Pergamon

Bioorganic & Medicinal Chemistry 6 (1998) 377–388

BIOORGANIC & MEDICINAL CHEMISTRY

Comparative Molecular Field Analysis of Non-steroidal Aromatase Inhibitors: An Extended Model for Two Different Structural Classes

Maurizio Recanatini* and Andrea Cavalli

Department of Pharmaceutical Sciences, University of Bologna, Via Belmeloro 6, I-40126 Bologna, Italy

Received 7 August 1997; accepted 7 October 1997

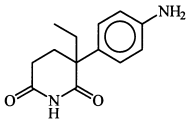
Abstract—Aromatase is a cytochrome P450 isozyme, whose inhibition is known to be therapeutically relevant in the treatment of the breast cancer. A comparative molecular field analysis (CoMFA) has been carried out on a series of non-steroidal aromatase inhibitors belonging to two different structural classes. One subset of compounds consists of fadrozole analogues and was studied in a previous work, from which a 'local' 3-D quantitative structure–activity relationship (QSAR) model for the inhibition of aromatase was obtained. In the present paper, that model is extended to include a second subset of compounds bearing a tetralone nucleus and acting at the same enzyme site with the same mechanism as the azoles. The critical alignment step has been solved by using two different steroidal inhibitors of aromatase as rigid templates, on which the non-steroidal compounds have been superimposed. The final 3-D QSAR models are discussed in terms of predictivity and some implications regarding the steric and electronic requirements of steroidal and non-steroidal inhibitors are pointed out.

Introduction

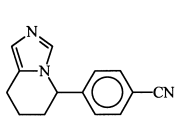
Aromatase inhibitors are among the most actively studied compounds in the field of antitumour agents, because of their role in the treatment of breast cancer.[1] They block the biosynthesis of estrogens at the last step, thus reducing the levels of circulating hormones, which are involved in the progression of the disease. Aromatase is a cytochrome P450 isozyme (P450 XIX) that can be inhibited either competitively or non competitively by various classes of both steroidal and non-steroidal compounds.[2] Two reviews by Banting cover in a comprehensive way many aspects of the medicinal chemistry of the inhibitors of aromatase.[3,4]

The former antiepileptic drug aminoglutethimide (**1**) is the first and most known non-steroidal inhibitor of aromatase acting through a characteristic competitive mechanism, that involves the coordination of the heme Fe^{2+} ion by the primary amino group. This kind of inhibitors bind to the cytochrome P450 enzyme in a way that one heteroatom (N, S, O) coordinates to the iron of the porphyrin ring causing a typical bathochromic shift in the UV-absorption spectrum Soret band of the heme.

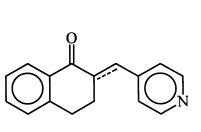
Other classes of compounds were successively identified to be able to block the enzyme via the same mechanism and, from the most important of them, the azoles, fadrozole (**2**) emerged, as a potent and specific antitumour agent.[5,6] On the other hand, in an independent work, Hartmann and his group, starting from earlier observations about the weak aromatase inhibitory activity of some flavone derivatives, developed a class of pyridyl-substituted tetralones (**3**), from which some interesting inhibitors of aromatase were discovered.[7]



1



2



3

Key words: QSAR; CoMFA; aromatase; cytochrome P450; breast cancer.
*Corresponding author. e-mail: mreca@alma.unibo.it

0968-0896/98/$19.00
PII: S0968-0896(97)10053-0

Using the 3-D QSAR technique called CoMFA (Comparative Molecular Field Analysis[8]), we recently developed a model, that allowed us to rationalize the structure–activity relationships (SAR) of a series of 29 azoles (Figure 1) and to identify two regions in the space around the reference molecule *S*-fadrozole, where steric and electronic modifications could favor (or disfavor) the aromatase inhibitory activity.[9] This model can be considered as a 'local' 3-D QSAR, in the sense that it describes the SAR of a series of congeners and, conse-

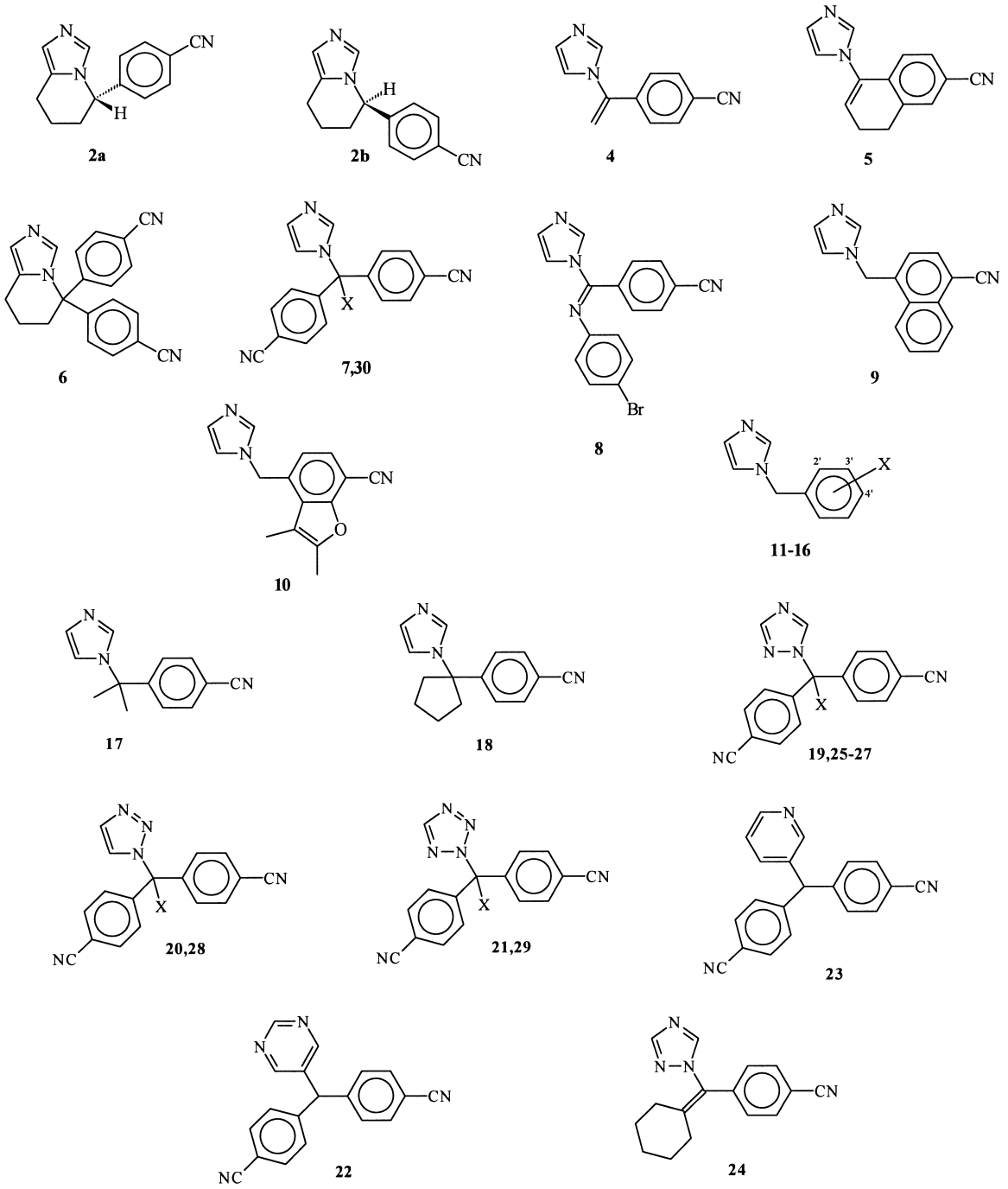


Figure 1. Structure formulae of the azole derivatives of the training set.

quently, it has no general applicability to other classes of compounds. Then, we thought it was worthwhile to try to extend the model, by including in it other compounds belonging to a different structural class, provided that they were tested and shown to act in the same way as the azoles of the first series. If the SAR of different series can be described by a single QSAR expression, one can start to build a general pharmacophore for the biological system under study, thus giving a substantial contribute to the design of new derivatives active on that system.

In this paper, we report the extension of our previous 3-D QSAR analysis,[9] to include in the same model a second series of compounds analogues of **3**. The aim of the work was to test the possibility of building a comprehensive 3-D QSAR model for the inhibition of aromatase by means of the CoMFA technique. It is a remarkable result that CoMFA succeeded in providing such a single consistent model for two structurally different series of analogues.

Methods

Datasets

The non-steroidal aromatase inhibitors considered in this work were divided into two datasets: the 'training set' (Table 1), on which the CoMFA analysis was actually performed, and the 'test set' used only for testing the predictive ability of the 3-D QSAR model. In turn, the training set is composed of two series, one made up by the 29 azole derivatives of Figure 1 (compds **2a**–**30** of Table 1) and the other comprising the 20 tetralone derivatives shown in Figure 2 (compds **31**–**50** of Table 1). The structures of the molecules of the test set are shown in Figure 3.

The observed biological activity data for all the compounds of this study were taken from the literature, both for the training set[7,10–14] and for the test set.[15–17] All the compounds were tested on human placental aromatase following the classical procedure of Thompson and Siiteri[18] and, moreover, all were reported to coordinate to the iron atom of the heme group present at the active site of the enzyme, as shown by the Type II difference spectra. It thus appeared reasonable to consider the reported IC_{50} values as sufficiently homogeneous data for use in the quantitative analysis of the SAR, even if they were generated in different laboratories and no common reference compound is reported.

Molecular modeling

Molecular modeling was performed by means of the SYBYL software[19] running on a Silicon Graphics Indigo2 workstation. Three-dimensional models of all the molecules, except for **2a** and **55**–**60** were built by assembling fragments from the SYBYL standard library. In the cases of *S*-fadrozole (**2a**), *R*- and *S*-vorozole (**55**, **56**) and the tetrahydrocyclopropanonaphthalene derivatives **57**–**60**, the crystal structures of **2a**,[20] **56**,[21] and **58**[17] were used as the starting geometries, respectively. Energy minimizations were performed with MAXIMIN2 (Tripos Force Field[22]) without considering the electrostatic charges.

The conformational space of each molecule was sampled by means of the RANDOMSEARCH procedure of SYBYL. In this Monte Carlo-like method, a number of conformations is generated by randomly rotating selected bonds of the molecules. The degree of completeness of the analysis can be checked by setting a number of parameters: energy cutoff, number of times one conformer is found and number of attempts to find a new conformer. How often the same conformation is found

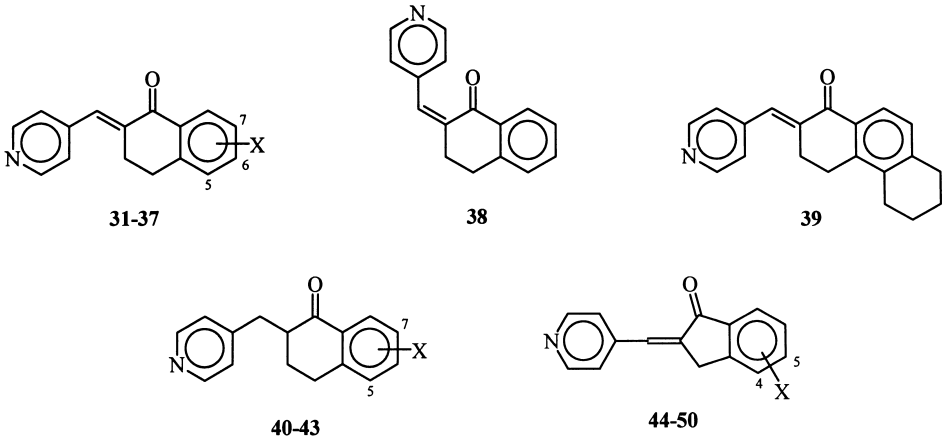


Figure 2. Structure formulae of the tetralone derivatives of the training set.

was suggested as an acceptable measure of completeness:[23] it is estimated that, if a conformer is found n times, then there is a $(1-(1/2)^n)$ probability that all possible conformations were found. In practice, the energy cutoff was set about 50 kcal/mol above the estimated total energy of the molecule, the maximum number of hits was six and the maximum number of attempts was 1000.

Table 1. Observed and calculated biological activities of the aromatase inhibitors of Figures 1 and 2 (training set)

Compound	X	Stereochemistry	pIC_{50obsd}	pIC_{50fit}[a]	Δ	$pIC_{50fitGRS}$[b]	Δ
2a	—	*S*	8.49	8.49	0.00	8.57	−0.08
2b	—	*R*	6.17	6.14	0.03	6.23	−0.06
4	—	—	9.15	8.83	0.32	9.03	0.12
5	—	—	8.54	8.79	−0.25	8.69	−0.15
6	—	—	6.63	6.59	0.04	6.59	0.04
7	H	—	8.46	8.64	−0.18	8.52	−0.06
8	—	—	8.11	8.05	0.06	8.12	−0.01
9	—	—	8.70	8.66	0.04	8.72	−0.02
10	—	—	8.46	8.49	−0.03	8.48	−0.02
11	2′-CN	—	6.52	6.51	0.01	6.48	0.04
12	3′-CN	—	6.00	6.08	−0.08	6.00	0.00
13	4′-Br	—	7.07	7.00	0.07	6.97	0.10
14	4′-COPh	—	8.00	8.22	−0.22	8.22	−0.22
15	$4'NO_2$	—	8.26	8.21	0.05	8.25	0.01
16	4′-CN	—	8.00	8.05	−0.05	7.97	0.03
17	—	—	7.68	7.72	−0.04	7.63	0.05
18	—	—	8.12	7.95	0.17	8.02	0.10
19	H	—	7.96	7.98	−0.02	7.94	0.02
20	H	—	7.92	7.89	0.03	7.97	−0.05
21	H	—	9.00	8.95	0.05	8.88	0.12
22	—	—	7.15	7.22	−0.07	7.20	−0.05
23	—	—	7.91	7.91	0.00	7.90	0.01
24	—	—	7.67	7.64	0.03	7.58	0.09
25	CH_3	—	8.05	7.92	0.13	7.99	0.06
26	C_2H_5	—	7.74	7.83	−0.09	7.84	−0.10
27	F	—	7.89	7.83	0.06	7.86	0.03
28	F	—	7.68	7.79	−0.11	7.59	0.09
29	F	—	7.52	7.64	−0.12	7.64	−0.12
30	F	—	8.24	8.15	0.09	8.20	0.04
31	H	—	5.03	5.29	−0.26	5.13	−0.10
32	$5\text{-}OCH_3$	—	5.20	5.45	−0.25	5.42	−0.22
33	$6\text{-}OCH_3$	—	5.05	5.07	−0.02	5.00	0.05
34	$7\text{-}OCH_3$	—	4.89	5.09	−0.20	4.99	−0.10
35	5-OH	—	6.16	5.88	0.28	5.99	0.17
36	6-OH	—	5.06	5.04	0.02	5.06	0.00
37	7-OH	—	5.52	5.31	0.21	5.34	0.18
38	—	—	4.92	4.81	0.11	4.92	0.00
39	—	—	5.00	5.20	−0.20	5.03	−0.03
40	$7\text{-}OCH_3$	(−)	5.02	5.07	−0.05	4.82	0.20
41	$7\text{-}OCH_3$	(+)	5.07	5.18	−0.11	5.10	−0.03
42	5-OH	(−)	6.68	6.55	0.13	6.67	0.01
43	5-OH	(+)	6.77	6.75	0.02	6.80	−0.03
44	H	*E*	4.60	4.68	−0.08	4.73	−0.13
45	H	*Z*	5.26	5.23	0.03	5.17	0.09
46	$4\text{-}OCH_3$	*E*	4.64	4.41	0.23	4.60	0.04
47	$4\text{-}OCH_3$	*Z*	5.43	5.55	−0.12	5.45	−0.02
48	$5\text{-}OCH_3$	*E*	5.32	5.09	0.23	5.23	0.09
49	$5\text{-}OCH_3$	*Z*	5.20	5.02	0.18	5.26	−0.06
50	5-OH	*E*	4.66	4.74	−0.08	4.76	−0.10

[a]Calculated from the non-crossvalidated conventional CoMFA model (five components).
[b]Calculated from the non-crossvalidated q^2-GRS CoMFA model (six components, 0.4 q^2 threshold).

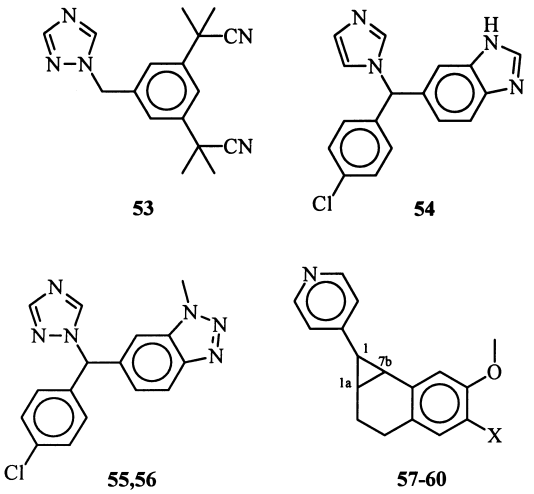


Figure 3. Structure formulae of the molecules of the test set.

The conformations generated by RANDOMSEARCH are minimized with MAXIMIN2 and they were further completely optimized by means of AM1.[24] Among the low energy conformations (within 5 kcal/mol from the global minimum) of each compound, one was chosen in the subsequent alignment procedure according to the criteria stated below.

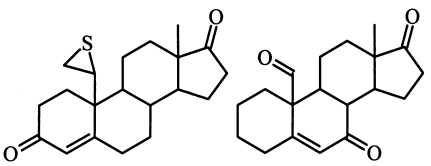
Alignment

The most critical point in the CoMFA procedure is the alignment of the molecules in the space, which is usually accomplished following some hypothesis of binding of the compounds to the site of action. In the azole model,[9] the starting assumption was that *S*-fadrozole shares two binding functions with the steroidal competitive inhibitor (19*R*)-10-thiiranylestr-4-ene-3,17-dione (**51**), that is, the iron-coordinating atom (the unsubstituted imidazolic nitrogen) and a H-bond acceptor group (the cyano function) mimicking the 17-keto function of steroids. Based on this hypothesis, first pointed out by Furet et al.,[11] and considering the rigid structure of *S*-fadrozole, all the other analogues were aligned on it atom-by-atom (i.e., searching the best fit between some atoms of *S*-fadrozole and the corresponding atoms of the molecules to be aligned). This alignment was not changed in the present study and details of it are reported in ref 9.

The most straightforward way to align the tetralone derivatives **31**–**50** seemed to superimpose them onto the same steroid **51** used for the azole series, considering the pyridyl nitrogen and the carbonyl group as the binding functions. However, such an alignment did not lead to any statistically significant result, that is, we could not obtain any 3-D QSAR model able to explain the variation of activity of the whole series of 49 aromatase inhibitors. Hartmann et al.[17] hypothesized that the OH or OCH_3 groups of the inhibitors could act as hydrogen bond acceptors in the same way as the 17-keto function of the steroids, and we tried a second alignment based on this theory. Again, no statistically significant relationship was obtained.

These preliminary negative results allow one to think that the tetralone derivatives, despite bearing the same interacting functions as the azoles, that is, an iron-coordinating atom and a hydrogen-bond acceptor group, do not bind in the same way to the enzyme. This means that, given the common anchoring point to the heme group, the H-bonding group interacts with different aminoacid residues of the protein. Such a hypothesis can also be supported by the observation that the tetralone derivatives as a series are definitely less active than the azoles, what might indicate a different binding mode to the enzyme.

In line with the strategy adopted by Furet et al.[11] for comparing the spatial orientations of fadrozole (**2**) and (19*R*)-10-thiiranylestr-4-ene-3,17-dione (**51**), we then looked for a steroid derivative that could justify a spatial orientation of the pharmacophoric functions of the tetralone derivatives different from that of the same functions of the azoles.



In a recent paper,[25] Numazawa et al. described a series of androstane derivatives inhibitors of aromatase, among which the androst-5-en-7,10,17-trione **52** showed a moderate activity ($IC_{50} = 60\,\mu M$). This compound, tested on human placental aromatase according to Thompson and Siiteri,[18] exhibits a competitive-type inhibition ($K_i = 13\,\mu M$) and was shown to be a mechanism-based ('suicide') irreversible inhibitor of the enzyme. This indicates that the compound forms a dissociable complex with the enzyme (binding non-covalently at the substrate site) followed by irreversible inactivation. Actually, the presence of the 7-keto group was not found necessary for a tight binding throughout the series of androstane analogues, even if it might be important in the formation of the reactive species.[25] However, considering the ability of **52** to bind at the aromatase active site, we assumed that this compound with its 7-C=O function can indicate the spatial orientation of a

second polar group involved in the binding of moderate to low affinity inhibitors at the enzyme active site and consequently we used it as a template on which to align the tetralone derivatives **31**–**50**.

The alignment was carried out by superimposing the pyridyl N of the tetralones onto the carbonyl O of the 10-oxo group and the C=O of the tetralones onto the 7-C=O of the steroid.

It might be observed that the oxygen atom of the steroidal 10-oxo function does not coordinate the heme iron as the tetralone pyridyl N does. On the other hand, it is generally accepted that an aldehydic intermediate is involved in the aromatization process of androgen substrates, that is centered on the attack of an iron-bound hydroperoxy group on the 19-C atom of the steroid.[3,4,26] The position of the carbonyl O can thus be assumed as representative of the position of a coordinating heteroatom in the close vicinity of the heme iron.

Another observation regards the oxygenated function of the tetralone derivatives used in the superimposition onto the 7-keto function of **52**. Again, we did not follow the Hartmann's hypothesis that considers the OH and OCH_3 groups as the H-bond acceptors involved in the ligand-enzyme interaction,[17] but used the tetralone C=O instead. This might be justified also on the basis of the rather weak effect of the former substitutions, as shown by compounds **31**–**34** and **44**–**50**.

The final overall alignment of all the 49 aromatase inhibitors was obtained via the superimposition of the two sets of molecules (**51** plus the azoles **2a**–**30** and **52** plus the tetralones **31**–**50**) by rms fitting the C atoms of the androstane nuclei of the two steroids. It resulted in an envelope of molecules orienting all the iron-coordinating atoms in the same region of space and the supposed H-bond accepting functions in two different regions, according to the location of either the 17-keto function of **51** or the 7-keto group of **52**, respectively. In Figure 4, the superimposition of the reference steroids **51** and **52** and the relative orientation of **2a** (an azole) and **31** (a tetralone) is shown.

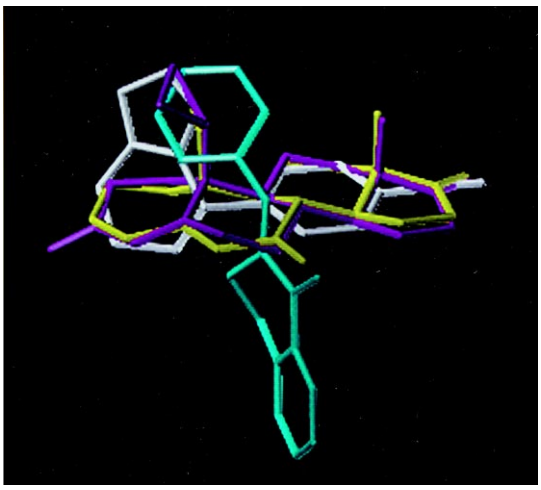


Figure 4. Superimposition of *S*-fadrozole **2a** (white) and the tetralone derivative **31** (cyan) onto the respective steroidal templates (19*R*)-10-thiiranylestr-4-ene-3,17-dione **51** (magenta) and androst-5-en-7,10,17-trione **52** (yellow).

The alignment of the molecules of the test set will be described and commented below (see Discussion).

CoMFA

A table was built containing the biological activity data of the aromatase inhibitors to be studied and the CoMFA columns. The latter were generated using an sp^3 C atom with a +1 charge as the probe.

For the conventional CoMFA, the region was created automatically and the default grid spacing was used (2 Å). The statistical analysis was performed by applying the PLS procedure to the appropriate columns and using the standard scaling method (COMFA_STD). Also, in order to reduce the actual number of column considered, energy cutoff values of 30 kcal/mol were selected for both electrostatic and steric fields and the minimum sigma value was set to 2.0. Crossvalidated PLS runs were carried out for establishing the optimum number of components to be used in the final fitting models. The number of crossvalidation groups was always equal to the number of compounds (leave-one-out technique) and the optimum number of components was chosen by considering the lowest standard error of prediction (s_{cross}).

In order to check the CoMFA and to ensure the reproducibility of the statistical results, we applied a method for region optimization recently developed by Cho and Tropsha.[27] This method, called cross-validated R^2-guided region selection (q^2-GRS), aims at creating a CoMFA region which is independent of the orientation of aligned molecules on the computer screen. Actually, it was noted that different orientations of the aggregate of aligned molecules lead to different q^2 values, resulting in poor reproducibility of the analyses.[27] The q^2-GRS routine is a process that, after breaking the original rectangular CoMFA region into 125 small boxes of equal size, performs a CoMFA for each of these newly generated regions and then selects those with a q^2 greater than a threshold value. The selected regions are combined to generate the master region on which the final CoMFA is performed.

In the q^2-GRS CoMFA, the energy cutoff values for the electrostatic and steric fields, as well as the minimum sigma were the same as in the conventional CoMFA (i.e., 30 kcal/mol and 2.0, respectively). The grid spacing of the 125 small boxes was 1.0 Å and we explored q^2 threshold values for selecting the individual small regions from 0.1 to 0.8.

Results

The results of both the conventional and the q^2-GRS CoMFA analyses for the training set are shown in Tables 2–4. In Table 2, standard errors and correlation coefficients associated to the crossvalidated PLS analyses of the conventional CoMFA calculated with numbers of components ranging from one to 10 are reported. Based on the criterion of the minimum s_{cross} value, the optimal number of components appears to be five.

As regards the statistics for the q^2-GRS analyses, the s_{cross} and the q^2 values for numbers of components ranging from one to 10 calculated at each of the q^2 threshold values from 0.1 to 0.8 are shown in Tables 3 and 4, respectively. It can be seen that the lowest standard errors were obtained with q^2 threshold values of 0.3–0.4 using six components. Actually, the differences between results obtained with 0.3 and 0.4 q^2 thresholds or with five or six components are quite small, such that the choice of one 'best' model is purely indicative.

In Table 5, the summary of the statistics of the selected conventional and q^2-GRS CoMFA models is reported,

Table 2. Statistical results of the conventional CoMFA for the training set

Components	1	2	3	4	5	6	7	8	9	10
s_{cross}	0.652	0.555	0.519	0.508	0.506	0.508	0.510	0.517	0.528	0.543
q^2	0.803	0.860	0.880	0.888	0.891	0.893	0.894	0.894	0.892	0.889
F	15.442	23.266	27.906	30.033	31.071	31.645	32.169	32.092	31.519	30.488

Table 3. Statistical results of the q^2-GRS CoMFA for the training set: s_{cross} values

Components	1	2	3	4	5	6	7	8	9	10
Threshold q^2										
0.1	0.513	0.512	0.520	0.524	0.637	0.552	0.532	0.525	0.519	0.512
0.2	0.638	0.554	0.524	0.515	0.506	0.504	0.505	0.507	0.521	0.524
0.3	0.634	0.557	0.526	0.515	0.503	0.500	0.507	0.507	0.514	0.524
0.4	0.634	0.558	0.529	0.518	0.506	0.499	0.504	0.505	0.516	0.520
0.5	0.635	0.560	0.531	0.521	0.509	0.505	0.512	0.511	0.520	0.526
0.6	0.628	0.559	0.545	0.532	0.519	0.526	0.526	0.526	0.532	0.538
0.7	0.595	0.560	0.577	0.594	0.613	0.636	0.639	0.639	0.638	0.639
0.8	0.583	0.545	0.581	0.587	0.620	0.637	0.659	0.644	0.651	0.654

Table 4. Statistical results of the q^2-GRS CoMFA for the training set: q^2 values

Components	1	2	3	4	5	6	7	8	9	10
Threshold q^2										
0.1	0.811	0.861	0.874	0.880	0.885	0.891	0.893	0.896	0.896	0.897
0.2	0.811	0.860	0.878	0.884	0.891	0.894	0.896	0.898	0.895	0.897
0.3	0.813	0.859	0.877	0.884	0.892	0.896	0.896	0.898	0.898	0.897
0.4	0.813	0.858	0.875	0.883	0.891	0.897	0.897	0.899	0.897	0.898
0.5	0.812	0.857	0.874	0.882	0.890	0.894	0.894	0.897	0.896	0.896
0.6	0.817	0.858	0.868	0.877	0.885	0.885	0.888	0.891	0.891	0.891
0.7	0.836	0.857	0.852	0.846	0.840	0.832	0.834	0.838	0.843	0.846
0.8	0.842	0.865	0.850	0.850	0.837	0.831	0.824	0.836	0.836	0.839

Table 5. Summary of conventional and q^2-GRS CoMFA analyses

	Conventional	q^2-GRS[a]
q^2	0.891	0.897
s_{cross}	0.506	0.499
F	31.07	32.97
r^2	0.991	0.996
s	0.146	0.100
Optimal no. of components	5	6
Steric contributions	0.456	0.445
Electrostatic contributions	0.544	0.555
r^2_{pred}	0.721	0.636

[a]q^2 threshold value: 0.4.

showing a substantial agreement between the two kinds of analysis. In both cases, the crossvalidated PLS runs gave almost equal values for q^2 and s_{cross}, while in the non-crossvalidated analyses, the standard error from the q^2-GRS CoMFA is lower than that from the conventional analysis. This indicates that the former model is able to reproduce more precisely the observed activity values, which might also be due to the greater number of components and to the much higher number of lattice point, that is, of variables (1210 in the conventional CoMFA, 4400 in the q^2-GRS CoMFA at 0.4 q^2 threshold).

In order to evaluate the ability of the CoMFA models to recalculate the original target values, i.e., to describe the three-dimensional structure–activity relationships of the whole series of aromatase inhibitors, one can observe the pIC_{50fit} and the Δ values reported in Table 1. The experimental inhibition data covering a range of 4.5 log units ($pIC_{50obsd} = 4.60$–9.15) are calculated with an average error of ± 0.28 by the conventional CoMFA model and of ± 0.18 by the q^2-GRS CoMFA. The plot of Figure 5 illustrates the relationship between observed and calculated data for the latter model.

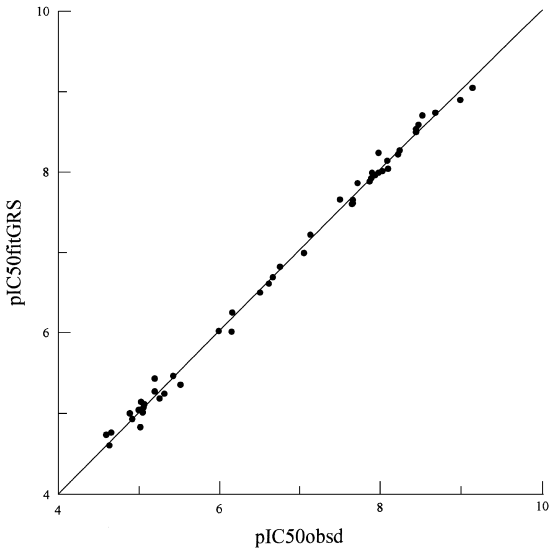


Figure 5. Plot of the calculated versus observed pIC_{50} from the non-crossvalidated q^2-GRS CoMFA model.

As regards the predictive ability of the model, it can be assessed by calculating the activity values of a test set of molecules which were not used in the derivation of the model itself. Compounds **53**–**60** (Figure 3) were used as the test set and their pIC_{50} values predicted by the selected CoMFA models are reported in Table 6. The pIC_{50pred} values are reasonably close to the respective pIC_{50obsd} values, with the exceptions of those for compounds **58** and **59**, which are greatly overestimated. A predictive r^2 (see ref 28) can be calculated for the test set by means of the expression

$$r^2_{pred} = (SD - PRESS)/SD$$

where SD is the sum of the squared deviations of each observed activity value for each molecule of the test set (pIC_{50obsd} values of Table 6) from the mean of the observed activity values of the training set (pIC_{50obsd} values of Table 1) and PRESS is the sum of the squared

Table 6. Observed and predicted biological activities of the aromatase inhibitors of Figure 3 (test set)

Compd	Stereochemistry	pIC_{50obsd}	pIC_{50pred}[a]	Δ	$pIC_{50predGRS}$[b]	Δ
53	—	7.82	8.02	−0.20	7.95	−0.13
54	*R*	7.35	7.05	0.20	6.81	0.54
55	*S*	8.85	8.38	0.47	8.24	0.61
56	*R*	8.40	8.04	0.36	7.94	0.46
57	1*S*,1a*S*,7b*S*	7.52	7.87	−0.35	7.44	0.08
58	1*R*,1a*R*,7b*R*	5.00	5.91	−0.91	5.97	−0.97
59	1*S*,1a*S*,7b*S*	6.03	7.61	−1.58	7.81	−1.78
60	1*R*,1a*R*,7b*R*	5.55	5.70	−0.15	5.87	−0.32

[a]Calculated from the non-crossvalidated conventional CoMFA model (five components).
[b]Calculated from the non-crossvalidated q^2-GRS CoMFA model (six components, 0.4 q^2 threshold).

deviations between predicted and observed values (Δ values of Table 6) for the molecules of the test set. The resulting r^2_{pred} values were 0.721 for the conventional CoMFA model and 0.636 for the q^2-GRS CoMFA.

Finally, the results of the comparative molecular field analysis can be illustrated by plotting the contour maps calculated from the QSAR equation expressing the non cross-validated model. The isocontours generated by interpolating the STDEV*COEFF terms of the conventional CoMFA equation for both the steric and electrostatic contributions are shown in Figure 6. CoMFA contour plots are shown surrounding regions in the space around the templates **2a** and **31** where steric and electrostatic properties of the molecules influence the aromatase inhibitory activity. The steric contours are

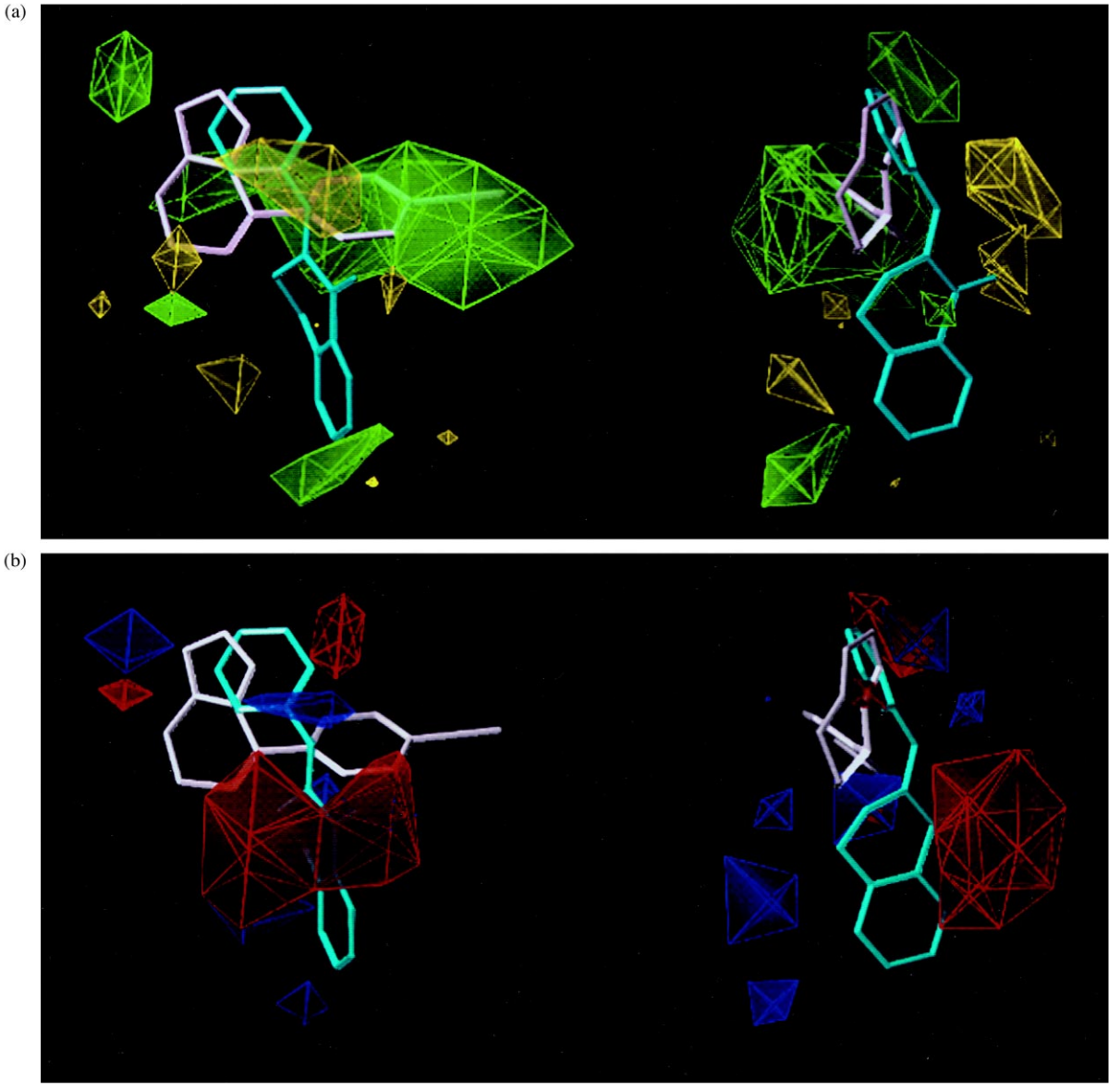


Figure 6. Orthographic views of the CoMFA contour maps: (a) steric contour: the region where increasing the volume increases activity is coloured green (0.024 level) and the region where increasing the volume decreases activity is coloured yellow (−0.023 level): (b) electrostatic contour: the red zone indicates increase of activity with increasing positive charge (0.030 level) and the blue contour indicates increase of activity with increasing negative charge (−0.030 level). The molecules of *S*-fadrozole (**2a**, white) and of the tetralone derivative **31** (cyan) are shown for reference.

colored green where addition of steric bulk increases the activity and yellow where an increase of the inhibitor's volume causes a decrease of activity; the electrostatic contours are red where activity increases with increasing positive charge and blue where an increase of negative charge is favorable.

Discussion

The application of the comparative molecular field analysis to a structurally varied set of 49 aromatase inhibitors (Figures 1 and 2, Table 1) resulted in a significant 3-D QSAR model, that was rendered reproducible, that is, independent of the orientation of the molecules envelope on the computer, by using a region selection procedure. This work extends our previous analysis[9] performed on a congeneric set of fadrozole (**2**) analogues and constitutes a further step towards a comprehensive 3-D QSAR model of the aromatase inhibitors.

The descriptive characteristics of the present model are good, as shown by the plot of Figure 5 and can be illustrated by the CoMFA contour maps of Figure 6.

As regards the steric contours [Figure 6(a)], they show a main region of positive (green) contribution to the activity that surrounds the fadrozole reference molecule. A great part of the green volume contains the phenyl ring of fadrozole and this is in agreement with the azole model,[9] where the *meta* and *para* positions of the same ring were indicated as suitable for steric substitution. Now, the green zone is more important and located around the azole reference molecule according to the higher inhibitory activity of the azoles with respect to the tetralones. The unfavorable (yellow) region of the CoMFA map is distributed in the space 'in front of' the molecules. In the previous model, it was centered over the benzylic C atom of the benzimidazole fragment of the azoles, in correspondence of the fourth substituent (H for fadrozole). In Figure 6(a), it seems to be split in some parts, probably due to fragments of the least active compounds among the azoles (**2b**, **6**) and of most of the tetralones. The relative contribution of the electrostatic field to the present model (Table 5) is higher than that to the azole model, and this lead to better defined electrostatic contours [Figure 6(b)]. The red volumes indicate regions around the molecules where increase of positive charge or, correspondingly, decrease of negative charge are preferred. The main red area seems an extension of the corresponding one revealed by the previous analysis[9] and justified by the presence of some alkyl residue. In the present case, it might be due also to the presence of both the phenyl rings of less active azoles and the 7-carbonyl O atoms of the tetralone derivatives. With respect to the whole series, these molecules are in the lower activity range and this might be the reason why these negatively polarized fragments appear to disfavor the inhibitory activity. Blue contours (negative polarization favored) appear mostly where azole fragments carrying π electrons are located.

The reliability of a QSAR model is usually assessed by testing its ability to predict the activity of new molecules acting on the same biological system. With this aim, we used the CoMFA models of Table 5 to calculate the aromatase inhibitory activity of eight molecules (Figure 3) already published, but not included in the training set. These compounds belong to both the azole (**53**–**56**) and the tetraline (**57**–**60**) series, but bear some structural characteristic that make them somewhat different from the molecules of the training set. As regards the alignment, the iron coordinating N atom was used as the anchoring center for all of them and one CN group of **53** was aligned to the 17-C=O of **51**. Compounds **54**–**56** show two possibilities for the H-bond accepting group, that is, the Cl atom and the -N=CH- or the -N=N- fragments of the benzazole ring. The electrostatic potential surfaces calculated for these molecules showed a negative potential area around the benzazole fragments comparable to that around the 17-C=O of **51** or the *para*-CN of fadrozole. The superimpositions of these groups allowed the *para*-Cl-phenyl rings of **54** and **56** to occupy a not unfavorable zone of space, while the same ring of **55** fell into the yellow negative steric region shown in Figure 6(a) and described above. This alignment lead to acceptable predictions of activity and to a correct ranking of the enantiomers **55** and **56** (Table 6). The enantiomeric pairs **57**, **58** and **59**, **60** show a tetralinic nucleus, but lack the carbonyl function characteristic of compounds **32**–**50**. We thus aligned them onto **51**, using the methoxy group as the H-bond acceptor. The ranking of the activity values was correct, even if the pIC_{50} values for **58** and **59** are rather mispredicted by both the conventional and the q^2-GRS CoMFA models (Table 6). This might be caused by an inherent weakness of the models or simply by the fact that such molecules as **57**–**60** exceed the parametric space explored by the training set. As a final observation regarding the predictivity of the 3-D QSAR models summarized in Table 5, it might be pointed out that the conventional CoMFA performs better than the q^2-GRS CoMFA (see the r^2_{pred} values). This might be a consequence of the region selection procedure: the resulting model is more finely tuned on the training set, but, on the other hand, tolerates less structural diversity in the molecules to be predicted than the conventional CoMFA model.

In a previous paper,[9] we tried to compare the three-dimensional results of our study with the aromatase model derived by Laughton et al.[29] The observations

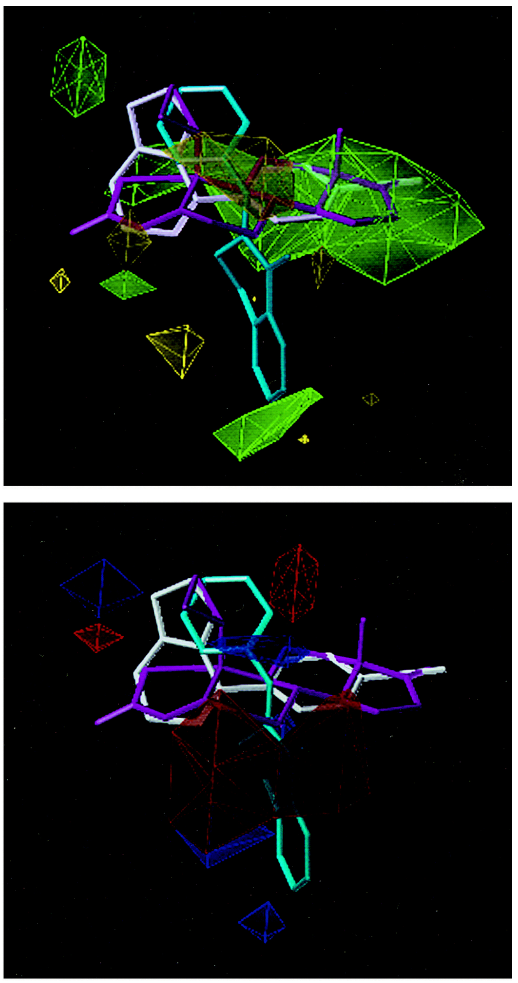


Figure 7. Views of the steric (a) and electrostatic (b) CoMFA contours shown with reference to **2a** (white), **31** (cyan) and **51** (magenta).

reported there seem to be confirmed by the extension of the investigation. Again, however, given the reliability of the 3-D QSAR model in terms of fit of the original data and predictivity, it would be desirable to refer its spatial characteristics, that is, the location of the steric and electrostatic contours, to the enzyme binding site. A detailed and generally accepted model of the cytochrome $P450_{arom}$ active site is still lacking, but an attempt could be done to compare the features of our model with the structure of a typical steroidal Type II aromatase inhibitor, like (19*R*)-10-thiiranylestr-4-ene-3,17-dione (**51**). We want to emphasize that such an extrapolation of the CoMFA results has to be done very carefully, in order to avoid an overinterpretation of the model, but, on the other hand, it might be a way to give a more general meaning to the structure–activity relationships. In Figure 7, the reference non-steroidal molecules **2a** and **31** are shown superimposed to the steroid **51**, together with the steric [Figure 7(a)] and electrostatic [Figure 7(b)] CoMFA contour maps. It can be noted that the positive steric volume (green) of the model surrounds the 'upper' edge and completely engulfs the D ring of the steroid, while the main negative yellow contour is located approximately over the β face of the 6 position of **51**. These observations can be compared with biological activity data of 16-α- and 17-α-steroids, which were reported to inhibit aromatase, even if with low affinity.[3,4] Moreover, Numazawa and Oshibe[30] showed the existence of a hydrophobic pocket of limited size in a region of the enzyme active site corresponding to the β-side of the steroid 6-position. As regards the position of the electrostatic contours with respect of the steroid **51** [Figure 7(b)], the red region indicating unfavorable effects by negatively charged functions contours an area where carbonyl groups of the 6-keto- and 7-keto-steroids are located. These compounds are known to be less active than the corresponding non oxygenated compounds.[3,24]

Conclusion

In conclusion, we have built an extended CoMFA model that accounts for the three-dimensional structure-activity relationships of two structurally different series of non-steroidal aromatase inhibitors. The goal was reached by aligning the compounds onto two steroidal inhibitors, thus hypothesizing two different binding modes for different classes of non-steroidal compounds. The model shows good descriptive and predictive capabilities and a region selection procedure was applied for ensuring its reproducibility. The results of the 3-D QSAR analysis, in terms of the individuation of regions in the space around the non-steroidal molecules where steric or electrostatic modifications can cause variation of activity, are in reasonable agreement with the SAR of steroidal inhibitors. Finally, this work shows the reliability of the CoMFA approach in the building of 3-D QSAR models for non-congeneric series of biologically active compounds.

Acknowledgements

The authors thank Dr S. J. Cho for supplying the script of the q^2-GRS routine. This work was supported by a grant of MURST.

References

1. Van Wauwe, J. P.; Janssen, P. A. J. *J. Med. Chem.* **1989**, *32*, 2232.

2. O'Reilly, J. M.; Brueggemeier, R. W. *Curr. Med. Chem.* **1996**, *3*, 11.

3. Banting, L.; Nicholls, P. J.; Shaw, M. A.; Smith, H. J. *Progr. Med. Chem.* **1989**, *26*, 253.

4. Banting, L. *Progr. Med. Chem.* **1996**, *33*, 147.

5. Browne, L. J.; Gude, C.; Rodriguez, H.; Steele, R. E.; Bhatnagar, A. *J. Med. Chem.* **1991**, *34*, 725.

6. Cheng, X.-M. *Annu. Rep. Med. Chem.* **1996**, *31*, 337.

7. Bayer, H.; Batzl, C.; Hartmann, R. W.; Mannschreck, A. *J. Med. Chem.* **1991**, *34*, 2685.

8. Thibaut, U. In *3D-QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 661–696.

9. Recanatini, M. *J. Comp.-Aided Mol. Des.* **1996**, *10*, 74.

10. Lang, M.; Batzl, C.; Furet, P.; Bowman, R.; Häusler, A.; Bhatnagar, A. S. *J. Steroid Biochem. Mol. Biol.* **1993**, *44*, 421.

11. Furet, P.; Batzl, C.; Bhatnagar, A.; Francotte, E.; Rihs, G.; Lang, M. *J. Med. Chem.* **1993**, *36*, 1393.

12. Hartmann, R. W.; Bayer, H.; Grün, G. *J. Med. Chem.* **1994**, *37*, 1275.

13. Bayer, H.; Hartmann, R. W. *Arch. Pharm.* **1991**, *324*, 833.

14. Bayer, H.; Hartmann, R. W. *Arch. Pharm.* **1991**, *324*, 815.

15. Koymans, L. M. H.; Moereels, H.; Vanden Bossche, H. *J. Steroid Biochem. Mol. Biol.* **1995**, *53*, 191.

16. Vanden Bossche, H.; Moereels, H.; Koymans, L. M. H. *Breast Cancer Res. Treat.* **1994**, *30*, 43.

17. Hartmann, R. W.; Bayer, H.; Grün, G.; Sergejew, T.; Bartz, U.; Mitrenga, M. *J. Med. Chem.* **1995**, *38*, 2103.

18. Thompson, E. A.; Siiteri, P. K. *J. Biol. Chem.* **1974**, *249*, 5364.

19. SYBYL Molecular Modeling System (Version 6.22), Tripos Ass., St. Louis, MO.

20. Van Roey, P.; Bullion, K. A.; Osawa, Y.; Browne, L. J.; Bowman, R. M.; Braun, D. G. *J. Enzyme Inhib.* **1991**, *5*, 119.

21. Peeters, O. M.; Schuerman, G. S.; Blaton, N. M.; De Ranter, C. J. *Acta Cryst.* **1993**, *C49*, 1958.

22. Clark, M.; Cramer III, R. D.; Van Opdenbosch, N. *J. Computat. Chem.* **1989**, *10*, 982.

23. Saunders, M. *J. Amer. Chem. Soc.* **1987**, *109*, 3150.

24. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Amer. Chem. Soc.* **1985**, *107*, 3902.

25. Numazawa, M.; Mutsumi, A.; Tachibana, M.; Hoshi, K. *J. Med. Chem.* **1994**, *37*, 2198.

26. Cole, A. P.; Robinson, C. H. *J. Med. Chem.* **1990**, *33*, 2933.

27. Cho, S. J.; Tropsha, A. *J. Med. Chem.* **1995**, *38*, 1060.

28. Waller, C. L.; Oprea, T.; Giolitti, A.; Marshall, G. R. *J. Med. Chem.* **1993**, *36*, 4152.

29. Laughton, C. A.; Zvelebil, M. J. J. M.; Neidle, S. *J. Steroid Biochem. Mol. Biol.* **1993**, *44*, 399.

30. Numazawa, M.; Oshibe, M. *J. Med. Chem.* **1994**, *37*, 1312.