data Bank (entry 2PRG). Twenty alternative models were built with MODELER (v5.0, Šali & Blundell, 1993) and fully optimized, after which the best model was selected based on the objective function implemented in MODELER. This model was optimized with CHARMm force field in Quanta (version 1998, Molecular Simulations Inc., Cambridge), using steepest descent for 25 steps followed by Newton Raphson minimization for 500 steps. During minimization, a distance dependent

electrostatic potential was used (ε=4R), the backbone atoms were fixed and a harmonic restraint was applied to the Cβ atoms.

SuperStar, based on PDB data, was used to calculate propensity maps for hydroxyl and aliphatic CH functional groups in the binding pockets of both the homology model and the crystal structures of the VDR-LBD complexes with the ligands **3**-**5**. All water molecules and ligands were removed from the crystal structures, after which hydrogen atom positions for the protein residues were calculated using InsightII (version 98, Molecular Simulations Inc., Cambridge, UK) at a pH of 7.4. The binding cavity search algorithm implemented in SuperStar was applied. The maps were calculated with a grid spacing of 0.7 Å. The SuperStar propensity maps were analyzed using InsightII.

Results & Discussion

Model building and comparison to the crystal structure

The sequence homology between the PPAR-γ receptor and the VDR was 33% similarity for residues I131-S427. For residues M226-S427, the homology is higher, *i.e.* 53% similarity. In Figure 3, the alignment of the VDR sequence to that of the PPAR-γ template is given, as well as the sequence of the VDR crystal structure. The N-terminus of the VDR-LBD has low sequence homology with the PPAR-γ and can therefore not be aligned properly. Since this part of the structure may interact with the ligand, it is important to model the amino acid stretch before helix 3 where possible. For this purpose, we used secondary structure prediction methods. These show that amino acids Q128-I134, T142-S148 and C151-Q152 are likely to adopt α-helix conformations. The secondary structure of the residues following the predicted helices, *i.e.* residues G164-L221, could not be assigned and therefore this amino acid stretch was left out of the model. Using residual sequence homology of the VDR-LBD sequence with that of the PPAR-γ template, the N-terminal sequence stretch containing the predicted α-helices was aligned to the template.

As can be seen in Figure 3, a shift of 7 residues for the start of helix 1 occurs in the alignment of the sequences of helices 1 and 2 of the homology model and the crystal structure of the VDR. The prediction of the N-terminal part of the sequence is better than that based on alignment of the sequence with the thyroid hormone receptor (Norman *et al.* 1999a). However, Yamamoto *et al.* (2000) were able to align the VDR sequence to that of the retinoic acid receptor-γ correctly using multiple sequence alignment. This shows that secondary structure prediction methods can help the alignment of two sequences in case the homology between template sequence and the aligned sequence is low. A slightly higher level of homology would be required than that existing between the sequences of the LBD's of the VDR and PPAR-γ, because the observed shift of seven residues is too large to correctly predict the involvement of Y143 in

binding calcitriol, as observed in the crystal structure of the VDR-LBD complex with calcitriol. We suggest that structure prediction is a useful extra tool for alignment.

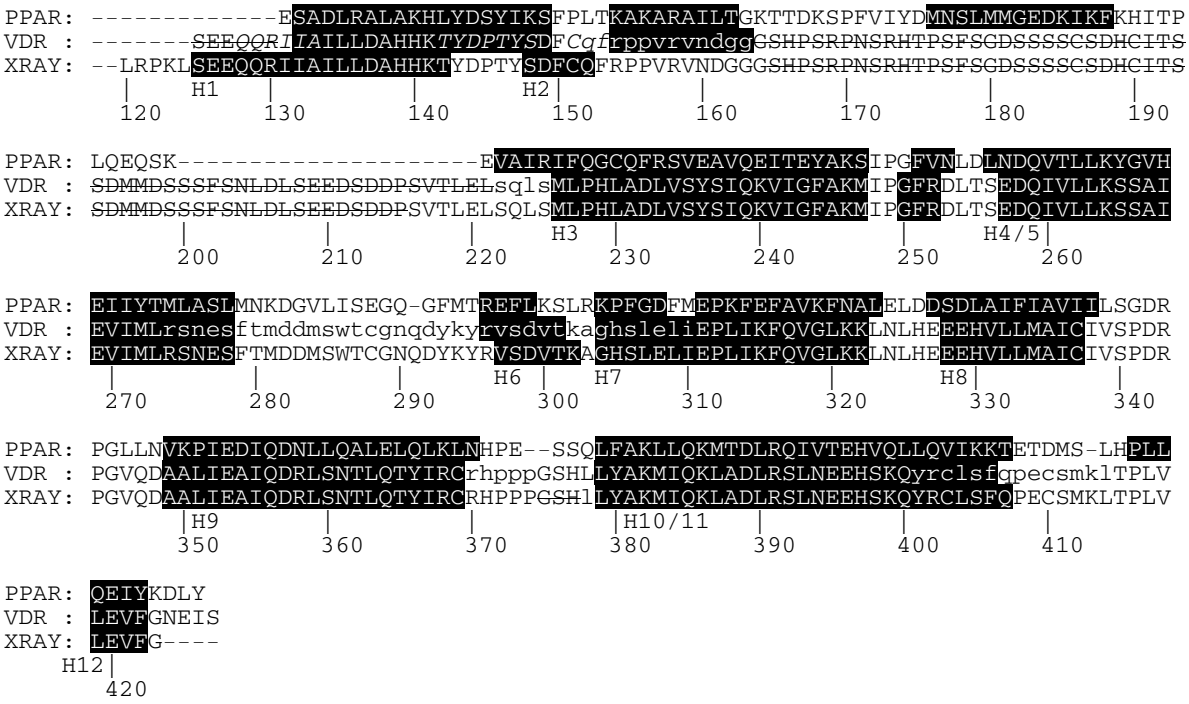


Figure 3. Alignment of the PPAR-γ (top), VDR homology model (middle) and the VDR X-ray structure (bottom) sequences. In all sequences, α-helices are indicated by white letters on black background. Residues that were not present in the homology model or in the X-ray structure are indicated by strike-through. Residues that were predicted to adopt and α-helical conformation by JPRED are made italic in the sequence of the homology model. Residues in the homology model with low structural similarity compared to the X-ray structure are shown in small caps.

An overlay of the cartoon representations of our homology model and the crystal structure of the VDR-LBD complex with calcitriol is given in Figure 4. The homology model consists of antiparallel helical triple sandwich topology typical for the nuclear receptor family. In the center of the model, a large cavity is situated, at a similar position as in the crystal structures of the ligand binding domains of VDR and other NR's. Mutation studies showed that residues S237 and H397 were found to be important for ligand binding (Väisänen *et al.*, 1998; Yamamoto *et al.*, 2000). In our homology model, these residues are located near the binding pocket region. The interface between helix 3 and helix 12, which packs against helix 3 upon ligand binding, consists of a hydrophobic core surrounded by several charged residues. Several charged groups are scattered over the dimerization surface. In conclusion, the topology and distribution of the

amino acid side chains were therefore in accordance with experimental biological data.

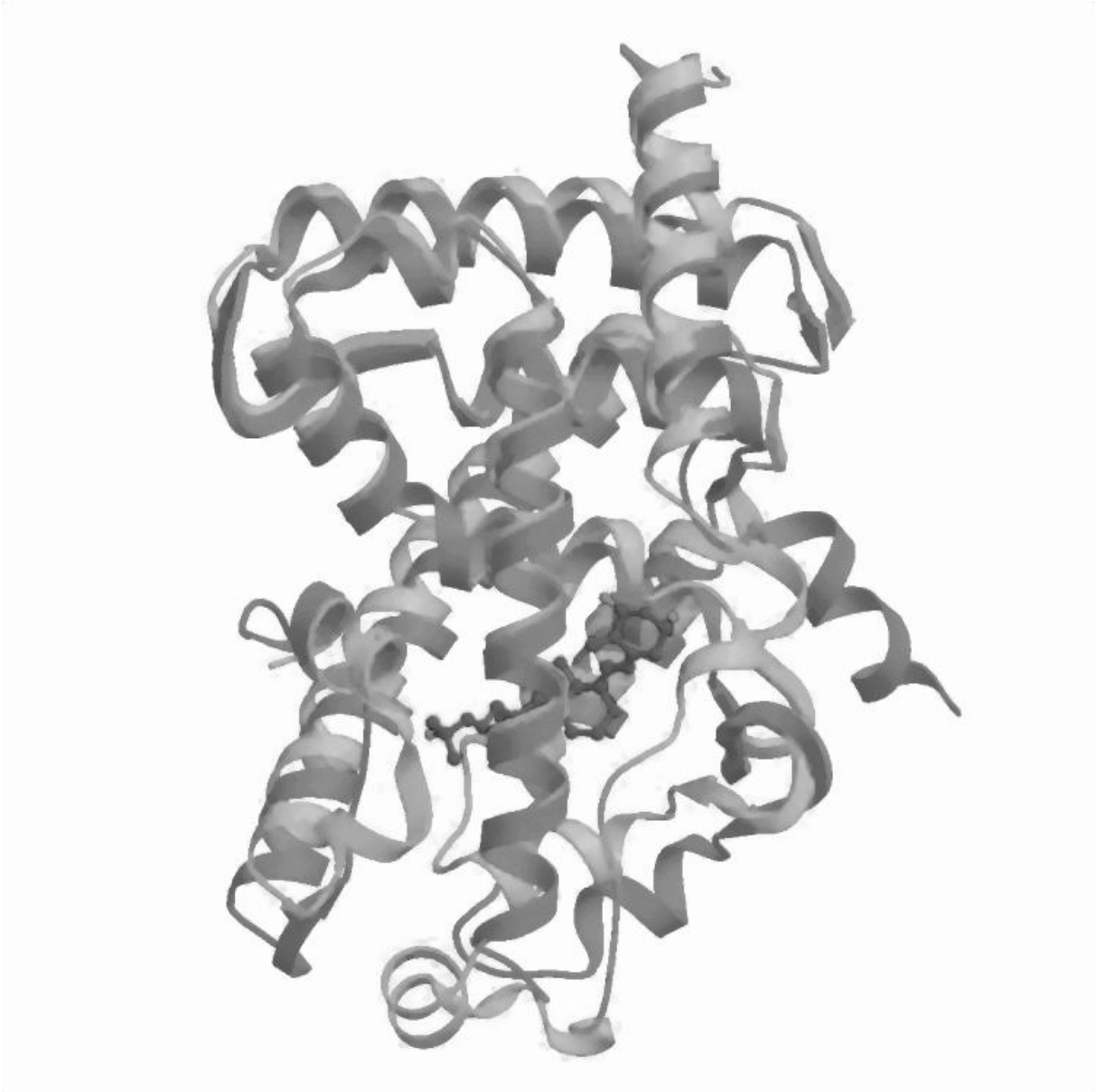


Figure 4. Overlay of the cartoon representation of our homology model and the crystal structure of the VDR. The model is shown in dark gray, the crystal structure in light gray. A ball-and-stick representation of calcitriol as bound in the crystal structure complex is shown as well.

Differences between the VDR-LBD homology model and the crystal structure occur in helix 2 and the β-strands between helices 5 and 6, which are neighboring the N-terminal loop region. These differences are associated with the low homology of the loop in the N-terminal sequence of the VDR-LBD with the PPAR-γ. The loop of the activation factor (AF-2), which is located at the C-terminus, also has low structural homology to the crystal structure. However, these differences occur far away from the ligand-binding pocket and are not

expected to impair the docking studies, in contrast to the region close to the β-strands.

Examination of the binding pockets

The SuperStar propensity maps of the aliphatic CH and OH groups, calculated for the binding pocket of the crystal structure of the VDR-LBD complex with calcitriol, are given in Figure 5. The CH and OH SuperStar maps calculated for the binding pockets of the VDR-LBD complexes with the analogs **4** and **5** are similar to those obtained from the crystal structure of the complex with calcitriol. The CH map (left panel in Figure 5) shows high propensities at the position the diene system. The OH map (right panel in Figure 5) shows high-propensity regions near, but not at the positions of the ring A hydroxyl groups of calcitriol. These hydroxyl groups form hydrogen bonds to water molecules that are present in the binding pocket of the crystal structures. The positions of these water molecules correspond to the high-propensity regions close to the β-strands and their positions are correctly predicted. The water molecules cannot at present be included as protein moieties in SuperStar. As a consequence, SuperStar has no knowledge on the presence of these mediating water molecules, which is why the hydroxyl groups are not predicted. The 25-OH group of calcitriol makes interactions with the protein directly and a region with very high propensity is indeed found in the SuperStar map at the corresponding position.

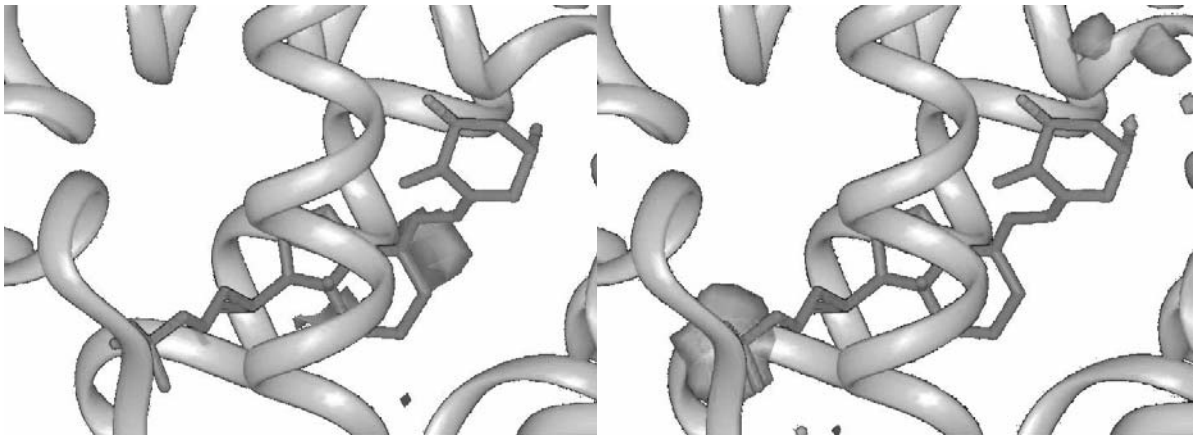


Figure 5. SuperStar propensity maps of CH (left panel) and OH (right panel) contact groups in binding pocket of the crystal structure of the VDR complexed to calcitriol. The CH grid maps are contoured at level 4.0 (shown in dark gray) and the OH grid maps are contoured at level 2.0 (shown in gray).

The SuperStar maps of the CH and OH groups calculated for the binding pocket of our homology model are given in Figure 6. In these Figures, the calcitriol ligand is in the same orientation as in the superposition of the homology model and the crystal structure as shown in Figure 4. A high-propensity peak is found in the CH map (left panel in Figure 6), close to the C21 methyl group. The OH

map (right panel in Figure 6) shows three regions where an OH group is likely to occur. Two of these are associated with a glutamate that is positioned in helix 1. The model in this region is less reliable because of low homology and therefore this part of the SuperStar map should be interpreted with care. The third high-propensity region was found near helix 11 and is considered more reliable, because this part of the binding pocket is not solvent-exposed and the alignment in this region was of sufficient quality. This OH hot-spot is located close to H397, which was identified as an important residue in binding calcitriol by mutation studies (Yamamoto *et al.*, 2000). Thus, it would seem reasonable to position one of the three hydroxyl groups onto this hot-spot. In addition to the anchors derived from the SuperStar calculations, mutation studies of the VDR provide additional clues on specific interaction sites of calcitriol. *E.g.*, S237 is very likely to interact with the ligand (Väisänen *et al.* 1998; Swamy *et al.*, 2000). Using the high-propensity region in SuperStar maps, the mutation data and the overall shape of the pocket, it would reasonable to position the VDR ligands as shown in Figures 4-6, *i.e.* the position found in the crystal structures.

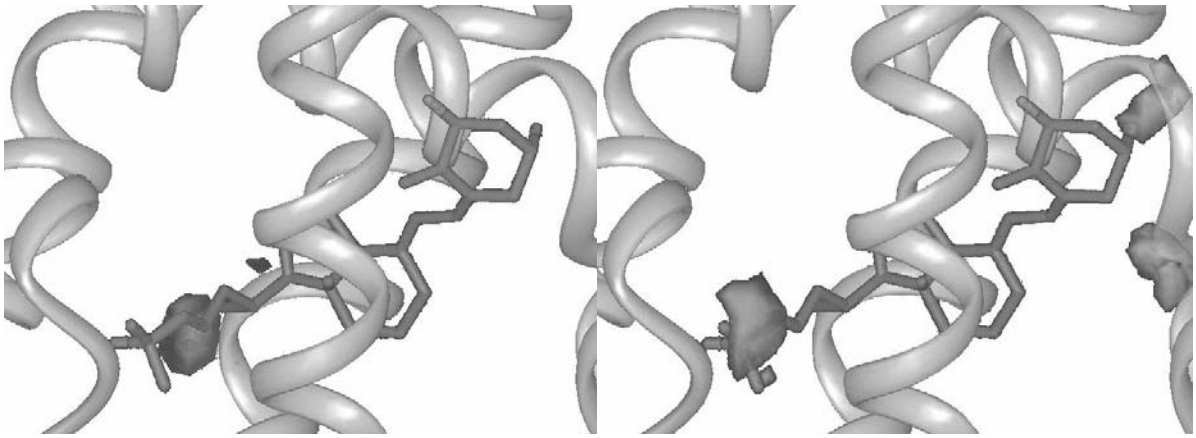


Figure 6. SuperStar propensity maps of CH (left panel) and OH (right panel) contact groups in the binding pocket of the homology model. The ligand from the crystal structure is shown, after superpositioning the homology model and the X-ray structure. The CH grid maps are contoured at level 4.0 (shown in dark gray), the OH grid maps are contoured at level 2.0 (shown in gray).

Conformation of bound calcitriol and its analogs

With the information provided by SuperStar and the mutation studies, a rough binding mode of calcitriol can be established. However, the details of the conformation that the ligand adopts in the binding pocket remain to be determined. Previous computational studies on the side chain conformations of calcitriol *in vacuo* indicate that in low energy conformations, the side chain of the natural isomer is extended and points away from the skeleton. In this side chain orientation, the 25-OH group is located in a region designated EA. This orientation corresponds to that found in the crystal structure of calcitriol (Suwinska & Kutner, 1996). In low energy conformations of the 20-epimer **4**,

the 25-OH group is found in a region near the edge of ring C, called region A. The EA conformation corresponds to the crystal structure conformations of several FF-MAS analogs with natural configuration at C20, the A conformation with that found in the crystal structure of compound **2**.

In contrast to the computational analysis of the ligands, the side chains in the crystal structures of the VDR-LBD complexes are pointing upwards with respect to the skeleton, towards the β-side, at an angle of 90° with respect to the plane through the skeleton rings. This conformation was not identified from the molecular mechanics studies and is somewhat in between the A and EA conformations. Apparently, this high-energy conformation is stabilized by interactions with the receptor. In contrast to what was expected from mutation and protein digestion studies upon binding of calcitriol and the 20-epimers (Peleg *et al.*, 1995; Liu *et al.*, 1997; Väisänen *et al.*, 1998; Väisänen *et al.*, 1999b), the binding of the latter is not accompanied by a substantial conformational change in the crystal structure of the complexes (Tocchini-Valentini *et al.*, 2001). In accordance with these crystal structures, the SuperStar calculations do not provide an alternative binding site for the 20-epimers, since no additional hydroxyl hot-spots were found in either the homology model or the crystal structures.

The fact that gemini binds to the VDR-LBD and induces a higher transcription level compared to calcitriol cannot be explained by the crystal structures of the VDR-LBD complexes. SuperStar also does not provide a clue since additional hot-spots for hydroxyl groups are not found in the calculated maps. However, the crystal structure of the LBD of the PPAR-δ complex with eicosapentaenoic acid (Xu *et al.*, 1999) provokes new ideas on the binding of the gemini. The hydrophobic tail of bound eicosapentaenoic acid is disordered over two pockets. One of the disorder components is bound similar to the way calcitriol binds to the VDR. The other component partly binds in the same pocket as the first component, but is bent such that it partly follows helix 3 towards its N-terminus, a region that is occupied by protein residues in the VDR structure but which may be accessible through an induced-fit mechanism. The Y-shaped pocket in the PPAR-δ structure is complementary to the shape of compound **6**. The region unoccupied in the VDR complex may be used by gemini to accommodate one of its side chains.

Implications on FF-MAS binding to its receptor

The bound conformation of calcitriol and its 20-epimers **4** and **5** to the VDR-LBD construct sheds new light on the possible binding conformation of FF-MAS. As shown in chapter 3 (see also Nes *et al.*, 1984), the crystal structures and solution conformations of MAS analogs with either natural or unnatural C20 configuration correspond to what was named the north-east and north-west orientations respectively (the EA and A regions in the discussion above). One

would expect calcitriol to adopt the conformation found in its crystal structure upon binding and not the conformation in which it is found in the VDR. Apparently, the protein environment stabilizes the bound conformation of calcitriol. The same feature could apply to the MAS compounds, which crystallize with the side chain in a similar conformation as calcitriol in its crystal structure, when they bind to their receptor. Molecular mechanics calculations *in vacuo* using the CFF91 force field in InsightII on different rotamers of the C17-C20 torsion, keeping the side chain in an extended conformation, show that the +60° and –60° conformations were 1.2 kcal/mole higher in energy than the 180° conformation. A similar analysis of the rotamers of the 22-oxo-20-epimer **2** shows that the conformation corresponding to the +60° conformation of MAS is 1.9 kcal/mole higher in potential energy than the northwest conformation found in its crystal structure. These energy differences are sufficiently low to be compensated for by protein-ligand interactions.

The analogy in the relation of side chain structure and resulting activity for MAS and the VDR ligands, suggests that MAS and its 22-oxo-20-epimer have similar conformational behavior of calcitriol and its analogs upon binding to their respective receptors. If this is the case, FF-MAS and its 22-oxo-20-epimer bind with their side chains perpendicular to the imaginary plane through the skeletal rings of the steroid skeleton, pointing towards the β-side of the molecules. Constricted analogs that direct the side chain in this direction have been synthesized and were found to be active, as shown in chapter 3. It is not unlikely that the 22-oxo-20-epimer of FF-MAS binds in a similar conformation as compound **5** due to the 22-oxo moiety in the side chain. The Δ^{24} double bond in the side chain of the 22-oxo-20-epimer cannot reach the same region in space as that of FF-MAS itself, but this is expected to be of minor influence on the affinity of the compound, since the hydrophobic contacts are of a less specific nature as the hydrogen bond network formed by the 25-OH group of calcitriol and its analogs.

Conclusions

FF-MAS might bind to its receptor in a similar way as calcitriol and its analogs bind to the VDR-LBD construct as found in the crystal structures of the respective complexes. Extrapolation of the bound conformation of the calcitriol and its 20-*epi* derivatives suggests that MAS and its analogs bind in a conformation in which the side chain is oriented perpendicular to the hypothetical plane through the skeletal rings. Only one conformation would then be responsible for the activity of the MAS compounds. This would be at odds with the observed conformations of MAS analogs in their crystal structures, which suggest two different conformations. For calcitriol and its analogs, the bound conformation, which is higher in energy, is probably stabilized by

binding to the VDR, a feature that could also apply to the MAS analogs when binding to their receptor.

Secondary structure prediction is helpful as an additional tool in the alignment of the sequences of the model and template proteins. Our homology model of the VDR-LBD has similar over-all topology and tertiary structure as determined in the crystal structures. SuperStar was quite powerful in the prediction of interaction hot-spots in the homology model, which can be considered a rough approximation of the crystal structure. It has correctly located the 25-OH and the C21 methyl groups, which are valuable anchoring points for positioning calcitriol and its analogs. In addition, the SuperStar maps did not provide evidence for a second binding site, in accordance with the crystal structures available for the VDR-LBD complexes. Finally, gemini may accommodate one of its side chains in a pocket analogous to a site occupied by bound eicosapentaenoic acid in the crystal structure of the PPAR-δ complex.

References

Allen, F.H. & Kennard, O. (1993). “3D search and research using the Cambridge Structural Database”, *Chem. Design Automation News* **8**, 31-37.

Allinger, N.L., Yuh, Y.H. & Lii, Y.-H. (1989). “Molecular mechanics. The MM3 force field for hydrocarbons. 1.”, *J. Am. Chem. Soc.* **111**, 8551-8566.

Bax, A. & Davis, D.G. (1985). “MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy”, *J. Magn. Reson.* **65**, 355-360.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000). “The Protein Data Bank”, *Nucleic Acid Res.* **28**, 235-242.

Boer, D.R., Kooijman, H., van der Louw, J., Groen, M., Kelder, J. & Kroon, J. (2000). “Calculated heats of formation of sterol diene isomers compared with synthetic yields of isomerisation reactions of $\Delta^{5,7}$ sterols”, *J. Chem. Soc. Perkin Trans. 2*, 1701-1704.

Böhm, H.-J. (1992a). “The computer program LUDI: a new method for the de novo design of enzyme inhibitors”, *J. Comput-aided Mol. Design* **6**, 61-78.

Böhm, H.-J. (1992b). “LUDI: rule-based automatic design of new substituents for enzyme inhibitors”, *J. Comput-aided Mol. Design* **6**, 592-606.

Boullion, R., Okamura, W.H. & Norman, A.W. (1995). “Structure-function relationships in the vitamin D endocrine system”, *Endocr. Rev.* **16**, 200-257.

Bruno, I.J., Cole, J.C., Lommerse, J.P.M., Rowland, R.S., Taylor, R. & Verdonk, M.L. (1997). “IsoStar: a library of information about nonbonded interactions”, *J. Computer-Aided Molec. Design* **11**, 525-537.

Bruno, I.J., Cole, J.C., Lommerse, J.P.M., Rowland, R.S., Smith, B., Taylor, R. & Verdonk, M.L. (1999). “IsoStar v1.2”. Cambridge Crystallographic Data Center, Cambridge. UK.

Brzozowski, A.M., Pike, A.C.W., Dauter, Z., Hubbard, R.E., Bonn, T., Engström, O., Öhman, L., Greene, G.L., Gustafsson, J.-Å. & Carlquist, M. (1997). “Molecular basis of agonism and antagonism in the oestrogen receptor”, *Nature* **389**, 753-758.

Byskov, A.G., Andersen, C.Y., Nordholm, L., Thogersen, H., Guoliang, X., Wassmann, O., Vanggaard Andersen, J, Guddal, E. & Roed, T. (1995). “Chemical structure of sterols that activate oocyte development”, *Nature* **374**, 559-562.

Byskov, A.G., Yding Andersen, C., Hossaini, A. & Guoliang, X. (1997). “Cumulus cells of oocyte-cumulus complexes secrete a meiosis-activating substance when stimulated with FSH”, *Mol. Reprod. Dev.* **46**, 296-305.

Byskov, A.G., Baltsen, M. & Yding Andersen, C. (1998). “Meiosis-activating sterols: background, discovery and possible use”, *J. Mol. Med.* **76**, 818-823.

Byskov, A.G., Yding Andersen, C., Leonardsen, L. & Baltsen, M. (1999). “Meiosis activating sterols (MAS) and fertility in mammals and man”, *J. Exp. Zool.* **285**, 237-242.

Caflisch, A., Miranker, A. & Karplus, M. (1993). “Multiple copy simultaneous search and construction of ligands in binding sites – application to inhibitors of HIV-1 aspartic proteinase”, *J. Med. Chem.* **36**, 2142-2167.

Clark, D.E., Frenkel, D., Levy, S.A., Li, J., Murray, C.W., Robson, B., Waszkowycz, B. & Westhead, D. R. (1995). “PRO_LIGAND: An approach to de novo molecular design. 1. Application of the design of organic molecules”, *J. Comput-aided Mol. Design* **9**, 13-32.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. & Barton, G.J. (1998). “JPred: a consensus secondary structure prediction server”, *Bioinformatics* **14**, 892-893.

Danziger, D.J. & Dean, P.M. (1989a). “Automated site-directed drug design: a general algorithm for knowledge acquisition about hydrogen-bonding regions at protein surfaces”, *Proc. Roy. Soc. ser. B* **236**, 101-114.

Danziger, D. J. & Dean, P.M. (1989b). “Automated site-directed drug design: the prediction and observation of ligand point positions at hydrogen-bonding regions on protein surfaces”, *Proc. Roy. Soc. ser. B* **236**, 115-124.

Dekel, N. (1995). “Molecular control of meiosis”, *Trends Endocrin. Metab.* **6**, 165-169.

Desvaux, H., Berthault, P., Birlirakis, N., Goldman, M. & Piotto, M. (1995). "Improved versions of off-resonance ROESY", *J. Magn. Reson., Ser. A* **113**, 47-52.

Dolle, R.E. & Kruse, L.I. (1986). "Improved preparation of (3β,5α,14α)-3-hydroxy-14-methylcholest-7-en-15-one. Synthesis of ergostenone and 20α-(hydroxymethyl)-pregnenone analogs", *J. Org. Chem.* **51**, 4047-4053.

Dolle, R.E., Schmidt, S.J., Eggleston, D. & Kruse, L.I. (1988). "Studies on the acid-catalyzed homonuclear steroidal diene isomerization", *J. Org. Chem.* **53**, 1563-1566.

Downs, S.M., Coleman, D.L., WardBailey, P.F. & Eppig, J.J. (1985). "Hypoxanthine is the principal inhibitor of murine oocyte maturation in a low molecular weight fraction of porcine follicular fluid", *Proc. Natl. Acad. Sci. U. S. A.* **82**, 454-458.

Duin, A. van. (1998). "Biogeochemical applications of a newly developed carbocation force field", PhD thesis, Delft University of Technology, The Netherlands.

Duisenberg, A.J.M. (1992). "Indexing in single-crystal diffractometry with an obstinate list of reflections", *J. Appl. Cryst*. **25**, 92-96.

Duisenberg, A.J.M. (1998). "Reflections on area detectors - *ab initio* calculation of single-crystal X-ray reflection contours", Thesis, Utrecht University, The Netherlands.

Eppig, J.J. & Downs, S.M. (1988). "Gonadotropin-induced murine oocyte maturation *in vivo* is not associated with decreased adenosine monophosphate in the oocyte-cumulus cell complex", *Gamete Res.* **20**, 125-131.

Færge, I., Grøndahl, C., Ottesen, J.L. & Hyttel, P. (2001). "Autoradiographic localization of specific binding of meiosis-activating sterol to cumulus-oocyte complexes from marmoset, cow, and mouse", *Biol. Reprod.* **64**, 527-536.

Ferguson, G., Siew, P.Y., Whalley, W.B. & Yeates, C.L. (1982). "22α,23α-Dibromo-12-methyl-18-nor-5α,17b-ergosta-8,11,13(14)-triene-3b-t-butyldimethyl silyl ether. C34H56Br2OSi.", *Cryst. Struct. Commun.* **11**, 775-780.

Friebolin, H. (1993). "Basic one- and two-dimensional NMR spectroscopy", 2nd Ed., VCH, Weinheim, Germany.

Goodford, P.J. (1985). "A computational procedure for determining energetically favourable binding sites on biologically important molecules", *J. Med. Chem.* **28**, 849-857.

Gonzalez, C., Rullmann, J.A.C., Bonvin, A.M.J.J., Boelens, R. & Kaptein, R. (1991). "Toward an NMR *R* factor", *J. Magn. Res.* **91**, 659-664.

Grøndahl, C., Lessl, M., Farge, I., Hegele-Hartung, C., Wassermann, K. & Ottesen, J.L. (2000). "Meiosis-activating sterol-mediated resumption of meiosis in mouse oocytes *in vitro* is influenced by protein synthesis inhibition and cholera toxin", *Biol. Reprod.* **62**, 775-780.

Grøndahl, C., Ottesen, J.L., Lessl, M., Faarup, P., Murray, A., Grønvald, F.C., Hegele-Hartung, C. & Ahnfelt-Rønne, I. (1998). "Meiosis-activating sterol promotes resumption of meiosis in mouse oocytes cultured *in vitro* in contrast to related oxysterols", *Biol. Reprod.* **58**, 1297-1302.

Grønvald, F.C., Faarup, P & Guddal, E. (1997). "Preparation of a cholestane derivative for regulating meiosis", *PCT. Int Appl.* WO 97/00883.

Guoliang, X., Byskov, A.G. & Yding Andersen, C. (1993). "Forskolin and the meiosis inducing substance synergistically initiate meiosis in fetal male germ cells", *Molec. Reprod. Dev*. **34**, 47-52.

Hegele-Hartung, C., Kuhnke, J., Lessl, M., Grøndahl, C., Ottesen, J., Beier, H.M., Eisner, S. & Eichenlaub-Ritter, U. (1999). "Nuclear and cytoplasm maturation of mouse oocytes after treatment with synthetic meiosis-activating sterol *in vitro*", *Biol. Reprod.* **61**, 1362-1372.

Honig B. & Nicholls, A. (1995). "Classical electrostatics in biology and chemistry", *Science* **268**, 1144-1149.

Hooft, R.W.W, Sander, C. & Vriend, G. (1996), "Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures", *Proteins* **26**, 363-376.

Jackson, S.M., Ericsson, J. & Edwards, P.A. (1997). "Signalling molecules derived from the cholesterol biosynthetic pathway", *Subcellular Biochemistry* **28** (Cholesterol: its functions and metabolism in biology and medicine), Ed. R. Bitman, Plenum Press, NY, USA, 1-21.

Janowski, B.A., Willy, P.J., Devi, T.R., Falck, J.R. & Mangelsdorf, D.J. (1996). "An oxysterol signalling pathway mediated by the nuclear receptor LXRα", *Nature* **383**, 728-731.

Jones, G., Willett, P. & Glen, R.C. (1995). "Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation", *J. Mol. Biol.* **245**, 43-53.

Jones, G., Willett, P., Glen, R.C., Leach, A.R. & Taylor, R. (1997). "Development and validation of a genetic algorithm for flexible docking", *J. Mol. Biol.* **267**, 727-748.

Kellogg, G.E., Semus, S.F. & Abraham, D. J. (1991). "HINT – a new method of empirical hydrophobic field calculation for CoMFA", *J. Comput-aided Mol. Design* **5**, 545-552.

Laskowski, R.A., Thornton, J.M, Humblet, C. & Singh, J. (1996). "X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins", *J. Mol. Biol.* **259**, 175-201.

Leonardsen, L., Strömstedt, M., Jacobsen, D., Kristensen, K.S., Baltsen, M., Andersen, C.Y. & Byskov, A.G. (2000a). "Effect of inhibition of sterol D14-reductase on accumulation of meiosis-activating sterol and meiotic resumption in cumulus-enclosed mouse oocytes *in vitro*", *J. Reprod. Fertil.* **118**, 171-179.

Leonardsen, L., Wiersma, A., Baltsen, M., Byskov, A.G. & Yding Andersen, C. (2000b). "Regulation of spontaneous and induced resumption of meiosis in mouse oocytes by different intracellular pathways", *J. Reprod. Fertil.* **120**, 377-383.

Lii, Y.-H. & Allinger, N.L. (1989). "Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics", *J. Am. Chem. Soc.* **111**, 8566-8575.

Liu, R.-M., Chillier, X.F.D., Kamalaprija, P., Burger, U. & Gålaçar, F.O. (1996). "Acid-catalysed backbone rearrangement of cholesta-6,8(14)-dienes", *Helv. Chim. Acta* **79**, 989-998.

Liu, Y.-Y., Collins, E.D., Norman, A.W. & Peleg, S. (1997). "Differential interaction of 1α,25-dihydroxyvitamin D_3 analogues and their 20-*epi* homologues with the vitamin D receptor", *J. Biol. Chem.* **272**, 3336-3345.

Liu, M. & Wang, S. (1999). "MCDOCK: a Monte Carlo simulation approach to the molecular docking problem", *J. Comput-aided Mol. Des.* **13**, 435-451.

Lu, Z., Guoliang, X., Byskov, A.G. & Yding Andersen, C. (2000). "Effects of amphotericin B and ketoconazole on mouse oocyte maturation: implications on the role of meiosis-activating sterol", *Mol. Cell. Endocrinol.* **164**, 191-196.

Majdic, G., Parvinen, M., Bellamine, A., Harwood Jr., H.J., Ku, W.W., Waterman, M.R. & Rozman, D. (2000). "Lanosterol 14α-demethylase (CYP51), NADPH-cytochrome P450 reductase and squalene synthase in spermatogenesis: late spermids of the rat express proteins needed to synthesize follicular fluid meiosis activating sterol", *J. Endocrin.* **166**, 463-474.

Matias, P.M., Donner, P., Coelho, R., Thomaz, M., Peixoto, C., Macedo, S., Otto, N., Joschko, S., Scholz, P., Wegg, A., Basler, S., Schafer, M., Egner, U. & Carrondo, M.A. (2000). "Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations", *J. Biol. Chem.* **275**, 26164-26171.

Midland, M.M., Plumet, J. & Okamura, W.H. (1993). “Studies of vitamin D (calciferol) and its analogs. Effect of C20 stereochemistry on the conformational profile of the side chains of vitamin D analogs”, *Bioorg. Med. Chem. Lett.* **3**, 1799-1804.

Miller, M.D., Kearsley, S.K., Underwood, D.J. & Sheridan, R.P. (1994). “FLOG: a system to select ‘quasi-flexible’ ligands complementary to a receptor of known three-dimensional structure”, *J. Comput-aided Mol. Design* **8**, 153-174.

Miranker, A. & Karplus, M. (1991). “Functionality maps of binding sites – a multiple copy simultaneous search method”, *Proteins: Struct. Funct. Genet.* **11**, 29-34.

Moor, R.M. & Warnes, G.M. (1979). “Regulation of meiosis in mammalian oocytes”, *Br. Med. Bull.* **35**, 99-103.

Mosquera, R.A., Rios, M.A. & Tovar, C.A. (1988). “Conformational analysis of 5Z- and 5E-vitamin D_3 dihydroderivatives by molecular mechanics”, *Theochem.* **45**, 125-133.

Nes, W. D., Wong, R.Y., Benson, M, Landrey, J.R. & Nes, W.R. (1984). “Rotational isomerism about the 17(20)-bond of steroids and euphoids as shown by the crystal structures of euphol and tirucallol”, *Proc. Natl. Acad. Sci. USA* **81**, 5896-5900.

Nicholls, A. & Honig, B. (1991). “A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation.”, *J. Comput. Chem.* **12**, 435-445.

Nolte, R.T., Wisely, G.B., Westin, S., Cobb, J.E., Lambert, M.H., Kurokawa, R., Rosenfeld, M.G., Willson, T.M., Glass, C.K. & Milburn, M. (1998). “Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma”, *Nature* **395**, 137-143.

Nonius (1998). Collect Software, Nonius, Delft, The Netherlands.

Norman, A.W., Adams, D., Collins, E.D., Okamura, W.H. & Fletterick, R.J. (1999a). “Three-dimensional model of the ligand binding domain of the nuclear receptor for 1α,25-dihydroxy-Vitamin-D_3”, *J. Cell. Biochem.* **74**, 323-333.

Norman, A.W., Song, X., Zanello, L., Bula, C. & Okamura, W.H. (1999b). “Rapid and genomic biological responses are mediated by different shapes of the agonist steroid hormone, 1α,25-$(OH)_2$-Vitamin-D_3”, *Steroids* **64**, 120-128.

Norman, A.W., Manchand, P.S., Uskokovic, M.R., Okamura, W.H., Takeuchi, J.A., Bishop, J.E., Hisatake, J.-I., Koeffler, H.P. & Peleg, S. (2000). “Characterization of a Novel Analogue of 1α,25-$(OH)_2$-Vitamin-D_3 with Two Side Chains: Interaction with Its Nuclear Receptor and Cellular Actions”, *J. Med. Chem.* **43**, 2719-2730.

Otwinowski, Z. & Minor, W. (1997). "Processing of x-ray diffraction data collected in oscillation mode", *Methods in Enzymology*, **276**: Macromolecular Crystallography, part A, C.W. Carter, Jr. and R.M. Sweet, Eds., Academic Press. 307-326.

Peleg, S., Sastry, M., Collins, E.D., Bishop, J.E. & Norman, A.W. (1995). “Distinct conformational changes induced by 20-epi analogues of 1a,25-dihydroxyvitamin D_3 are associated with enhanced activation of the vitamin D receptor”, *J. Biol. Chem.* **270**, 10551-10558.

Pitzer, K.S. & Gwinn, W.D. (1942). “Energy levels and thermodynamic functions for molecules with internal rotation. I. Rigid frame with attached tops”, *J. Chem. Phys.* **10**, 428-440.

Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). “Predicting receptor-ligand interactions by an incremental construction algorithm”, *J. Mol. Biol.* **261**, 470-489.

Renaud, J.P., Rochel, N., Ruff, M., Vivat, V., Chambon, P., Gronemeyer, H. & Moras, D. (1995). “Crystal structure of the RAR-gamma ligand-binding domain bound to all- trans retinoic acid”, *Nature* **378**, 681-689.

Rochel, N., Wurtz, J.M., Mitschler, A., Klaholz, B. & Moras, D. (2000). "The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand", *Mol. Cell* **5**, 173-179.

Ruan, B., Wilson, W.K. & Schroepfer Jr., G.J. (1998). "An alternative synthesis of 4,4-dimethyl-5a-cholesta-8,14,24-trien-3β-ol", and intermediate in sterol biosynthesis and a reported activator of meiosis and of nuclear orphan receptor LXRα", *Bioorg. Med. Chem. Lett.* **8**, 233-236.

Šali, A. & Blundell, T.L. (1993). "Comparative protein modelling by satisfaction of spatial restraints", *J. Mol. Biol.* **234**, 779-815.

Schroepfer Jr., G.J. (1982). "Sterol biosynthesis", *Ann. Rev. Biochem.* **51**, 555-585.

Schultz, R.M. (1988). "Regulatory functions of protein phosphorylation in meiotic maturation of mouse oocytes *in vitro*", *Prog. Clin. Biol. Res.* **267**, 137-151.

Sheldrick, G.M. (1997a). SHELXL97. Program for crystal structure refinement. University of Göttingen, Germany.

Sheldrick, G.M. (1997b). SHELXS97. Program for crystal structure solution. University of Göttingen, Germany.

Sluis, P. van der & Spek, A.L. (1990). "BYPASS: an effective method for the refinement of crystal structures containing disordered solvent regions", *Acta Cryst.* **A46**, 194-201.

Spek, A.L. (2001). PLATON. A multi-purpose crystallographic tool. Utrecht University, The Netherlands.

Stewart, J.J.P. (1990). *MOPAC 6.00*, QCPE 455.

Strömstedt, M., Waterman, M.R., Haugen, T.B., Taskén, K., Parvinen, M. & Rozman, D. (1998). "Elevated expression of lanosterol 14α-demethylase (CYP51) and the synthesis of oocyte meiosis-activating sterols in postmeiotic germ cells of male rats", *Endocrinology* **139**, 2314-2321.

Suwinska, K. & Kutner, A. (1996). "Crystal and molecular structure of 1,25-dihydroxycholecalciferol", *Acta Cryst. B* **52**, 550-554.

Swamy, N., Xu, W., Paz, N., Hsieh, J.-C., Haussler, M., Maalouf, G.J., Mohr, S.C. & Ray, R. (2000). "Molecular modelling, affinity labelling, and site-directed mutagenesis define the key points of interaction between the ligand-binding domain of the vitamin D nuclear receptor and 1α,25-dihydroxyvitamin D_3", *Biochemistry* **39**, 12162-12171.

Tanenbaum, D.M., Wang, Y., Williams, S.P. & Sigler, P.B. (1998). "Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains", *Proc. Natl. Acad. Sci. USA* **95**, 5998-6003.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic Acids Res.* **22**, 4673-80.

Tocchini-Valentini, G., Rochel, N., Wurtz, J.M., Mitschler, A. & Moras, D. (2001). "Crystal structures of the vitamin D receptor complexed to superagonist 20-epi ligands", *Proc. Natl. Acad. Sci. USA* **98**, 5491-5496.

Törnell, J., Billig, H. & Hillensjö, T. (1991). "Regulation of oocyte maturation by changes in ovarian levels of cyclic nucleotides", *Hum. Reprod.* **6**, 411-422.

Väisänen, S., Duchier, C. & Mäenpää, P.H. (1998)."Putative helices 3 and 5 of the human vitamin D_3 receptor are important for the binding of calcitriol", *FEBS Lett.* **440**, 203-207.

Väisänen, S., Duchier, C., Rouvinen, J. & Mäenpää, P.H. (1999a). "The Importance of the putative Helices 4 and 5 of human vitamin D_3 receptor for conformation and ligand binding", *Biochem. Biophys. Res. Commun.* **264**, 478-482.

Väisänen, S., Ryhänen, S., Saarela, J.T.A. & Mäenpää, P.H. (1999b). "Structure-function studies of new C-20 epimer pairs of vitamin D_3 analogs", *Eur. J. Biochem.* **261**, 706-713.

Verdonk, M.L., Cole, J.C. & Taylor, R. (1999). "SuperStar: a knowledge-based approach for identifying interaction sites in proteins", *J. Mol. Biol.* **289**, 1093-1108.

Verdonk, M.L., Cole, J.C., Watson, P., Gillet, V, & Willett, P. (2001). "SuperStar: improved knowledge-based interaction fields for protein binding sites", *J. Mol. Biol.* **307**, 841-859.

Wagner, R.L., Apriletti, J.W., McGrath, M.E., West, B.L., Baxter, J.D. & Fletterick, R.J. (1995). "A structural role for hormone in the thyroid hormone receptor", *Nature* **378**, 690-697.

Weinmann, R., Einspahr, H.M., Krystek, S.R., Jr., Sack, J,S., Salvati, M.E., Tokarski, J.S., Wang, C., Attar, R.M. (2001). "Crystallographic structure of the androgen receptor ligand binding domain and pharmacological applications", *PCT Int. Appl.* WO 01/27622.

Wenckens, M., Grønvald, F. & Hansen, J.B. (1998). "Synthesis of meiosis-activating sterols containing fluorine", *Acta Chem. Scand.* **52**, 503-507.

Wertz, D.H. & Allinger, N.L. (1979). "Conformation analysis-129. Heats of formation and thermodynamic parameters for hydrocarbons, calculated by the molecular mechanics method including the effects of molecular vibrations", *Tetrahedron* **35**, 3-12.

Wiersma, A., Hirsch, B., Tsafriri, A., Hanssen, R.G.J.M., Van de Kant, M., Kloosterboer, H.J., Conti, M. & Hsueh, A.J.W. (1998). "Phosphodiesterase 3 inhibitors suppress oocyte maturation and consequent pregnancy without affecting ovulation and cyclicity in rodents", *J. Clin. Invest.* **102**, 532-537.

Williams, S.P. & Sigler, P.B. (1998). "Atomic structure of progesterone complexed with its receptor", *Nature* **393**, 392-396.

Wilson, W.K. & Schroepfer Jr., G.J. (1988). "Acid-catalyzed isomerization of 7-dehydrocholesterol benzoate. A revised mechanism and an improved synthetic procedure", *J. Org. Chem.* **53**, 1713-1719.

Wilson, A.J.C. (1992) Ed. "International Tables for Crystallography", Volume C, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Wireko, F.C., Kellogg, G.E. & Abraham, D.J. (1991). "Allosteric modifiers of hemoglobin. 2. Crystallographically determined binding-sites and hydrophobic binding interaction analysis of novel hemoglobin oxygen effectors", *J. Med. Chem.* **34**, 758-767.

Xu, H.E., Lambert, M.H., Montana, V.G., Parks, D.J., Blanchard, S.G., Brown, P.J., Sternbach, D.D., Lehmann, J.M., Wisely, G.B., Willson, T.M., Kliewer, S.A. & Milburn, M.V. (1999). "Molecular recognition of fatty acids by peroxisome proliferator-activated receptors", *Mol. Cell* **3**, 397-403.

Yamada, S., Yamamoto, K., Masuno, H. & Ohta, M. (1998). "Conformation-function relationship of vitamin D: conformational analysis predicts potential side-chain structure", *J. Med. Chem.* **41**, 1467-1475.

Yamamoto, K., Sun, W.Y., Ohta, M., Hamada, K., DeLuca, H.F. & Yamada, S. (1996). "Conformationally restricted analogs of 1α,25-dihydroxyvitamin D_3 and its 20-epimer: compounds for study of the three-dimensional structure of Vitamin D responsible for binding to the receptor", *J. Med. Chem.* **39**, 2727-2737.

Yamamoto, K., Ooizumi, H., Umesono, K., Verstuyf, A., Bouillon, R., DeLuca, H.F., Shinki, T., Suda, T. & Yamada, S. (1999). "Three-dimensional structure-function relationship of vitamin D: side chain location and various activities", *Bioorg. Med. Chem. Lett.* **9**, 1041-1046

Yamamoto, K., Masuno, H., Choi, M., Nakashima, K., Taga, T., Ooizumi, H., Umesono, K., Sicinska, W., VanHooke, J., DeLuca, H.F. & Yamada, S. (2000). "Three-dimensonal modeling of and ligand docking to vitamin D receptor ligand binding domain", *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1467-1472.

Yding Andersen, C., Baltsen, M. & Byskov, A.G. (1999). "Gonadotropin-induced resumption of oocyte meiosis and meiosis-activating sterols", *Curr. Top. Dev. Biol.* **41**, 163-185.

Yoshida, Y., Yamashita, C., Noshiro, M., Fukuda, M., & Aoyama, Y. (1996). "Sterol 14-demethylase P450 activity expressed in rat gonads: contribution to the formation of mammalian meiosis-activating sterol", *Biochem. Biophys. Res. Commun.* **223**, 534-538.

Summary

Meiosis Activating Sterols (MAS) are key regulatory factors in the meiotic cell cycle. Two compounds in this family, Follicular Fluid-MAS (FF-MAS) and Testicular-MAS (T-MAS), have been isolated and identified. FF-MAS was found to initiate the resumption of the arrested development of immature egg cells (*i.e.* primary oocytes), whereas T-MAS regulates the continuous generation of sperm cells. The biological function of these molecules makes them promising leads in the development of novel contraceptive agents as well as drugs that remedy fertility problems that are related to the processes in which the MAS compounds are involved. The work described in this thesis aims to explore the structural and physicochemical properties of MAS, in particular of FF-MAS and derived compounds and the implications of these properties on the MAS activity.

The syntheses of MAS analogs often yield a mixture of double bond isomers. The relative abundancies of isomer products vary and are expected to depend on the stabilities of the intermediate isomers that are formed during the reactions. In **chapter 2**, the stabilities of these isomers are evaluated using molecular mechanics. Heats of formation were calculated for series of isomers that could in principle be formed during synthesis. It was found that the relative synthetic yields of double bond isomers of MAS analogs correspond well with the calculated heat of formation differences.

The side chain conformations of several active and inactive synthetic MAS analogs were studied in the solid phase and in solution. Five crystal structures of FF-MAS derivatives were solved. For three of these, NOESY and ROESY NMR experiments were performed. This work is described in **chapter 3**. The crystal structures reveal that the conformations of side chains of MAS analogs are analogous to those of other sterols with similar side chain structure retrieved from the CSD. After superposition of the skeletal carbon atoms of the molecules in the X-ray structures determined by us, the positions of the side chains of an active and an inactive compound are found to overlap, whereas the side chains of two other active compounds are located in different regions in space, respectively. It is clear that conclusions on the active conformation cannot be drawn from the crystal structures. The solution conformations of the side chains of three compounds, two of which are active and one is inactive, are similar to those found in the corresponding crystal structures. For one of the active compounds, two conformations appear to exist in solution. The less abundant conformation corresponds to the conformation found for the other active compound and it is therefore a candidate for the active conformation. Aromatic C ring analogs of FF-MAS have also been shown to be active. The side chain in these compounds is directed towards another region of space with respect to the skeleton. This conformation may also be able to induce FF-MAS activity. The

extended conformation of MAS compounds, with natural configuration at C20, as found in crystal structures and in solution is probably not the active conformation.
The crystal structure conformations of skeletons of different MAS double bond isomers are very similar. A conformational change induced by the difference in double bond positions is therefore not likely to be the determining factor for activity. Evaluation of the electrostatic potential (ESP) on the solvent accessible surfaces of several FF-MAS isomers show a distinct local negative ESP region near the double bond system, as shown in **chapter 4**. The average potential of the local ESP and the distance of the center of this region to the hydroxyl oxygen atom required for activity appear to be to –9 kJ/mol and 8 Å, respectively. In conclusion, activity differences of sterol double bond isomers appear to be related to electrostatic effects, rather than steric effects.

The analogy of the side chain structure-activity relationships of MAS compounds to calcitriol analogs was exploited by building a homology model of the Vitamin D Receptor (VDR). Possible binding sites for calcitriol and its analogs were analyzed using SuperStar, a program that predicts favorable sites of interaction around molecules. SuperStar initially used intermolecular interaction data derived from the Cambridge Structural Database (CSD) as knowledge base. For the prediction of hot-spots in a protein binding pocket, protein data bank (PDB) source data are considered to represent the protein-ligand binding properties better than CSD data.

The implementation of PDB source data and the comparison of PDB and CSD source data are discussed in **chapter 5**. It was found that non-bonded interaction geometries in X-ray structures of low molecular weight compounds (CSD) are generally very similar to those observed in protein ligand interfaces (PDB). However, hydrophobic contacts are more likely to occur in the PDB crystal structures compared with those in crystal structures from the CSD, whereas hydrophilic contacts are more likely to occur in CSD crystal structures. Non-bonded intermolecular interactions from the PDB are therefore more appropriate as source data for the prediction of favorable interaction sites in hydrophobic binding pockets of a protein. Another advantage of PDB source data is that the uncertainty of the atomic positions of the protein crystal structure model is inherently included in the source data. In some cases, however, it may be worthwhile to use CSD based data, because CSD structures often contain the positions of the hydrogen atoms, which is not the case in for PDB derived data. An example is the prediction of hot-spots near residues with unknown protonation states, such as histidines and carboxylates.

The construction of a homology model of the VDR and its use in the prediction of the binding modes of calcitriol and several of its analogs is discussed in **chapter 6**. The position of calcitriol and two 20-*epi* analogs as found in the X-ray structure of the VDR-calcitriol complex, which was published after the homology model was built, could be well reproduced using

SuperStar maps of the binding pocket of the VDR homology model and mutation studies on residues close to the binding pocket. The orientation of the side chain of calcitriol bound to the VDR corresponds to that of the aromatic C ring structure retrieved from the CSD, as described in chapter 3. It can be expected that FF-MAS binds to its receptor in a similar way as calcitriol, given the analogy in side-chain structure and the resulting activity found for analogs of both compounds. As such, the bound conformation of calcitriol provides a second indication that the side-chain conformation of the aromatic C ring structure may be able to bind to the MAS receptor. The conformation exhibited by bound calcitriol and the rigidified aromatic C ring structure is accessible to FF-MAS and its 22-oxo-20-epimer.

From the work described in this thesis, a pharmacophoric model for the MAS compounds starts to emerge. Apparently, a rod-shaped molecule is needed, with a 3β-OH group attached to one end and an unpolar side chain attached to the other. A negative electrostatic potential on the rod-shaped skeleton is required at a distance of 8 Å from the hydroxyl group and at 5 Å from the start of the aliphatic chain. Two conformations of this side chain are possibly responsible for the activity of the MAS compounds. Evidence for one of these was derived from studies of the conformation of MAS compounds in the solid state and in solution. The other conformation is implicated by the conformation found for the bound state of calcitriol and from the crystal structure of an aromatic ring structure. Thus, it appears that the conformation of MAS compounds with natural side-chain stereochemistry found in crystal structures is not responsible for activity. Rather, the studies presented in this thesis indicate that either of the conformations discussed above, or perhaps both of them, are able to induce MAS action.

Samenvatting

Meiose activerende sterolen (MAS) reguleren de celcyclus gedurende meiose. Tot dusver zijn twee verbinding geïsoleerd en gekarakteriseerd. Een ervan is afkomstig uit de follikel (FF-MAS), de ander uit de testes van de stier (T-MAS). Uit onderzoek blijkt dat FF-MAS de ontwikkeling van nog niet volgroeide eicellen initieert en dat T-MAS de continue spermatogenese reguleert. Vanwege hun biologische functie zijn deze twee verbindingen veelbelovende leads voor de ontwikkeling van nieuwe typen anticonceptiva en farmaceutica die vruchtbaarheidsproblemen verhelpen die gerelateerd zijn aan de processen waarin de MAS verbindingen een rol spelen. Het onderzoek beschreven in dit proefschrift heeft als doel de structurele en fysisch-chemische eigenschappen van MAS, in het bijzonder FF-MAS, te bestuderen en de hun implicaties op activiteit in kaart te brengen.

De syntheses van MAS en analoge verbindingen resulteren vaak in een mengsel van isomeren met dubbele bindingen op verschillende posities. De relatieve opbrengsten van de isomere producten verschillen van reactie tot reactie en zijn naar verwachting gerelateerd aan de stabiliteiten van de intermediaire isomeren die gevormd worden tijdens de reacties. In **hoofdstuk 2** wordt de berekening van de stabiliteiten van deze intermediairen met behulp van moleculaire mechanica beschreven. Voor een reeks van isomeren, welke in principe gevormd kunnen worden tijdens de reacties, zijn vormingsenthalpieën berekend. De relatieve opbrengsten van de isomeren blijken sterk gerelateerd te zijn aan de verschillen in berekende vormingswarmten.

De zijstaartconformaties van verscheidene actieve en inactieve gesynthetiseerde MAS analoga zijn bestudeerd, zowel in de vaste toestand als in oplossing. Vijf kristalstructuren van MAS derivaten zijn bepaald. Voor drie van deze verbindingen zijn NOESY en ROESY NMR experimenten uitgevoerd. Dit werk is beschreven in **hoofdstuk 3**. Uit vergelijking van de skeletconformaties in de kristalstructuren van de vijf MAS analoga met kristalstructuren van sterolen, opgenomen in de Cambridge Structural Database (CSD), blijkt dat deze conformaties voor isomere sterolen sterk overeenkomen. De koolstofatomen in de skeletten van de vijf MAS analoga, waarvan de structuren zijn opgehelderd, zijn gesuperponeerd. Uit deze superpositie blijkt dat de oriëntatie van de zijstaarten ten opzichte van het skelet van één van de actieve verbindingen overeenkomt met één van de inactieve analoga. De zijstaartposities van de andere actieve verbindingen verschillen van deze oriëntatie en van die van elkaar. Het moge duidelijk zijn dat uit de kristalstructuuranalyses geen conclusies getrokken kunnen worden betreffende de actieve conformatie van de MAS verbindingen. De zijstaartconformaties zoals bepaald uit de NMR analyses van drie van de verbindingen, waarvan er één inactief is en twee actief zijn, komen overeen met de conformaties in de kristalstructuren. Één van de actieve

verbindingen komt waarschijnlijk voor in twee voorkeursconformaties, waarvan één overeenkomt met de conformatie gevonden voor het andere actieve molecuul. Deze conformatie is een mogelijke kandidaat voor de actieve conformatie. De positie van de zijstaarten van verbindingen met een aromatische C-ring, waaronder ook actieve verbindingen zijn gevonden, verschillen van deze conformatie. De zijstaartconformatie van deze verbindingen kan ook verantwoordelijk zijn voor de activiteit van de MAS verbindingen. Het is onwaarschijnlijk dat de gestrekte conformatie van MAS verbindingen met een natuurlijke configuratie van C20, zoals gevonden in hun respectievelijke kristalstructuren, de actieve conformatie is.

De skeletconformaties van de isomeren van MAS, met verschillende posities van de dubbele bindingen, verschillen weinig. Daarom is het niet waarschijnlijk dat deze conformatie de bepalende factor is voor activiteit. Analyse van de electrostatische potentiaal (ESP) op het oplosmiddel-toegankelijke oppervlak van verscheidene FF-MAS isomeren gaf aan dat een lokaal negatieve ESP aanwezig is dichtbij de dubbele bindingen, zoals beschreven in **hoofdstuk 4**. Een gemiddelde potentiaal van de lokale ESP van -9 kJ/mol en een afstand van het centrale punt van dit gebied tot het zuurstofatoom van de hydroxyl groep van ongeveer 8 Å zijn nodig voor activiteit. Hieruit blijkt dat verschillen in activiteit van MAS isomeren met verschillende posities van de dubbele bindingen meer gerelateerd zijn aan electrostatische dan aan sterische effecten.

De analogie in structuur-activiteitsrelaties van de zijstaartketens van de MAS verbindingen enerzijds en die van calcitriol en analoga daarvan anderzijds was de aanleiding om een homologiemodel te bouwen van de vitamine D receptor (VDR), de endogene receptor van calcitriol. Om de bindingspositie van calcitriol en analoga te vinden in het homologiemodel, is gebruik gemaakt van SuperStar, een programma dat gunstige posities voor interacties probeert te vinden in de buurt van moleculen. SuperStar gebruikte in eerste instantie intermoleculaire interactiegegevens uit de CSD als basis. Voor het voorspellen van gunstige interactieposities in een eiwitbindingsplaats wordt het gebruik van interactiegegevens uit de Protein Data Bank (PDB) geprefereerd, omdat verwacht wordt dat deze de eiwit-ligand interacties beter kunnen representeren.

De implementatie van gegevens afgeleid uit de PDB en de vergelijking tussen PDB en CSD brongegevens worden bediscussiëerd in **hoofdstuk 5**. Uit deze vergelijking bleek dat de interactie geometriëen in PDB en CSD structuren over het algemeen goed overeenkomen. De relatieve frequenties van de verschillende typen interacties verschillen daarentegen. Hydrofobe contacten zijn waarschijnlijker in de PDB, terwijl hydrofiele contacten waarschijnlijker zijn in de CSD. Intermoleculaire interacties afgeleid uit de PDB zijn daarom beter geschikt voor het voorspellen van gunstige interacties in hydrofobe bindingsplaatsen. Een tweede voordeel van PDB brondata is dat de onzekerheid in de atoomposities in kristalstructuren van eiwitten impliciet aanwezig is in de gegevens. In sommige gevallen is het echter de moeite waard om CSD brondata

te gebruiken, omdat in CSD structuren de posities van de waterstofatomen vaak bepaald zijn, wat niet het geval is in PDB structuren. Een voorbeeld is het voorspellen van gunstige interactieposities in de buurt van residuen met onbekende protonatietoestand, zoals histidine en carboxylaten.

De constructie van een homologie model van de VDR en het gebruik ervan voor de voorspelling van de bindingsmotieven van calcitriol en enkele analoga is beschreven in **hoofdstuk 6**. De positie van calcitriol and twee 20-epi homologen in de kristalstructuren van de betreffende complexen, gepubliceerd na de constructie van ons homologiemodel, kon nauw gereproduceerd worden met behulp van de SuperStar berekeningen aan de ligandbindingsplaats in de VDR en mutatiestudies van aminozuren rond dit gebied. De oriëntatie van de zijstaart van gebonden calcitriol komt overeen met de conformatie van de zijstaart van de aromatische C-ring structuur waarvan een kristalstructuur aanwezig is in de CSD, zoals beschreven in hoofdstuk 3. Gegeven de analogie in zijstaartstructuur en de resulterende activiteit van MAS en calcitriol verbindingen kan verondersteld worden dat FF-MAS aan zijn endogene receptor bindt op gelijke manier als calcitrol. In die hoedanigheid is de gebonden conformatie van calcitriol een tweede indicatie dat de zijstaartconformatie van de aromatische C-ring structuur mogelijk de actieve conformatie is. De conformatie van gebonden calcitriol en de gerigidiseerde aromatische C-ring structuur is energetisch toegankelijk voor FF-MAS en de corresponderende 22-oxo-20-epimeer.

Uit het werk beschreven in dit proefschrift begint een farmacofoor model van de MAS verbindingen uit te kristalliseren. Een cylindervormig molecuul is nodig als uitgangspunt, met een 3β-OH groep en een alifatische zijstaart gebonden de twee uiteinden. Een negatieve elektrostatische potentiaal op een afstand van 8 Å van de hydroxyl groep en 5 Å van het begin van de zijstaart is een tweede voorwaarde. Twee conformaties van de zijstaart kunnen verantwoordelijk zijn voor de acitiviteit. Aanwijzingen voor één hiervan zijn afgeleid uit studies van MAS verbindingen in vaste vorm en in oplossing. De kristalstructuurconformatie van calcitriol gebonden aan de VDR en die van een aromatische C-ring structuur impliceren de tweede conformatie. Het is waarschijnlijk dat de zijstaartconformatie van MAS verbindingen met natuurlijke zijstaartconfiguratie niet verantwoordelijk is voor de activiteit. Het onderzoek beschreven in dit proefschrift duidt er juist op dat één van de zijstaartconformaties hierboven genoemd, of wellicht beide, MAS functionaliteit kunnen induceren.

Curriculum Vitae

De schrijver van dit proefschrift werd geboren op 3 april 1972 te Ambon (Indonesië). Na het behalen van de het VWO diploma aan het Revius lyceum te Doorn in 1990, begon hij aan de studie scheikunde aan de Rijksuniversiteit Utrecht, later hernoemd tot Universiteit Utrecht. Tijdens de studie is onderzoek verricht aan ceramide monolagen bij de vakgroep Biochemie van Membranen, aan vloeibaar-kristallijne eigenschappen van cyclohexylidenen bij de vakgroep Fysisch Organische Chemie en aan peptoïden bij de vakgroep Medicinal Chemistry aan de Faculteit Farmacie. In januari 1997 werd de studie afgerond met als specialisatie kristal- en structuurchemie. In februari 1997 begon hij als Assistent in Opleiding aan de vakgroep Kristal- & Structuurchemie van de Faculteit Scheikunde aan de Universiteit Utrecht onder begeleiding van Prof.dr. Jan Kroon (promotor), Dr. Huub Kooijman en Dr. Jan Kelder (co-promotores). Een deel van de promotietijd werd doorgebracht in Cambridge (UK) onder begeleiding van Dr. Marcel Verdonk, alwaar de auteur meewerkte aan een methode om eiwit-ligand interacties te beschrijven. Een ander deel van het werk werd uitgevoerd binnen de Molecular Design & Informatics groep binnen Organon N.V. te Oss, onder begeleiding van Dr. Jan Kelder en Dr. Milou Kouwijzer. De promotie is afgerond onder begeleiding van Prof.dr. Ton Spek. Het onderzoek uitgevoerd in het kader van de promotie is gepresenteerd op verscheidene congressen in Lissabon, Glasgow, Kopenhagen en op een summer school Erice (Sicilië, Italïe).

Publications

D.R. Boer, H. Kooijman, J. van der Louw, M. Groen, J. Kelder & J. Kroon. “Comparison of crystal structure conformations of five follicle fluid meiosis activating sterol related active and inactive compounds”, accepted for publication in *Acta Cryst. C*.

D.R. Boer, H. Kooijman, J. van der Louw, M. Groen, J. Kelder & J. Kroon. “Relation between the molecular electrostatic potential and activity of some FF-MAS related sterols”, accepted for publication in *Bioorg. & Med. Chem*.

D.R. Boer, J. Kroon, J.C. Cole, B. Smith & M.L. Verdonk. “SuperStar: Comparison of CSD and PDB-based interaction fields as a basis for the prediction of protein-ligand interactions”, accepted for publication in *J. Mol. Biol.*

D.R. Boer, H. Kooijman, J. van der Louw, M. Groen, J. Kelder & J. Kroon (2000). “Calculated heats of formation of sterol diene isomers compared with synthetic yields of isomerisation reactions of $\Delta^{5,7}$ sterols”, *J. Chem. Soc. Perkin Trans. 2*, pp. 1701-1704 (2000).

E. ten Grotenhuis, R.A. Demel, M. Ponec, D.R. Boer, J.C. van Miltenburg & J.A. Bouwstra. (1996). “Phase behavior of stratum corneum lipids in mixed Langmuir-Blodgett monolayers”, *Biophys. J.* **71**, 1389-1399.

Dankwoord

Mijn dank gaat allereerst uit naar mijn helaas veel te vroeg overleden promotor, Jan Kroon. Ondanks het feit dat je dit niet lezen zal, wil ik me toch in de tweede persoon tot je richten. Ik heb bijzonder veel bewondering gehad voor de manier waarop je je hebt ingezet voor de mensen in je groep. Daarnaast zorgde je inspirerende manier van besturen voor een bijzonder prettige werksfeer, waarin humor een belangrijke plaats innam, zonder dat dit ooit ten koste ging van de wetenschappelijke inhoud en het niveau van het onderzoek. Op wetenschappelijk vlak waardeer ik je inbreng zeker ook, tenslotte was dit boekje er niet gekomen zonder jou. Daarnaast heb je me de mogelijkheid geboden om ook buiten de vakgroep rond te kijken, onder andere in Cambridge.

Alle bovengenoemde uitspraken gelden evenzeer voor mijn co-promotor, Huub. Je positieve instelling en het feit dat je altijd paraat stond om de dagelijkse wetenschappelijke beslommeringen te bespreken waardeer ik zeer. Je tips gaven vaak waardevolle nieuwe gezichtspunten van waaruit het onderzoek verder uitgebouwd kon worden. Helaas is het pillion-wezen niet erg van de grond gekomen. Ton Spek wil ik bedanken voor zijn slagvaardige optreden na de leemte die het overlijden van Jan achterliet en het overnemen van de promotie zelf. Niet in de laatste plaats gaat mijn dank uit naar Marjan Hogenbirk, op organisatorisch en persoonlijk vlak een belangrijke spil in het vakgroepsgebeuren.

Ook Jan Kelder, Marinus Groen, Milou Kouwijzer en Jaap van der Louw wil ik bedanken, allereerst voor de geboden stoffen, maar vooral voor hun inzet en meedenken gedurende de hele promotietijd. Ten minste twee hoofdstukken zijn ontstaan uit ideeën door jullie aangedragen. Daarnaast is het werk voor een hoofdstuk grotendeels uitgevoerd onder jullie kundige supervisie. De laatste twee jaar hebben jullie de werkbesprekingen en het reviseren van teksten voornamelijk tussen de eigenlijke werkzaamheden door ingepast, iets wat ik zeer waardeer.

Their are several people from the Cambridge Crystallographic Data Centre I’d like to thank. First of all Marcel Verdonk, for his supervision during the two-month stay on British territory. Both in the scientific sense as well as on a personal level I enjoyed an inspirational collaboration, which has resulted in the achievement of all (three) goals I had planned before coming over. I would also like to thank Steven Salisbury and Robin Taylor for making the project possible. The support and company of Andreas Bergner, Jason Cole, Barry Smith, Ian Bruno, Paul Watson, Vanessa Williams, Gwenda Kyd, Karen Lipscomb and Claire Laskey were greatly appreciated and made my stay a pleasant one.

Mede dankzij Jan, maar vooral ook door de mensen zelf, is de Kristal & Structuurchemie groep een zeer prettige groep geweest om in te werken. Ik heb veel plezier beleefd aan de ontspannen manier van leven van Jeroen, Jean,

Andrew, Ton L. en Raimond, de bisous van Anne, het oriëntatievermogen van Ton L., de pauzes bij Arie, het enthousiasme van Toine, het Nederlands van de Martins, Stephane en Sjors, het geblaf van Ab, de overtuigende ideeën van Piet, de nog overtuigender ideeën van Bogos en Martin L., de grappen van Martin W., het winnen van een squash partij van Stephane, het veel vaker verliezen van een squash partij van Stephane, het getover en de eerlijkheid van Barend, de borrels met ‘de harde kern’, het organisatietalent en de barbecues van Marjan, de wandelingen met Clasien, het shoppen met Carien, de pogingen tot het starten van roddels met Clasien (helaas nooit gelukt), de motor van Huub (heb ´m vaak zien staan), de congressen in Lissabon en Glasgow, de ring in Erice, de hulp van Bouke, Loes, Toine, Ton Spek, Arie en Jan Kanters, de onverwachte kanten van Lucy, het Spaans met Sjors en Martin, de tennispartijen met de fam. Kroon en consorten, de expertise van Erik, de salsa van Dianne, de lach van Allison enz.enz.enz.

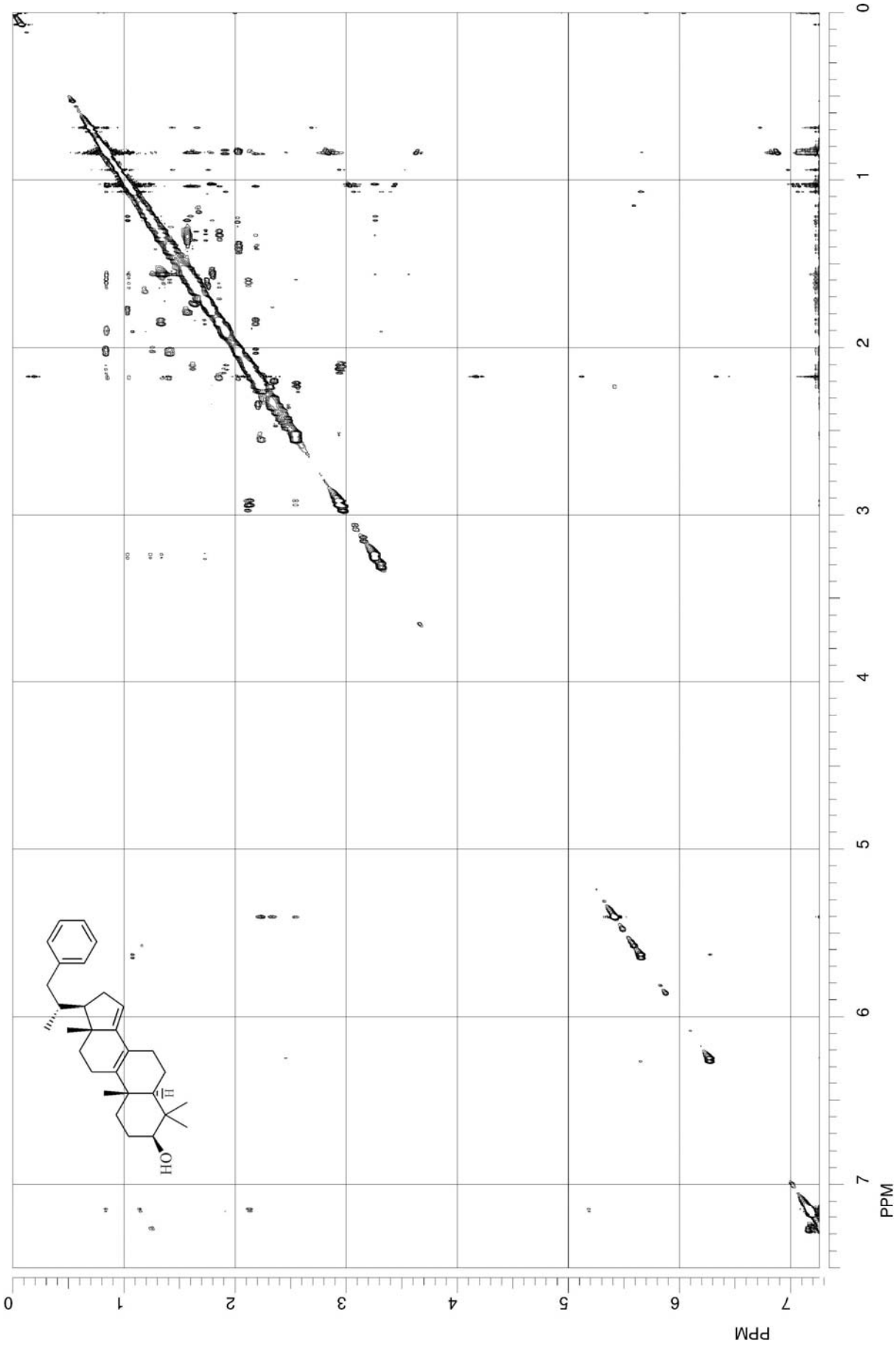
Buiten de vakgroep lopen ook goede wetenschappers en leuke mensen rond, waaronder de BOC’ers en de NMR’ers. Zonder iedereen bij naam te noemen, zeer veel dank voor suggesties, discussies en gezelschap!

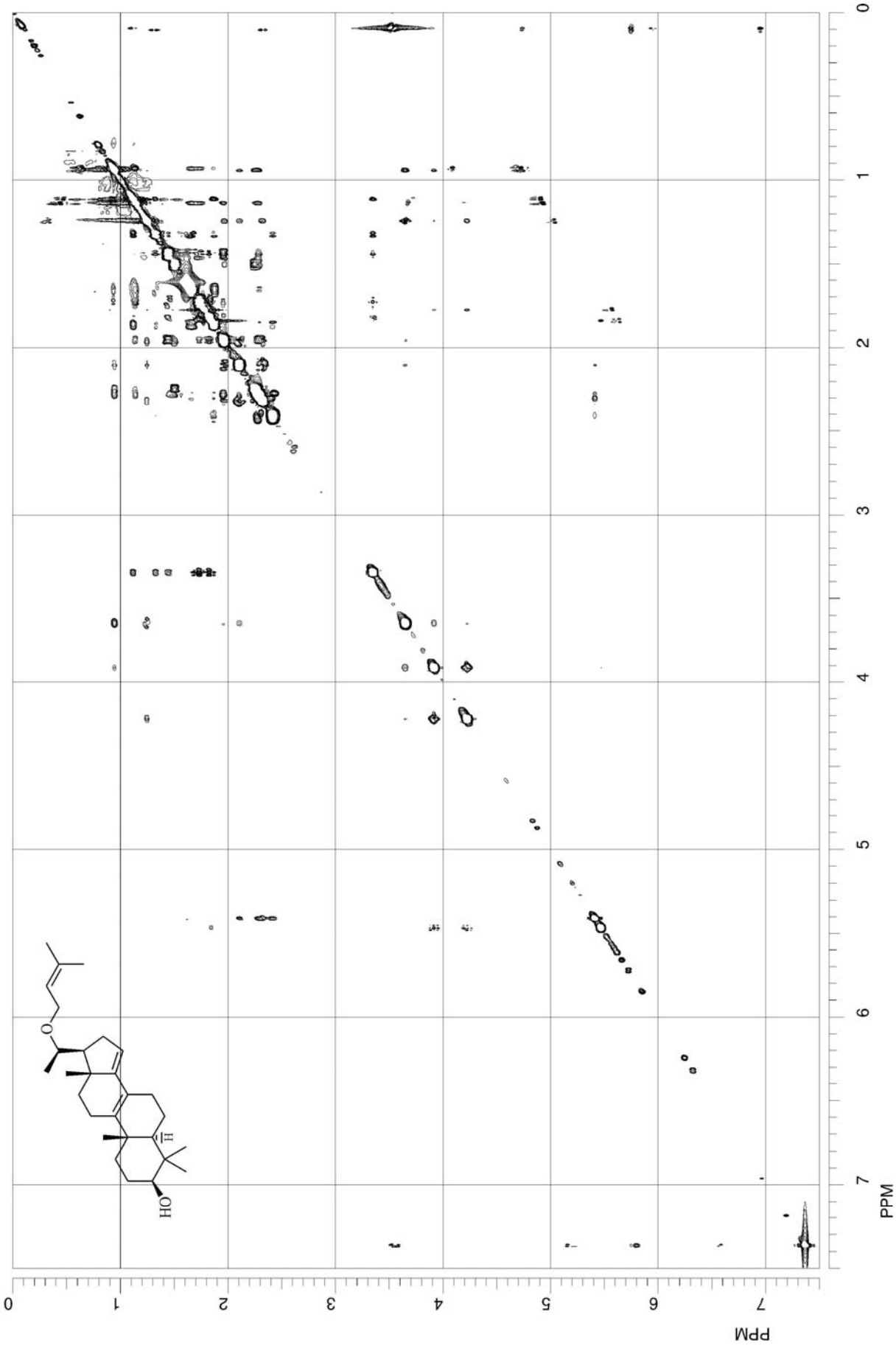
Het leven buiten de wetenschap wil ook nog wel eens aangenaam zijn, in het bijzonder in gezelschap van mensen die een koffiekaatje effect hebben, in de zin dat de scherpe kantjes bijgeslepen worden. Stephan, Brenda, Lineke, Arein, Anne, Lieke, Ben, Lysette, Geert & Annemiek, Felix & Sandy, Bart, Alwin & Simone, Nanet, Elien, Jacco, Eric, Ivo en last-but-not-least Anna, wil ik bij name noemen.

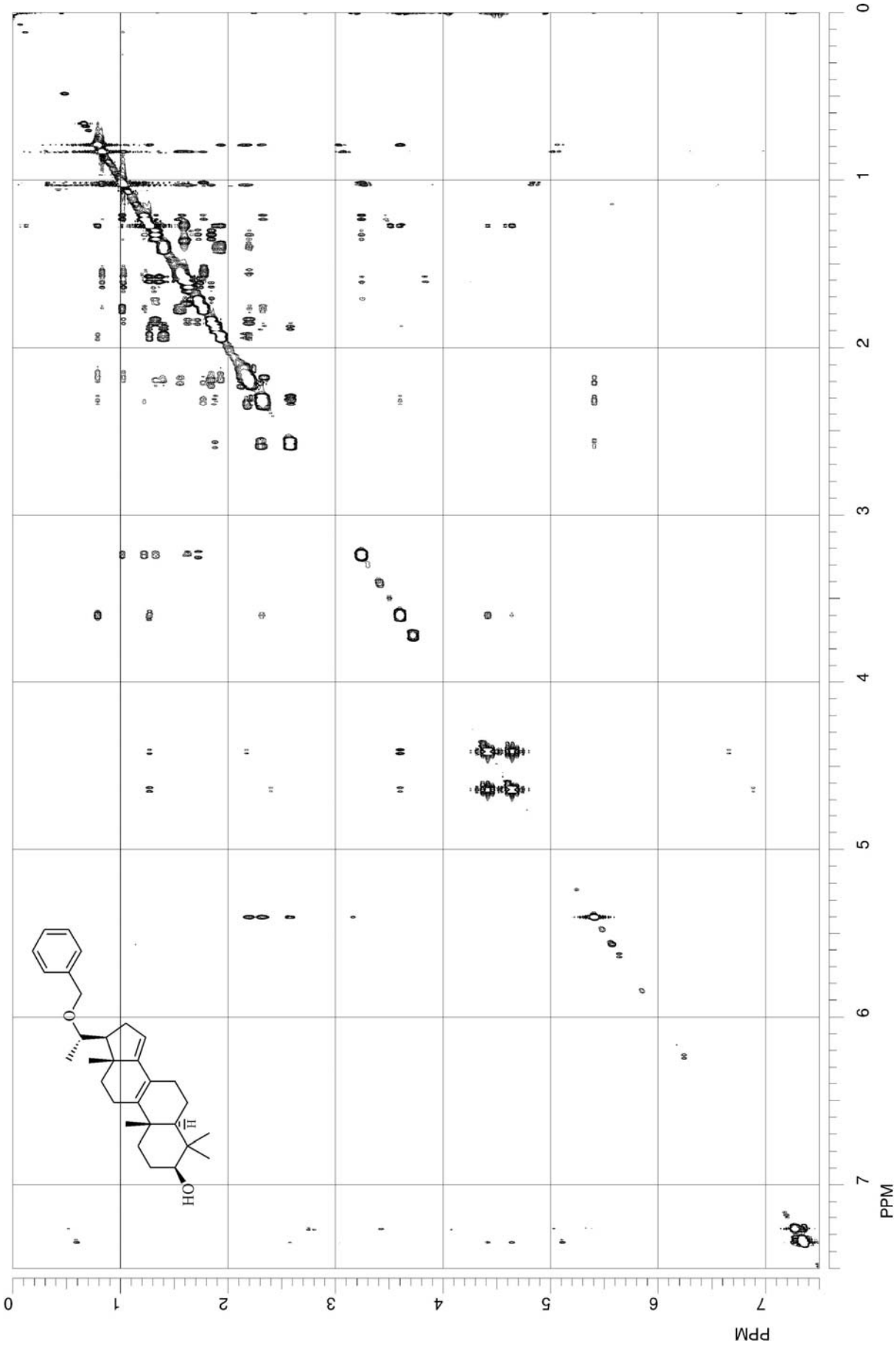
Op de laatste plaats bedank ik mijn ouders en familie voor hun onvoorwaardelijke steun. Het is geruststellend om te weten dat er altijd een plek is om op terug te vallen en ik prijs mij gelukkig met het feit dat ik die heb.

Appendix

NOESY and ROESY NMR spectra of three FF-MAS analogs.


PPM
0
1
2
3
4
5
6
7
HO
H


PPM
PPM
HO
H
O


PPM
PPM
HO