Eur. J. Med. Chem. 35 (2000) 787−796

S0223523400001689/REV

**Review**

# Application of (quantitative) structure–activity relationships to progestagens: from serendipity to structure-based design

Roberta Bursi[a*], Marinus B. Groen[b]

[a]Molecular Design & Informatics Department, N.V. Organon, P.O. Box 20, 5340 BH Oss, The Netherlands
[b]Medicinal Chemistry Department, N.V. Organon, P.O. Box 20, 5340 BH Oss, The Netherlands

Received 11 May 2000

**Abstract** – Progestagens are drugs, which are widely used in hormonal contraception and in hormone-replacement therapy. Since the natural hormone, progesterone, lacks oral activity, much effort has been devoted to finding analogues with improved oral activity and, preferably, higher potency and selectivity. A crystal structure of the hormone binding domain (HBD) region of the progesterone receptor (PR) could only be obtained recently. For more than forty years the process of designing new progestagens could therefore only be guided by the knowledge of the structure of the ligand and its corresponding in vitro/in vivo activities. While in early days chemical intuition and simple statistics (structure–activity relationship – SAR) were leading the drug design process, in later days more complex statistics and visualization tools have become routinely part of quantitative structure–activity relationship (QSAR) studies. The present review aims to provide a general overview of the strategies, efforts and achievements of synthetic and computational chemists in more than forty years of development of progestagens.


**progestagens / contraception / hormone replacement therapy / quantitative structure–activity relationship / hormone binding domain**

## 1. Historical introduction

Early in the 20th century it was found that the corpus luteum in mammals is required for the nidation of fertilized eggs and for maintenance of pregnancy. The compound responsible for this effect was found to be a steroid hormone for which, in 1935, the now generally accepted name progesterone (**1**) was proposed. Much interest in progesterone was generated by its ability to inhibit ovulation in rats, rabbits and guinea-pigs. This observation ultimately led to the discovery of the contraceptive pill ('The Pill'), which is generally attributed to the American biologist Pincus (1903–1967). Since progesterone itself has poor drug properties, much research has been devoted to the discovery of progesterone mimics (progestagens, alternatively called progestins or progestogens) with improved oral bioavailibility and potency. Over the years several highly orally active progestagens have been discovered, leading to widespread use of these hormones in contraception, but also in HRT (hormone replacement therapy), in treatment of certain cancers, in gynaecological disorders, etc.

Because of their limited amount of flexibility, progestagens, as steroids in general, have always represented an optimal target for structural activity relationship studies. It is therefore not surprising that the rational design of these compounds has evolved from sheer trial and error to various levels of sophistication. In this paper the various SAR approaches which have been applied to progestagens are reviewed.

## 2. From the origin till the end of the eighties

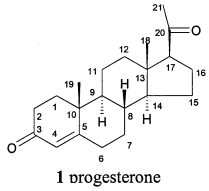
### *2.1. Laying the foundations*

While progesterone originally was a scarce and very expensive compound, it became readily available in the 1940s after the pioneering work by Marker on the production of steroid hormones from sapogenins. A serious drawback of the natural hormone, however, is its

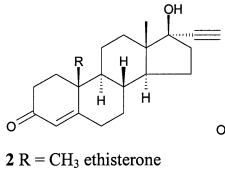
* Correspondence and reprints:
r.buma@organon.oss.akzonobel.nl

poor bioavailibility on oral administration due to rapid metabolism in the liver: over 99% of the orally administered dose is metabolized in the liver before it reaches the general circulation. The primary goal of the research on new progestagens was therefore to design orally active compounds. Prior to the development of in vitro assays in the 1970s the design of new progestagens was guided by in vivo progestagenic activity. Many different tests have been employed for this purpose, which often makes direct comparison of the potency of compounds tested by different research groups difficult. A reliable and widely accepted test for progestagenic activity is the McPhail test in oestrogen primed immature rabbits. In this test the proliferation of the endometrium is checked upon progestagenic treatment. After a few days of such treatment the differentiation of the endometrium is analysed by autopsy and scored from 0–4, depending on the strength of the effect (progestagenic potency) observed.

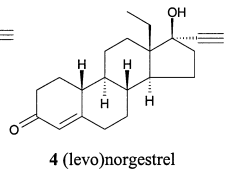
Early research benefited from the application of new synthetic methods in steroid chemistry as well as from insight in the metabolism of progesterone. The first factor is responsible for the serendipitous discovery of ethisterone (**2**) in 1938 by Inhoffen [1]. This compound unexpectedly turned out to possess oral progestagenic activity, but its practical use was limited due to its androgenic activity. Nevertheless, the finding that the combination of a 17α-ethinyl and a 17β-hydroxyl group appears to mimic the 17β-acetyl group of progesterone and confers a measure of metabolic stability on the steroid has made ethisterone the prototype for a very large family of progestagens. Thus, the 19-nor derivative, norethisterone or norethindrone (**3**), reported in 1954 by Djerassi [2], turned out to be a much more potent and selective progestagen, which is still in use today. The 19-nor-18-homo derivative of **2** was reported in 1964 by Smith [3] and turned out to be even more potent. The natural enantiomer, known as levonorgestrel (**4**), is still widely used today.



**1** progesterone



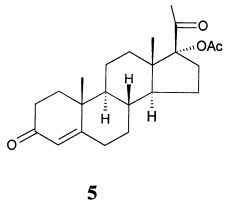
**2** R = $CH_3$ ethisterone
**3** R = H norethisterone



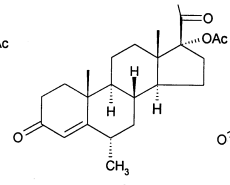
**4** (levo)norgestrel

The metabolic degradation of progesterone was shown to involve primarily reduction of the 20-keto group and the 4,5-double bond, followed by reduction of the 3-keto group.

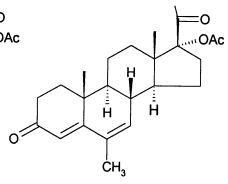
This knowledge allowed the rational design of orally active analogues of progesterone: the 20-keto group could be protected by additional substituents in close proximity, e.g. at C-17 (acyl, alkyl, halogen), C-16 (alkyl, cycloalkyl) and C-21 (OH, alkyl, halogen). Similarly reduction of the A-ring is slowed down by substituents at C-6 (methyl, halogen) and by extending the enone with an additional double bond. Examples are 17α-acetoxyprogesterone (**5**) [4], medroxyprogesterone acetate (**6**) [5], megestrol acetate (**7**) [6], chlormadinone acetate [7], and cyproterone acetate [8].



**5**

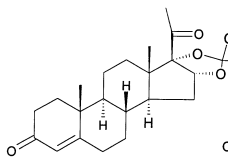


**6**

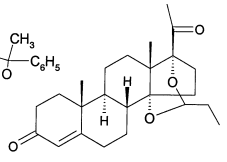


**7**

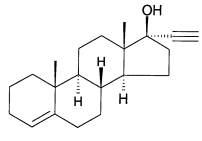
Oral activity was also improved by inversion of the stereochemistry at C-9 and C-10, resulting in so called retrosteroids such as dydrogesterone [9]. Apart from improved oral bioavailibility, these compounds turned out, in many cases, to have higher intrinsic progestational activity in comparison to progesterone as well. Further orally active progestagens discovered were acetals derived from 16α, 17α-dihydroxyprogesterone (algestone), such as the acetophenide **8** [10], and from 14α, 17α-dihydroxyprogesterone, such as proligestone **9** [11]. Also lynestrenol (**10**), the 3-desoxoderivative of norethisterone turned out to be an orally active progestagen [12] due to metabolic conversion into **3**.



**8**



**9**



**10**

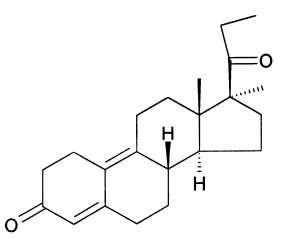
In 1968 over 20 progestagens had found practical use and several hundreds more had been tested. The data have been compiled comprehensively by Neumann [13].

These data provided a sound basis for SAR studies in later decades. Some qualitative conclusions, with regard to SAR that could be drawn were: 1) hydrogen bond accepting substituents at C-3 and C-17 are required; 2) very polar substituents are unfavourable at almost all other positions; 3) small lipophilic substituents are allowed in many positions and often favourable; 4) large substituents are allowed underneath the steroid D-ring,

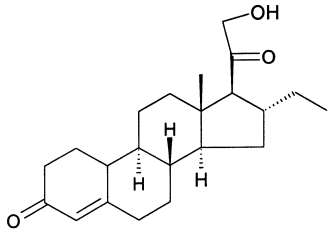
suggesting the presence of a sizeable pocket in the, at the time putative, progesterone receptor.

It should be noted that up to this point the SAR studies were based on in vivo activities, which were not only dependent on the intrinsic activities of the compounds being studied, but also on metabolic stability and pharmacokinetic properties. Therefore, the SAR relationships found were actually describing a combination of properties rather than a single molecular characteristic. This was changed in the early 1970s when it was discovered that certain tissues contain a protein that binds progesterone with high affinity and was subsequently identified as the progesterone receptor. This discovery made it possible to determine the receptor binding affinity of many steroids via a relatively fast and easy in vitro assay [14–20]. In most cases binding affinities were determined by competition with progesterone, promegestone (R 5020) (**11**) or Org 2058 (**12**) and reported as percentages relative to the reference compound.

For many steroids the receptor binding data have been tabulated in graphical [17] or numerical form [16, 21, 22].




**11**




**12**

*2.2. SAR methods*

The idea of exploring the relationship between molecular structure and physical chemical properties was first introduced in 1935 by Hammett [23] and further developed by Hansch and Fujita [24] in 1964. By introducing the substituent constant π (octanol/water partition coefficient) in Hammett's equation they paved the way for the use of physical-chemical parameters (lipophilicity π, the Taft parameters Es and σ, etc.) for the prediction of biological activities. This type of analysis (parameters π, Es, σ, etc. in combination with a multiple linear regression technique) became rapidly famous as the Hansch analysis and it played a pivotal role in (Q)SAR studies for decades. In the same year, Free and Wilson [25] published a mathematical model based on the additive contribution of the individual fragments present into the molecule to the corresponding activity. This approach, the Free-Wilson approach, was applied with reasonable success to congeneric data sets through the years as well.

At the same time that these SAR techniques were developed, steroids were synthesized in large numbers and tested. Because of their limited flexibility and great chemical diversity, these compounds became an immediate target for SAR studies. A sudden urgency to come to the molecular descriptors or physical-chemical parameters, which would describe their activities, rapidly developed.

As a consequence, a debate soon arose. In 1973, Teutsch and Shapiro [26] published a work on a series of Δ6-6-substituted progestagens where they emphasized the importance of the steric influence of substituents on their activities. One year later, Wolff and Hansch [27] criticized their work by showing that by means of the multiparameter regression technique applied to a set of 13 steroids the hydrophobic and the electron withdrawing descriptors were the best descriptors.

In answer to Hansch, Topliss and Shapiro [28] showed that Hansch analysis was biased by the omission of one outlier and that if the whole original dataset of 14 steroids was considered a two-term equation including hydrophobic and steric effects would best describe the activities. The same year, 1975, Wolff, Hansch, Kollman and Duax [29] repeated the same analysis on another dataset of 9α glucocorticoids and progestagens and showed that their activities were best described by a combination of hydrophobicity and molar refractivity parameters.

One year later, Coburn and Solo [30] criticized again the original work of Wolff and Hansch on Δ6-6-substituted progestagens by repeating the analysis with hydrophobic, electronic and many steric parameters. The conclusion was that steric parameters are important for the description of steroidal activities.

In 1977, Lee et al. [31] performed a step-wise linear regression analysis on the binding affinities to the progesterone receptor of a set of 55 progestagens where they showed how strongly the molecular surface area is correlated with hydrophobicity in this class of compounds.

This debate was not conclusive with respect to the best (combination of) descriptor(s) to be used for steroids in SAR studies, but (a posteriori) it reflects the amount of attention that this new class of compounds was receiving.

Methods other than Hansch and Free-Wilson analyses such as pattern recognition methods were also applied [32]. Moriguchi applied the Adaptive Least-Squares (ALS) and the Linear Discriminant Analysis (LDA) methods to rank with moderate success several datasets of compounds (among which steroids). These methods turned out to be most useful when potencies are ordered

in an ordinal (sequential) scale or when the activity is given by the kind of action.

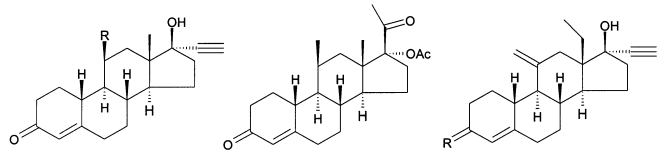
### *2.3. Crystal structures and steroid binding mechanism models*

While the 'Δ6-6-substituted progestagens' debate was progressing, a large number of crystal structures of steroids were determined with the objective of elucidating the working mechanisms of these compounds. The first attempts to rationalize the A, B, C and D rings puckering and strain energies in terms of the substituent or saturation present in the steroidal skeleton were performed with little success. The work of Altona [33, 34] may serve as a typical example. Some time later, Duax, not satisfied with Altona's results, decided to approach the problem statistically by analysing A and D ring puckering as a function of substituents in large data sets of crystal structures.

Duax theory on the working mechanism of steroids started to develop from these studies [35–38]. In his view, subtle conformational changes in the steroidal skeleton were determining the biological activities of these compounds and upon binding all steroids would assume the same 'active' conformation. Further, the A ring (and the substituent on position 3 on it) would determine the binding of the steroid to the receptor, while the D ring would determine its function (agonistic and antagonist activity). This was first concluded from the observation that most chemical and structural variation, on average, is present on the D ring, while only few possibilities for binding are present on the A ring. The D ring would therefore induce the allosteric form of the receptor. Contact of the D ring with a base pair in DNA for activation and transcription factors was also considered.

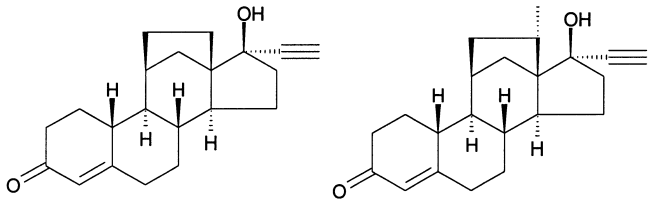
While Duax was persisting in his vision on the working mechanisms of steroids, other ideas contrasting with such vision and supporting the 'induced-fit model' started to develop. The work of Zeelen [39, 40] is one example. From his and other QSAR studies, Zeelen concluded that there is not one specific conformation of the rings in steroids that corresponds to the 'active' conformation. Depending on substituent(s) and/or saturation(s) present on the skeleton the steroid will adopt a conformation, which will be adapted (induced-fit) in the binding cavity of the receptor. In other words, steroids will not adopt upon binding the same 'binding' conformation, but the substituents and/or structural modifications present on the steroidal skeleton will determine their conformations. Such conformations will be accommodated in the binding cavity through a synergistic fit between ligand and receptor. Through the years, many studies (from the work of Bohl and Kaufmann [41, 42] till recently the work of Broess and Groen [43, 44]) strongly supported the induced-fit model.

11-Substituted steroids form a case in point. It was found that 11β-substitution (alkyl, halogen) in many cases leads to greatly enhanced progestational activity. Also short 11-alkylidene substituents, e.g. a methylene group, give a major increase in activity, but not an 11α-substituent [45]. This led to the development of several highly active progestagens, which have found widespread use, i.e. norgestomet **(14)** in veterinary medicine, etonogestrel **(15)** for contraception via non-oral routes and desogestrel **(16)** for oral contraception.




**13** **14** **15** R = O, **16** R = $H_2$

It was found by X-ray crystallography that the steroid skeleton in **13** (R = alkyl, halogen) is curved downward markedly in comparison with the parent compound. This bending of the steroid is obviously due to the repulsion between the 11β substituent and the angular methyl group at C-13. It was speculated that this bending put the 3-keto group in a more favourable position for binding to the progesterone receptor. However, recently norethisterone derivatives **17** and **18** were synthesized, which are bridged analogues of **13**, and they were found to be equally potent progestagens, in spite of the fact that the bridge caused a curvature in the 'wrong' upward direction.




**17** **18**

### *2.4. Molecular mechanics calculations and $^{13}$C-chemical shifts*

While crystal structures of steroids were determined, Allinger [46] was testing his first versions of the MM force field on them. Structural factors (bond lengths, angles and dihedral angles) as well as reaction rates for

congeneric series were calculated with MM and compared with the corresponding experimental data.

Allinger's attempt to obtain correct in silico structures of steroids could not pass unobserved. Duax [37, 38] performed an extended analysis on conformer populations of steroids from crystal structures and MM calculations. His conclusions were that crystal structures outperform MM calculations because MM yielded correct bond distances and angles, but deviated as much as 50° from the crystal structure in the estimate of the dihedral angles. Relative energies of compounds turned out to be off by at least a factor of 10. Duax concluded that the program of Allinger, MM2p, needed improvement and that the information taken from crystallography could help that process.

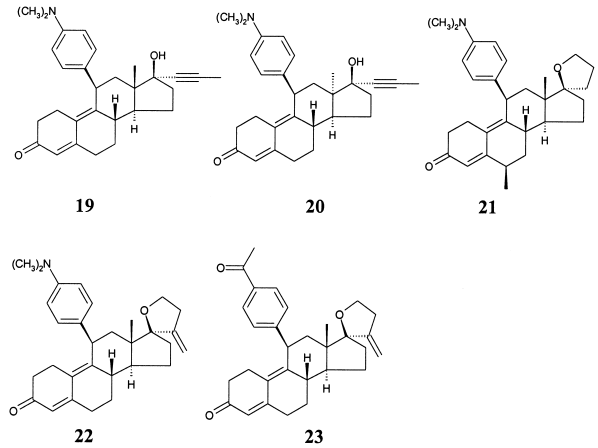
The work of Allinger was a breakthrough. Schneider and Gschwendtner [47–49] performed several MM calculations on steroids in order to come to a better understanding of the substituent effects on the different rings of the steroidal skeleton and more specifically of the polar and steric through-bond transmission change mechanisms. In these investigations they often combined the MM calculations with the $^{13}$C-NMR chemical shifts of the steroidal skeletons. In this sense, the work of Wray [50] in 1981 had been inspirational. In his study on a small congeneric set of progestagens, Wray empirically investigated the effect of substituents on the $^{13}$C-chemical shifts of the steroidal skeletons. He found a consistency in shifts for the same substituents. One year later, Schneider and Gschwendtner applied the same approach to larger data sets and their conclusion supported the usefulness of $^{13}$C-NMR measurements in detecting small changes in geometries and electron densities in steroids.

Hoppen and Hammann [51] unsuccessfully tried to correlate the PR and AR binding affinities of a set of eight progestagen derivatives of norethisterone with their corresponding $^{13}$C-NMR chemical shifts. This result is not surprising considering the limited amount of compounds analysed and the empirical approach followed.

Despite the pioneering work of Hall and Sanders in 1980 [52], who published the first conformational $^{1}$H-NMR study on steroids, at this point in time the relatively simple interpretability of $^{13}$C-NMR shifts was preferred to the more complex $^{1}$H-NMR shifts in SAR studies.

### *2.5. PR receptor mapping*

The ultimate objective of every (Q)SAR study is not only limited to the identification of correlations between structure and activity, but it is also extended to the mapping of the active site of the studied target. A significant impact on these studies was provided by the discovery of progesterone antagonists. Unlike antagonists for, e.g., the oestrogen and androgen receptors, antagonists for the progesterone receptor have remained unknown for a very long time. It was only in 1982 that Roussel Uclaf reported the first high affinity antagonists. The prototype for this class of compounds is RU 486 or mifepristone (**19**), which, however, is also a very potent glucocorticoid antagonist. Since the mixed profile of RU 486 was considered a drawback for certain therapeutic applications, much research was devoted to find more selective anti-progestagens. Some examples are **20–23** [53, 54].




First attempts to map the progesterone receptor were performed by Doré [55], Ojasoo [56] and Zeelen [57] in the late eighties. Doré performed a correspondence analysis on the selectivity of a set of steroids with respect to four different nuclear receptors (AR, PR, ER and GR). By projection of the principal components, receptor maps were obtained. In Ojasoo's study, the authors tried to map the PR and AR receptors by looking at the crystal structures of chemically different progestins and androgens and at their corresponding receptors. They came to the conclusion that space is present in both receptors around the 3-keto, 11β, 7α and, in PR, C21 positions. Hydrophobic pockets are distributed around them all. At the time that this paper was written, several receptors (ER, GR and PR) were cloned and sequenced. In Zeelen's receptor mapping study of PR it was concluded that a large hydrophobic pocket is present under the D ring and that possible structural modifications of the steroidal skeleton should take place under the D ring.

## 3. From the late 1980s till now

In the 1980s, major advances were made in characterizing the progesterone receptor, the biological target for progestagens. Modern biotechnology enabled the cDNA cloning of all steroid hormone receptors, including the progesterone receptor. The human progesterone receptor was found to occur in two forms, the full-length receptor, PR B, consists of 933 amino acids, while the truncated version, PR A, consists of 769 amino acids [58]. From then on the knowledge of steroid hormone action on a molecular level has dramatically increased [59]. Receptor cloning has also been of particular importance to the discovery and design of new progestagens. In the first place cell constructs have been prepared which contain both the progesterone receptor and a suitable reporter gene coupled to a promoter which is responsive to the progesterone receptor/agonist complex. The so-called transactivation assays, based on these constructs, lend themselves for rapid screening of large numbers of compounds, including non-steroids. In the second place it became possible to prepare pure receptor protein in relatively large amounts, which made it possible, at least in principle, to determine the three-dimensional structure via X-ray crystallography, high field NMR, etc. While the size of the complete receptor thus far has resisted such an approach, the hormone-binding domain of the progesterone receptor has been crystallized in the presence of progesterone as the ligand, and the structure determined by X-ray crystallography [60].

In the SAR field, the Comparative Molecular Field Analysis (CoMFA) method [61] undoubtedly represents a milestone in the development of three-dimensional-QSAR methods. Cramer first published it in 1988 and since then it has become of wide-spread use amongst medicinal chemists. As introduced in 1988, however, CoMFA (although powerful and elegant) needed several technical/statistical improvements, and because of its intrinsic requirement for molecular superposition, it determined one way of performing 3-D-QSAR studies, i.e. with alignment. It is not surprising that from 1988 till now many QSAR developments have been devoted either to improve CoMFA or to derive CoMFA-like approaches, or to develop 3-D-QSAR methods which, in contrast to CoMFA, would not require a molecular alignment.

Further, the 1990s were also characterized by the development of several receptor mapping techniques, spanning a wide range of complexity (from simple, CoMFA, to sophisticated, PARM).

### *3.1. 3-D-QSAR with alignment*

In CoMFA, activities are predicted in terms of differences in steric and electrostatic interaction energies (molecular fields) between an atom probe and every atom in a molecule and every molecule in the data set. The interaction energies are calculated on a three-dimensional grid, whose points are set to an arbitrary distance within the CoMFA box. Compounds are by default centred in this box and superimposed on each other. Once the 3-D molecular fields are calculated for all molecules and accurate activities (mostly binding affinities or potencies) are available, the multivariate linear regression technique partial least square is used to derive a statistical model.

The statistical results can be further visualized back on the 3-D grid in terms of steric and electrostatic contours, which suggest where compounds should be modified in steric and electrostatic terms to enhance or decrease their activities. Because of this, CoMFA finally had become the tool that medicinal chemists had been waiting for for a long time. As introduced in 1988, however, the method suffered from several limitations such as a grid-dependence of the statistical results (absence of rotational and translational invariance), a very stiff Lennard-Jones potential for calculating steric interactions, and a partial charge type dependence of the electrostatic interactions [62]. Further CoMFA contours are not transferable because they are entirely dependent on the data set investigated.

Some of these limitations were successfully dealt with by the work of Baroni [63] to improve variable selection in partial least squares analyses, the work of Klebe [64] to improve the calculations of the electrostatic and steric interactions by Gaussian smoothing of the Lennard-Jones and Coulomb potentials and the work of Cho [65] to ensure the rotational and translational invariance of the statistical quantities.

The steroid data set on which CoMFA was first published became quickly known as the ‘CoMFA data set’. Many new methods were subsequently tested on this data set to compare performances. The work of Oprea [66] is an interesting example, where he compared the minimal steric difference [67] approach with CoMFA. The results were not conclusive, but CoMFA was criticized for the too repulsive cut-off and for the stiffness of the Lennard-Jones potential. In 1994, Jain [68] published Compass, a method in which steric and electrostatic interactions are sampled close to the surface of compounds by means of neural networks. Its performance was tested on the steroid benchmark against CoMFA and the similarity index approach. Compass performed better.

The intrinsic requirement of an alignment and the need of finding the 'active' conformation were, however, criticized.

In 1996, Kellogg et al. [69] introduced the E and HE topological fields in 3-D-QSAR. In a CoMFA-like approach they substituted the scoring function used in CoMFA with the calculated E and HE fields. Their results showed that these fields perform better than the steric and electrostatic molecular fields. However, alignment is always needed and the interpretation of the E and HE contours is less intuitive than the corresponding contours in CoMFA.

### 3.2. 3-D-QSAR without alignment

Molecular superposition requires knowledge of the binding mode of a ligand towards its target. Unfortunately, this mode is not always known. In this case a theoretical superposition must be derived. This process is not always straightforward and objective, which suggests that analyses and corresponding conclusions based on theoretical alignments should be performed with great care. Developing a 3-D-QSAR method that is independent from molecular superposition is undoubtedly one of the greatest challenges in the field of QSAR. Several attempts have already been performed. In 1996, Silverman and Platt [70] published the Comparative Molecular Moment Analysis (CoMMA), where moments of inertia, dipole and quadruple moments are used as molecular descriptors. The statistical results compared favourably with CoMFA. Visualization of the results was not considered.

In 1997, simulated infrared spectra were used by Ferguson [71] in the EVA approach to predict binding affinities. EVA correlations compared well with CoMFA.

In the same year, Bravi [72] published the MS-WHIM (Molecular Surface-Weighted Holistic Invariant Molecular) indices. The WHIM indices had already been introduced earlier [73]. In the original WHIM work, the Cartesian coordinates of the nuclei were used to calculate roto-translational invariant molecular moments. In the extended MS approach, the same indices were calculated from Connolly molecular surface points. The approach was tested on the steroid benchmark and compared with CoMFA, Compass and the Carbo similarity indices. The results were more than favourable. Within this approach, however, visualization of the statistical results remains a critical issue.

Full exploitation of molecular spectra in QSAR studies was only recently published by Bursi [74], after almost 20 years the work of Wray on $^{13}$C-NMR spectra in SAR studies. In the novel Comparative Spectra Analysis (CoSA) approach, experimental $^{1}$H-NMR, mass, and IR spectra and simulated $^{13}$C-NMR and IR spectra were used alone or in combination to predict the binding affinities of a set of progestagens. The results compared more than favourably with CoMFA, strongly supporting the use of spectroscopic fields in QSAR studies.

### 3.3. QSAR without 3-D

Although dominated by CoMFA, in the nineties QSAR studies continued to be performed by means of a great variety of methods.

One example is the work of Good [75] who implemented the Carbo index in a large variety of ways, and by means of the corresponding similarity matrices predicted the biological affinities of different data sets. The program GOLPE [63] was used as variable selector. The conclusion of this study was that similarity matrices should be more often used in QSAR studies.

A data set of progestagens was also investigated by means of a combination of genetic algorithm (GA) and neural networks (NN) [76]. In this study molecular descriptors were selected by means of GA and correlations with the binding affinities were obtained by means of NN. Conclusions were that a non-linear technique leads to better results on this data set than partial least squares or other linear regression techniques. The interpretability of the results, however, remained very difficult.

Novak [77] has recently published a study in which he presented several UV spectra for steroids of great interest, such as progesterone and testosterone. Assignment of the UV transitions were performed by means of AM1 MO calculations. Ionization potentials were further determined and a SAR study was attempted where ionization energies were correlated with receptor binding affinities. Unfortunately, no proper SAR could be derived on the basis of the available data.

### 3.4. Receptor mapping

In the 1990s many different approaches were developed to map the active sites of drug targets from a data set of ligands. These approaches were mostly applied to targets whose crystal structures were still unknown. Crystallization of the progesterone (PR) nuclear receptor and of some other nuclear receptors was achieved only recently [60]. The mapping of the progesterone receptor ligand-binding domain was therefore highly desirable.

One of CoMFA's great advantages is the visualization of the statistical results in terms of contours around the molecules. When binding data are considered for biological activity, these contours can also be seen as fingerprints

of the target's active site. In this sense, CoMFA belongs as well to receptor mapping approaches. Many CoMFA studies were therefore performed on PR alone or on PR and other nuclear receptors to obtain ideas on its active site or ideas on the differences between different active sites. Unfortunately, as published most of these studies were poor.

Another interesting receptor site model is Doweyko's Hypothetical Active Lactice (HASL) approach [78], where a four-dimensional lattice is build around each molecule in the data set. The four dimensions are based on the three Cartesian coordinates of the atoms and a fourth dimension, which can be a physical-chemical property of choice. The multidimensional lattices, which are generated, are used to compare molecules with each other and to generate QSARs. The results discussed in this study were encouraging.

In 1995, Hahn published a specific type of receptor site model, i.e., the Receptor Surface Model (RSM) [79]. Based on the idea of the active analogue approach, a dataset of active compounds is used to build the surface of the active site. Subsequently, 3-D energetics descriptors can be calculated from the interactions between RSM and ligand and used in the corresponding QSAR. A different selection of active compounds will obviously lead to a new RSM and, therefore, to new descriptors. Genetic algorithm is then used to select which descriptors (and therefore RSM models) are mostly valuable for the QSAR study considered. The approach was compared with CoMFA and Compass on different data sets. On the 'CoMFA data set' RSM does not perform better than the other approaches.

In 1998, the Pseudo Atomic Receptor Model (PARM) [80, 81] was published. PARM is, in fact, an improvement of Walter's Genetic Evolved Receptor Model (GERM) [82]. As in the other approaches, PARM needs a molecular alignment as well. Around the surface of each molecule a grid of points is built. At each grid point interaction energies are calculated between a given atom type and the closest atom in the molecule. Fifteen atom types are considered amongst which no atom at all is also considered. All possible combinations of atom types on the grid of points are considered, which leads to several models. At this point, in GERM, genetic algorithm evolves the individual combinations to the best combinations that are valuable for QSARs. In PARM the process is guided by a charge-dependent evolution of the individuals, where complementary charges between ligand and receptor are assumed. Partial charges are assigned to the atom types and only individuals or combinations of them, which are complementary to the partial charges in the molecules or molecule model, further evolve. PARM was tested amongst others on a steroid data set and compared with CoMFA. The cross-validated results as well as the predictive ability of the method were satisfactory.

## 4. Conclusions

Progestagens are steroid hormones which can be used in several therapeutic areas such as contraception, hormone replacement therapy and gynaecological disorders.

From the start they represented a very interesting class of compounds because of their relative rigidity and chemical diversity. As soon as reasonable amounts of progestagens were synthesized, they became objects of a great variety of studies, not only for synthetic and computational medicinal chemists, but also for computational chemists in general, crystallographers, and statisticians.

It is therefore not surprising that while historically looking at the origin and further developments of progestagens, we ended up looking to a large extent at the historical development of the (Q)SAR field.

It is undoubtedly true that we have come a long way in the process of understanding the properties and mechanisms of action of this class of compounds. Nowadays we can develop very potent and selective steroidal and non-steroidal progestagens, we can visualize the active site of the ligand-binding domain of the PR receptor and progesterone bound to it and we can derive reliable and robust QSAR models using a great variety of molecular descriptors from field to spectral descriptors.

The challenge is, however, not over. Genomics and chip technologies drive nowadays the search for new targets, combinatorial chemistry (CC) and high-throughput screening (HTS) dominate the lead discovery process, and high-speed synthesis (HSS) and early ADME (Adsorption Distribution Metabolism Excretion) accelerate the lead optimization cycles. Despite all this, independently from the level of sophistication at which we are operating, we are continuously confronting ourselves with the concept of correlating structures with activities whenever we need to come to a full understanding of a given molecular mechanism. In this sense alone, the knowledge and the experience that have been acquired in more than sixty years of (Q)SAR developments should be treasured.

## References

[1] Inhoffen H.H., Logemann W., Hohlweg W., Serini A., Chem. Ber. 71 (1938) 1024–1030.

[2] Djerassi C., Miramontes L., Rosenkranz G., Sondheimer F., J. Am. Chem. Soc. 76 (1954) 4092–4094.

[3] Smith H., Hughes G.A., Douglas G.H., et al., J. Chem. Soc. (1964) 4472–4492.

[4] Junkmann K., Arch. Exp. Pathol. Pharmakol. 223 (1954) 244–253.

[5] Babcock J.C., Gutsell E.S., Herr M.E., Hogg J.A., Stucki J.C., Barnes L.E., Dulin W.E., J. Am. Chem. Soc. 80 (1956) 2904–2905.

[6] Ringold H.J., Perez Ruelas J., Batres E., Djerassi C., J. Am. Chem. Soc. 81 (1957) 3712–3716.

[7] Brückner K., Hampel B., Johnsen U., Chem. Ber. 94 (1961) 1225–1240.

[8] Laurent H., Schulz G., Wiechert R., Chem. Ber. 102 (1969) 2570–2582.

[9] Westerhof P., Reerink E.H., Recl. Trav. Chim. Pays-Bas 79 (1960) 771–783.

[10] Fried J., Sabo E.F., Grabowich P., Lerner L.J., Kessler W.B., Brennan D.M., Borman A., Chem. Ind. (London) (1961) 465–466.

[11] Van der Sijde D., Kooreman H.J., Jaitly K.D., Marx A.F., J. Med. Chem. 15 (1972) 909–914.

[12] De Winter M.S., Siegmann C.M., Szpilfogel S.A., Chem. Ind. (London) (1959) 905.

[13] Neumann F., Handbuch der Experimentellen Pharmacologie, Band XXII/1, Springer Verlag, 1968.

[14] McGuire J.L., Bariso C.D., Shroff A.P., Biochemistry 13 (1974) 319–322.

[15] Smith H.E., Smith R.G., Toft D.O., Neergaard J.R., Burrows E.P., O'Malley B.W., J. Biol. Chem. 249 (1974) 5924–5932.

[16] Kontula K., Janne O., Vihko R., de Jager E., de Visser J., Zeelen F.J., Acta Endocrinologica 78 (1975) 574–592.

[17] Raynaud J.P., Ojasoo T., Bouton M.M., Philibert D., Receptor Binding as a Tool in the Development of New Bioactive Steroids, in: Ariens J. (Ed.), Drug Design, vol. VIII, Academic Press, 1979.

[18] Raynaud J.P., Bouton M.M., Moguilewsky M., Ojasoo T., Philibert D., Beck, Labrie F., Mornon J.P., J. Steroid Biochem. 12 (1980) 143–157.

[19] Bergink E.W., Hamburger A.D., de Jager E., van der Vies J., J. Steroid Biochem. 14 (1981) 157–183.

[20] Bergink E.W., van Meel F., Turpijn E.W., van der Vies J., J. Steroid Biochem. 19 (1983) 1563–1570.

[21] Seeley D.H., Wang W.-Y., Salhanick H.A., J. Biol. Chem. 257 (1982) 13359–13366.

[22] Neelima, Seth M., Bhaduri P., Arzneim.-Forsch. (1986) 151–188.

[23] Hammett L.P., Chem. Rev. 17 (1935) 125–136.

[24] Hansch C., Fujita T., J. Am. Chem. Soc. 86 (1964) 1616–1626.

[25] Free S.M., Wilson J.W., J. Med. Chem. 7 (1964) 395–399.

[26] Teutsch G., Weber L., Page G., Shapiro E.L., Herzog H.L., J. Med. Chem. 16 (1973) 1370–1376.

[27] Wolff M.E., Hansch C., J. Med. Chem. 17 (1974) 898–899.

[28] Topliss J.G., Shapiro E.L., J. Med. Chem. 18 (1975) 621–623.

[29] Wolff M.E., Hansch C., Giannini D.D., Kollman P.A., Duax W.L., Baxter C.C., J. Steroid Biochem. 6 (1975) 211–214.

[30] Coburn R.A., Solo A.J., J. Med. Chem. 19 (1976) 748–754.

[31] Lee D.L., Kollman P.A., Marsh F.J., Wolff M.E., J. Med. Chem. 20 (1977) 1139–1146.

[32] Moriguchi I., Komatsu K., Matsushita Y., Anal. Chim. Acta 133 (1981) 625–636.

[33] Altona C., Geise H.J., Romers C., Tetrahedron 23 (1967) 439–463.

[34] Altona C., Geise H.J., Romers C., Tetrahedron 24 (1968) 13–32.

[35] Duax W.L., J. Steroid Biochem. 15 (1981) 41–47.

[36] Duax W.L., J. Tox. Envir. Hlth. 4 (1978) 205–227.

[37] Duax W.L., Griffin J.F., Rohrer D.C., X–Ray Crystallography and Drug Action, in: Horn A.S., de Router C.J. (Eds.), Course Int. Sch. Crystallogr. 9th, Oxford University Press, Oxford, 1984, pp. 405–426.

[38] Duax W.L., Griffin J.F., Rohrer D.C., in: Zalewski R.I., Skolik J.J. (Eds.), Natural Products Chemistry, 1984, pp. 3985–3996.

[39] Van der Broek A.J., Broess A.I.A., Heuvel M.J., de Jongh H.P., Zeelen F.J., Steroids 30 (1977) 481–510.

[40] Zeelen F.J., Biol. Act. Chem. Struct. (1977) 147–159.

[41] Bohl M., Kaufmann G., Hubner M., Reck G., Kretschmer R.-G., J. Steroid Biochem. 23 (1985) 895–900.

[42] Bohl M., Simon Z., Vlad A., Kaufmann G., Ponsold K., Z. Naturforsch. 42c (1987) 935–940.

[43] Broess A.I.A., Groen M.B., Hamersma H., Bioorg. Med. Chem. Lett. 7 (1997) 2925–2928.

[44] Broess A.I.A., Groen M.B., Hamersma H., Bioorg. Med. Chem. Lett. 7 (1997) 2929–2934.

[45] Van der Broek A.J., Broess A.I.A., Heuvel v.d. M.J., Jongh de H.P., Leemhuis J., Schönemann K.H., Smits J., Visser de J.N., Vliet van P., Zeelen F.J., Steroids 30 (1977) 481–510.

[46] Allinger N.J., Lane G.A., J. Am. Chem. Soc. 96 (1974) 2937–2941.

[47] Schneider H.-J., Gschwendtner W., Weigand E.F., J. Am. Chem. Soc. 101 (1979) 7195–7198.

[48] Schneider H.-J., Buchheit U., Gschwendtner W., Lonsdorfer M., Mol. Struct. Biol. Act. (1982) 165–179.

[49] Schneider H.-J., Gschwendtner W., J. Org. Chem. 47 (1982) 4216–4221.

[50] Wray V., Tetrahedron 37 (1981) 777–780.

[51] Hoppen H.-O., Hammann P., Acta Endocrinol. 115 (1987) 406–412.

[52] Hall L.D., Sanders J.K.M., J. Am. Chem. Soc. 102 (1980) 5703–5711.

[53] Teutsch G., Philibert D., Progesterone Antagonists in Reproductive Medicine and Oncology, Human Reproduction, vol. 9 Suppl. 1, Oxford University Press, Oxford, 1994, pp. 12–32.

[54] Schoonen W.G.E.J., Vermeulen G.J., Deckers G.H., Verbost P.M., Kloosterboer H.J., Curr. Top. Steroid Res. 2 (1999) 15–54.

[55] Doré J.-C., Gilbert J., Ojasoo T., Raynaud J.-P., J. Med. Chem. 29 (1986) 54–60.

[56] Ojasoo T., Delettré J., Mornon J.P, Turpin-VanDycke C., Raynaud J.P.J., Steroid Biochem. 27 (1987) 255–269.

[57] Groen M.B., van der Heuvel M.J., Deckers G.H.J., Kloosterboer H.J., Kelder J., Zeelen F.J., Organon Internal Communication, 1988.

[58] Misrahi M., Atger M., d'Auriol L., Loosfelt H., Meriel C., Fridlansky F., et al., Biochem. Biophys. Res. Commun. 143 (1987) 740–748.

[59] Beato M., Cell 35 (1998) 335–344.

[60] Williams S.P., Sigler P.B., Nature 393 (1998) 392–395.

[61] Cramer R.D., Patterson D.E., Bunce J.D., J. Am. Chem. Soc. 110 (1988) 5959–5967.

[62] Bursi R., Grootenhuis P.D.J., J. Comput.-Aided Mol. Des. 13 (1999) 221–232.

[63] Baroni M., Constantino G., Cruciani G., Riganelli D., Valigi R., Clementi S., Quant. Struct.-Act. Relat. 12 (1993) 9–20.

[64] Klebe G., Abraham U., Mietzner T., J. Med. Chem. 37 (1994) 4130–4146.

[65] Cho S.J., Tropsha A., J. Med. Chem. 38 (1995) 1060–1066.

[66] Oprea T.I., Ciubotariu D., Sulea T.I., Simon Z., Quant. Struct.-Act. Relat. 12 (1993) 21–26.

[67] Simon Z., Rev. Roumaine Chimie 37 (1992) 323–327.

[68] Jain N., Koile K., Chapman D., J. Med. Chem. 37 (1994) 2315–2327.

[69] Kellogg G.E., Kier L.B., Gaillard P., Hall L.H., J. Comput.-Aided Mol. Des. 10 (1996) 513–520.

[70] Silverman B.D., Platt D., J. Med. Chem. 39 (1996) 2129–2140.

[71] Ferguson A.M., Heritage T., Jonathon P., Pack S.E., Phillips L., Rogan J., Snaith P.J., J. Comput.-Aided Mol. Des. 11 (1997) 409–422.

[72] Bravi G., Gancia E., Mascagni P., Pegna M., Todeschini R., Zaliani A., J. Comput.-Aided Mol. Des. 11 (1997) 79–92.

[73] Todeschini R., Lasagni M., Marengo E., J. Chemom. 8 (1994) 263–272.

[74] Bursi R., Dao T., van Wijk T., de Gooyer M., Kellenbach E., Verwer P., J. Chem. Inf. Comp. Sci. 39 (1999) 861–867.

[75] Good C., Peterson S.J., Richards W.G., J. Med. Chem. 36 (1993) 2929–2937.

[76] Van Helden S., Hamersma H., van Geerestein V.J., Genetic Algorithms in Molecular Modelling, Academic Press, London, 1996, pp. 159–192.

[77] Novak I., Kovac B., Biophys. Chem. 78 (1999) 233–240.

[78] Doweyko A.M., J. Med. Chem. 31 (1998) 1396–1406.

[79] Hahn M., Rogers D., J. Med. Chem. 38 (1995) 2091–2102.

[80] Chen H., Zhou J., Xie G., J. Chem. Inf. Comput. Sci. 38 (1998) 243–250.

[81] Santagati M., Chen H.M., Santagati A., Modica M., Guccione S., Uccello Barretta G., Balzano F., in: Gundertofte K., Jorgensen F.S. (Eds.), Molecular Modelling and Prediction of Bioactivity, Kluwer Academic/Plenum Publishers, New York, 2000, pp. 443–450.

[82] Walters D.E., Hinds R.M., J. Med. Chem. 37 (1994) 2527–2531.