and its Illinois Division. A doctoral dissertation award by the NIU Graduate School is acknowledged by P.L.T. We thank Gill Gladding and the ADD Program, NINCDS, NIH, for providing the biological testing.

Registry No. 19, 4630-75-5; 1b, 98818-30-5; 2a, 6276-34-2; 2b, 363-27-9; 2c, 92697-39-7; 2d, 92697-40-0; 2e, 98735-82-1; 2f, 98735-90-1; 2h, 98735-91-2; 3a, 91642-30-7; 3b, 6278-76-8; 4a, 4624-95-7; 4b, 98735-92-3; 4c, 98735-81-0; 4d, 98735-93-4; 4f, 98735-89-8; 4g, 98818-29-2; 4h, 98735-96-7; 4i, 98735-83-2; 4j, 98735-94-5; 4k, 98735-95-6; 4l, 98735-84-3; 5, 28360-59-0; 6a, 98735-86-5; 6b, 98735-85-4; 7a, 98735-87-6; 7b, 98735-88-7.

Correspondence Analysis Applied to Steroid Receptor Binding

Jean-Christophe Doré,*† Jacques Gilbert,† Tiiu Ojasoo,‡ and Jean-Pierre Raynaud‡

CERCOA-CNRS, BP 28, 94320 Thiais, and Roussel Uclaf, BP 120.07, 75007 Paris, France. *Received February 12, 1985*

The relative binding affinities of 48 steroids for four classes of hormone receptor (progestin, PR; androgen, AR; glucocorticoid, GR; mineralocorticoid, MR) have been analyzed by correspondence analysis. The steroids were, for the most part, derivatives of nortestosterone, differing by their degree of unsaturation, by the presence or absence of a 17α-ethynyl group, and by the length of the C-13 alkyl substituent. Derivatives of norprogesterone were included as reference compounds. Distribution maps visualizing the results of the mathematical analysis revealed that the majority of the test steroids were within the zone of influence of AR and PR and had limited affinity for GR and MR. Overall lack of specificity and enhanced affinity for GR and MR were induced by increasing unsaturation and by the presence of a C-13 ethyl group. The general and specific conclusions of the analysis confirm and extend previous intuitive and partial interpretations of the data. Correspondence analysis, however, has the advantage of taking into account the sum total of the available information, without any preconceived notion of the relative importance of a specific structural feature or biological parameter and, furthermore, enables simultaneous representation on a single graph of the receptor and steroid fields. The present example demonstrates the use of this type of methodology in processing routine screening data involving multiple parameters.

In general, one of the first steps in the determination of the potential usefulness of a molecule as a drug is screening its activity with respect to a large number of biological targets in comparison to several reference substances or structural analogues. Thus, information on the molecule is contained in a plethora of data requiring careful analysis.

This analysis can take several forms:

(a) The intrinsic complexity of the relationships among the different parameters can be partially disregarded, and the results can be analyzed methodically by, for instance, establishing maxima and minima and comparing rows and columns. The data are subjected to the rational reasoning of the experimentator whose approach, however logical, nevertheless remains intuitive.

(b) Several authors have described multiparametric models for structure–activity relationship studies with the aim of designing new drugs with preselected characteristics.[1] The common feature of their approaches is the search for the molecular descriptors (structural fragments, quantum chemical indices, physicochemical properties ...) that can discriminate among various activities.

(c) Our approach attempts to analyze the data in their overall complexity without any preconception of the particular importance, for instance, of a structural descriptor as opposed to any other feature of the molecule. The molecules position themselves relative to each other on confronting the chemical and biological fields, and the resultant configuration is analyzed. This analysis therefore does not tend toward a knowledge of the activity of the nth + 1 molecule but toward an understanding of how the molecules relate to each other with regard to the biological activities under study.

To find out how the items of the two fields are arranged in an n-dimensional system, we used factorial analysis but, in order to conserve the probabilistic nature of the system, we adopted correspondence analysis (CA) using χ^2 metrics[2] instead of principal component analysis (PCA) using a covariance matrix. The advantages of CA are at least threefold: (a) CA confers the same importance to the two fields (rows and columns), which can therefore be represented together on the same distribution map. In this way, observations can be correlated directly. (b) CA enables each representative point to be broken down into its constituent parts (principle of distribution equivalence) without affecting the factors. (c) In comparison to other factorial methods, CA does not overemphasize the elements that carry the most weight in the system. The use of this approach in the study of structure–activity relationships is licit in so far as the activity K_{ij} of molecule i in relation to biological parameter j is a function of the frequency of association between ligand and receptor. The outcome of the analysis takes the form of one or more distribution maps confronting the various items. These maps are read by the scientist in conjunction with the mathematical parameters necessary for their interpretation.

We have already used this type of approach in a few specific cases[3] but shall describe its full import in this paper by illustrating its application to the study of the

† CERCOA-CNRS.
‡ Roussel Uclaf.

(1) Franke, R. "Theoretical Drug Design Methods"; Elsevier: Amsterdam, 1984. Mager, P. P. "Multidimensional Pharmacochemistry: Design of Safer Drugs"; Academic Press: New York, 1984. Mardia, K. V.; Kent, J. T.; Bibby, J. M. "Multivariate Analysis"; Academic Press: New York, 1979.
(2) Hill, M. O. *Appl. Statist.* **1974**, *23*, 340. Benzēcri, J. P. L'analyse des Donnēes. II L'Analyse des Correspondances, Vol. 2, Dunod, Paris, 1983.
(3) Dorē, J. C.; Miquel, J. F. *C.R. Hebd. Seances Acad. Sci.* **1981**, *293*, 1061. Dorē, J. C.; Gilbert, J.; Crastes de Paulet, A.; Michel, F.; Miquel, J. F. *C.R. Hebd. Seances Acad. Sci.* **1982**, *294*, 731. Dorē, J. C.; Miquel, J. F.; Mrlina, G.; Calmon, J. P. *C.R. Hebd. Seances Acad. Sci.* **1983**, *297*, 125. Labia, R.; Morand, A.; Verchēre-Bēaur, C.; Dorē, J. C. *J. Antimicrob. Chemother.* **1983**, *11* (Suppl. A), 147. Dorē, J. C.; Marçot, B.; Pillon, D.; Viel, C. *C.R. Acad. Agri. Fr.* **1984**, *70*, 649.

0022-2623/86/1829-0054$01.50/0

interactions between a variety of steroid ligands and five hormone receptors. Over the last 15 years Roussel-Uclaf has screened in vitro over 1000 steroids for binding affinity for the estrogen, progestin, androgen, glucocorticoid, and mineralocorticoid receptors in order to find highly selective ligands and eventually specific drugs with minimal hormonal side effects.[4] In one particular study, the effect of the introduction either of methyl groups, of an ethynyl group, or of conjugated double bonds (or the combined introduction of several of these features) into the nortestosterone molecule was examined.[5] Several of the compounds thus obtained have been used as contraceptive or anabolic agents.[6] We shall in this paper illustrate the application of our factorial analysis to these data.

A brief outline of correspondence analysis is as follows. The n observations (corresponding to the 48 steroids, $n = 48$) are distributed in the *receptor field* R^p ($p = 4$, i.e. PR, AR, GR, MR). The position of each steroid within this field is defined by the ratio $f_{ij}/f_{i\cdot}$, where the relative frequency of binding between the steroid i and the receptor j is $f_{ij} = k_{ij}/\sum_{ij}k_{ij}$ and where $f_{i\cdot}$ corresponds to the marginal relative frequency of the steroid i for the various receptors $f_{i\cdot} = \sum_j f_{ij}$. Each point defining a steroid is thus endowed with an information content that encases responses to all four receptors (receptor profile). Two steroids will be close neighbors in this field if their binding profiles are comparable. In the *steroid field* (R^n) the scatter of the p receptor points is defined, by symmetry, by the formula $f_{ij}/f_{\cdot j}$.

To represent these two sets of points, with a minimum loss in information, principal projection axes are established as in principal component analysis by determining eigenvalues (λ) and eigenvectors (V_x). A symmetrical matrix is constituted of the distances $S_{jj'}$ between receptor pairs (χ^2 distance) which, in our example, is as follows:

$$R = \{S_{jj'}\} = \left[\sum_{j=1}^{n} \frac{1}{f_{i\cdot}} \frac{f_{ij}f_{ij'}}{\sqrt{f_{\cdot j}f_{\cdot j'}}}\right] = \begin{bmatrix} 0.652 & 0.299 & 0.164 & 0.168 \\ & 0.506 & 0.108 & 0.132 \\ & & 0.142 & 0.103 \\ & & & 0.162 \end{bmatrix}$$

The calculation is performed by solving equations of the type $[R] - \lambda[x] = 0$ and $[R][V_x] = \lambda_x[V_x]$ (diagonalization of the symmetric matrix) and, in CA, is simplified by the fact that one of the sets of points is given by the matrix $[R] = [M][M]'$. The permutation of the indices is thus equivalent to transposing the matrix to the other set of points $[M]'[M]$ with the same eigenvalues as R.

The coordinates of the various receptor points for the factorial axes are easily deduced from these parameters by

$$\varphi_{\alpha j} = \lambda_\alpha^{1/2} V_{\alpha j}/f_{\cdot j}^{1/2}$$

where the coordinate φ of receptor j for the factorial axis α is equal to the square root of the corresponding nontrivial eigenvalue λ_α multiplied by the corresponding eigenvector $V_{\alpha j}$ divided by the square root of the marginal relative frequency of receptor j toward the various steroids. Transition formulas, whose symmetry indicate that the two fields are represented equivalently, give the coordinates of the steroids for the different factors.

$$\varphi_{\alpha i} = (1/\lambda_\alpha^{1/2}) \sum_{j=1}^{p} (f_{ij}/f_{i\cdot})\varphi_{\alpha j}$$

$$\varphi_{\alpha j} = (1/\lambda_\alpha^{1/2}) \sum_{i=1}^{n} (f_{ij}/f_{\cdot j})\varphi_{\alpha i}$$

The end result is a projection of the different points onto the factorial planes that have been constituted by taking the axes 2 × 2.

To facilitate interpretation of the results, two parameters are defined: (a) Absolute contributions (AC) denote the contribution of an item of the two fields to the variance explained by a factor. The ACs are calculated as follows: for the steroid field $AC_\alpha(i) = f_{i\cdot}\varphi^2_{\alpha i}/\lambda_\alpha 100$; for the receptor field, $AC_\alpha(j) = f_{\cdot j}\varphi^2_{\alpha j}/\lambda_\alpha 100$. The sum of the absolute contributions of all the items in both fields for the α-axis is 100%. (b) Relative contributions (RC) correspond to the correlation item/factor and denote to what extent a factor explains the dispersion of an item (steroid or receptor). These parameters are calculated by the formulas below that evaluate the distances d from the centers of gravity of the two sets of points (G for the steroids and H for the receptors): for the steroid field, $RC_\alpha(i) = \varphi^2_{\alpha i}/d^2_p(i,G)$; for the receptor field, $RC_\alpha(j) = \varphi^2_{\alpha j}/d^2_n(j,H)$. They correspond to the square of the cosine of the steroid i or the receptor j for the α-axis. The sum of the relative contributions of each point of a same field to the various factors is thus equal to 1.

For further details on this type of approach, which are beyond the scope of the present paper, the reader is referred to more specialized publications.[2]

The present calculations were performed on a microcomputer (16/32 bits of 655 kbytes of central memory, Hewlett-Packard 9836) using an analysis of correspondence program rewritten by us in BASIC on the basis of a Fortran ANACOR software.[2] The factorial maps were drawn directly on a digital plotter with a precision of $^1/_{100\text{th}}$ in. (but have been redrawn by a professional artist for the purposes of this article).

Results

Experimental Data. Table I gives the relative binding affinities (RBAs) of the test steroids for the estrogen (ER), progestin (PR), androgen (AR), glucocorticoid (GR), and mineralocorticoid (MR) receptors. The molecules were chosen from the Roussel-Uclaf chemical bank and display a certain homogeneity since the majority differ by only a single stepwise substitution. No molecules were expressly synthesized for the study. A quick glance at the results shows that none of the steroids bind to the estrogen receptor—these data were therefore not included in the analysis—and that the sample reflects a population where androgen and progestin binding predominate over glucocorticoid and mineralocorticoid binding.

Correspondence Analysis. The original data matrix of the RBAs of the 48 steroids for PR, AR, GR, and MR

(4) Raynaud, J. P. In "Medicinal Chemistry"; Mathieu, J., Ed.; Elsevier: Amsterdam, 1977; Vol. V, p 451. Ojasoo, T.; Raynaud, J. P. *Cancer Res.* **1978**, *38*, 4186. Raynaud, J. P.; Ojasoo, T.; Bouton, M. M.; Philibert, D. In "Drug Design"; Ariëns, E. J., Ed.; Academic Press: New York, 1979, Vol. VIII, p 169. Raynaud, J. P.; Ojasoo, T. In "Steroid Hormone Receptors: Structure and Function"; Eriksson, H., Gustafsson, J. Å., Eds.; Elsevier: Amsterdam, 1983; p 141.

(5) Delettrē, J.; Mornon, J. P.; Lepicard, G.; Ojasoo, T.; Raynaud, J. P. *J. Steroid Biochem* **1980**, *13*, 45.

(6) Azadian-Boulanger, G.; Bucourt, R.; Nēdēlec, L.; Nominē, G. *Eur. J. Med. Chem.* **1975**, *10*, 353. Sakiz, E.; Azadian-Boulanger, G.; Ojasoo, T.; Laraque, F. *Contraception* **1976**, *14*, 275. Azadian-Boulanger, G.; Secchi, J.; Laraque, F.; Raynaud, J. P.; Sakiz, E. *Am. J. Obstet. Gynec.* **1976**, *125*, 1049. Grandadam, J. A.; Sheid, J. P.; Jobard, A.; Dreux, H.; Boisson, J. M. *J. Anim. Sci.* **1975**, *41*, 969.

(7) Ojasoo, T.; Raynaud, J. P. In "Steroids and Endometrial Cancer"; Jasonni, V. M., et al., eds.; Raven Press: New York, 1983; p 11. Delettrē, J.; Mornon, J. P.; Ojasoo, T.; Raynaud, J. P. In "Perspectives in Steroid Receptor Research"; Brescia-ni, F., Ed.; Raven Press: New York, 1980; p 1.

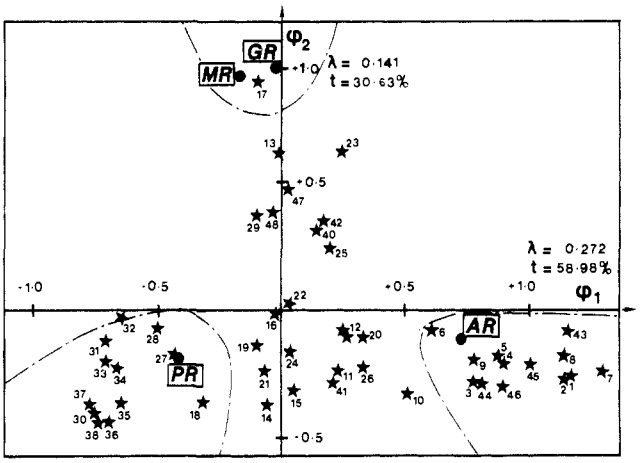


Figure 1. $\varphi_1\varphi_2$ distribution map of the 48 steroids in Table I: PR, progestin receptor binding; AR, androgen receptor binding; MR, mineralocorticoid receptor binding; GR, glucocorticoid receptor binding.

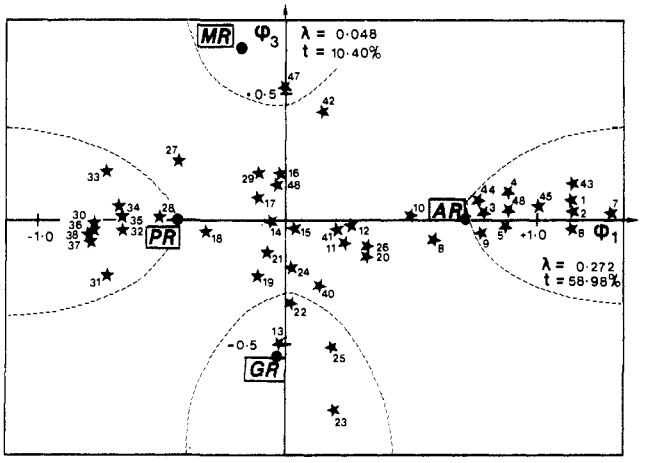


Figure 2. $\varphi_1\varphi_3$ distribution map of the 48 steroids in Table I.

was studied directly by correspondence analysis. The structure of the system thus formed was established and this 4-D system was reduced to several 2-D plots of two fields (molecules and receptors) depicting the scatter of the different items. This analysis sheds light principally on the specificity of binding of the test steroids with regard to the four receptors and with regard to each other, *regardless of the amplitude of the response* (affinity) which will be taken into consideration later.

Calculation of the three factorial axes φ_1, φ_2, and φ_3 showed that they account respectively for 59.0%, 30.6%, and 10.4% of the total variance of the system.

Organization of the Receptor Field. The main characteristic of the population under study is the dualism between progestin and androgen binding affinity (Table II) which constitutes more than 99% of the first factorial axis (φ_1) (i.e., sum of the absolute contributions (AC), 34.7 + 64.5). The relative contribution (RC) of androgen binding toward this axis ($\cos^2 \varphi_1$) is 0.98 and of progestin binding, 0.86. The second factorial axis (φ_2) is characterized to 86% by the corticoid nature of the molecules (glucocorticoid ($\cos^2 \varphi_1 = 0.76$) and mineralocorticoid (0.65)) and is opposed to 11% mostly by their progestin nature. The third factorial axis (φ_3), which accounts for $^1/_{10\text{th}}$ of the total variance opposes mineralocorticoid and glucocorticoid binding which amount to 99.7% of the variance of the axis.

Three different distribution maps may be considered: φ_1 vs. φ_2 representing 89.6% (59 + 30.6) of the total variance of the system (Figure 1), φ_1 vs. φ_3 representing 69.4%

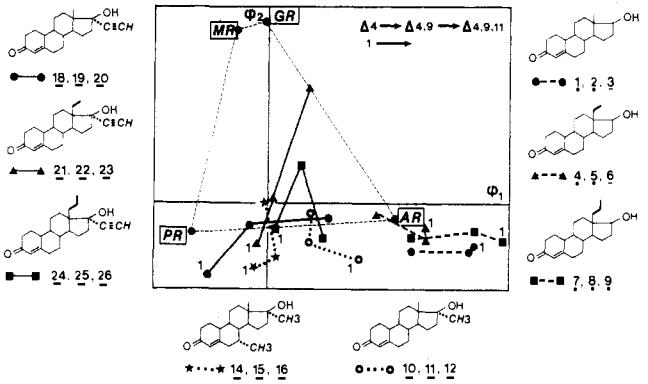


Figure 3. $\varphi_1\varphi_2$ distribution map of homologous unsaturated steroids (Δ^4 monoenes (numbered 1), $\Delta^{4,9}$ dienes, $\Delta^{4,9,11}$ trienes). Molecules belonging to the same series have been joined up in the order monoenes → dienes → trienes using the symbols indicated in the margin. The points have been transcribed from Figure 1.

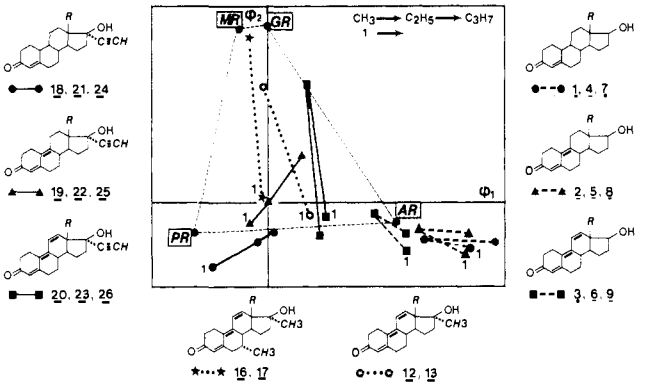


Figure 4. $\varphi_1\varphi_2$ distribution map of homologous C-13 alkyl substituted steroids (methyl (numbered 1), ethyl, propyl). Molecules belonging to the same series have been joined up in the order methyl → ethyl → propyl.

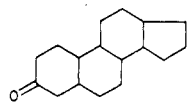
(59 + 10.4) of this variance (Figure 2), and φ_2 vs. φ_3 (41% = 30.6 + 10.4 of the variance; not shown).

Relationships between the Receptor and Molecular Fields. One of the aims of the analysis is to distinguish the molecules that are the most specific for a particular receptor in order to establish the structural features that they have in common. As seen in Figure 1, the molecules under study fall into several clusters: a group of 10 or so molecules that are predominantly progestins; a group of 10 or so molecules that are predominantly androgens; a more or less equivalent number of molecules with dual AR/PR specificity located beneath the φ_1 axis between the two poles; a few molecules that are rather more corticoid in nature. Just one molecule (17) can be considered truly corticoid, the others displaying a decidedly mixed (AR/PR/corticoid) tendency.

The least specific molecules within the trapezium defined by the four poles can be confronted by considering their relative positions on the $\varphi_1\varphi_3$ (Figure 2) and $\varphi_2\varphi_3$ (not shown) distribution maps.

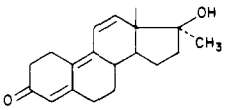
In order to spare the reader a rather fastidious description of the interest represented by each of the 48 molecules, we have extracted from Figure 1 two groups of molecules depicted in Figures 3 and 4. In Figure 3 are represented all molecules differing from each other by their degree of unsaturation; in other words, they are homologous Δ^4 monoenes, $\Delta^{4,9}$ dienes, and $\Delta^{4,9,11}$ trienes. Figure 4, on the other hand, depicts sets of molecules differing by the substituent in position C-13 (methyl, ethyl, propyl). We shall, thus, consider in turn the effects of increasing

Table I. Relative Binding Affinities



no.	Δ	C-13	C-17α	C-17β	ER[a]	PR	AR	MR	GR
1	4	CH_3		OH	<0.1	20	154	1.6	0.5 (2)
2	4,9	CH_3		OH	<0.1	17	134	0.3	1.6
3	4,9,11	CH_3		OH	<0.1	74	197	1.3	2.9
4	4	C_2H_5		OH	<0.1	34	126	4.7	2.1
5	4,9	C_2H_5		OH	<0.1	26	93	1.3 (2)	5.1
6	4,9,11	C_2H_5		OH	<0.1	86	172	5.6 (2)	20
7	4	C_3H_7		OH	<0.1	4.5	108	0.4	2.1
8	4,9	C_3H_7		OH	0.2 (2)	4.6	42	<0.1	1.9 (2)
9	4,9,11	C_3H_7		OH	0.1 (1)	38	105	0.7	6.0
10	4	CH_3	CH_3	OH	<0.1	100	146	0.6 (1)	1.5 (2)
11	4,9	CH_3	CH_3	OH	<0.1	71	64	0.4 (2)	6
12	4,9,11	CH_3	CH_3	OH	<0.1	208	204	18	26
13	4,9,11	C_2H_5	CH_3	OH	<0.1	230	143 (2)	36 (2)	155
14 (7α-CH_3)	4	CH_3	CH_3	OH	<0.1	214	108	2.1	1.4
15 (7α-CH_3)	4,9	CH_3	CH_3	OH	0.1 (1)	198	122	1.7 (2)	6.1
16 (7α-CH_3)	4,9,11	CH_3	CH_3	OH	0.1 (2)	306	180	52	22
17 (7α-CH_3)	4,9,11	C_2H_5	CH_3	OH	0.1 (2)	236	124 (2)	139	145
18	4	CH_3	C≡CH	OH	<0.1	156	43	0.2	2.7 (2)
19	4,9	CH_3	C≡CH	OH	<0.1	42	19	0.1 (2)	5.7
20	4,9,11	CH_3	C≡CH	OH	<0.1	63	70	1.8	11
21	4	C_2H_5	C≡CH	OH	<0.1	170	84	0.6	14
22	4,9	C_2H_5	C≡CH	OH	<0.1	68	41	0.4 (2)	17
23	4,9,11	C_2H_5	C≡CH	OH	<0.1	76	83	3.2 (2)	77
24	4	C_3H_7	C≡CH	OH	<0.1	73	44	0.1 (1)	9.8
25	4,9	C_3H_7	C≡CH	OH	0.22 (2)	11	10	<0.1 (1)	5.5
26	4,9,11	C_3H_7	C≡CH	OH	0.16	61	66	0.3 (1)	6.4
27	4	CH_3			<0.1	190	37	21	0.1
28	4,9	CH_3			<0.1	218	29	17	17
29	4,9,11	CH_3			<0.1	274 (2)	138	86	62
30	4	CH_3			<0.1	252	3.5	3.6 (2)	0.4 (2)
31	4,9	CH_3			<0.1	226	2.3	3.9 (2)	23
32	4,9,11	CH_3			<0.1	376	14	27	35
33	4	CH_3		$COCH_3$	<0.1	230	6.4	20 (2)	0.6 (2)
34	4,9	CH_3		$COCH_3$	<0.1	181	8.8	10	5.1 (2)
35	4,9,11	CH_3		$COCH_3$	<0.1	230	16	4.8	1.5 (2)
36	4	CH_3	CH_3	$COCH_3$	<0.1	317	5.5	1.5	0.7 (2)
37	4,9	CH_3	CH_3	$COCH_3$	<0.1	230	1.1	1.5 (2)	4.5
38	4,9,11	CH_2	CH_3	$COCH_3$	~0.1 (2)	220	1.9	0.7	2.4

struct modifcn to



no.		ER	PR	AR	MR	GR
39[a]		<0.1	208	204	18	26
40	3-deoxo	<0.1	0.7 (1)	0.6 (1)	0.1 (1)	0.3 (1)
41	no substituent at C-17	<0.1	9.9	8.5	0.1 (2)	0.5 (1)
42	C=O at C-17	<0.1	0.6 (1)	0.6 (1)	0.3 (1)	<0.1 (1)
43	2-*gem*-dimethyl	<0.1	1.0	14	0.8	0.4
44	6-*gem*-dimethyl	<0.1	69	178 (2)	3.0 (1)	1.2 (1)
45	2-methyl	<0.1	13	66	1.5 (2)	1.5
46	4-methyl	<0.1	27	89	0.2 (1)	0.5 (2)
47	A-nor	<0.1	229	166	127	38
48	2-oxo	<0.1	262	158	84	68

[a] ER, PR, AR, MR, GR: Estrogen, progestin, androgen, mineralocorticoid, and glucocorticoid receptors, respectively. The relative binding affinities (RBA) of estradiol, progesterone, testosterone, aldosterone, and dexamethasone for these receptors were arbitrarily taken as equal to 100. The RBA values for the test compounds are the means of three or more experiments unless otherwise indicated in brackets. [b] Note: compound 39 is the same as 12.

unsaturation and C-13 alkylation on receptor binding specificity before attempting to present an overall analysis.

The relative and absolute contributions of molecules 1–26 to the three factorial axes are given in Tables III and IV.

(a) Effects of Increasing Unsaturation. To interpret the data and analysis, it is necessary to confront the $\varphi_1\varphi_2$

Table II. Contributions to the Factorial Axes (Receptor Field)

	relative $\cos^2\theta$			absolute %		
	φ_1	φ_2	φ_3	φ_1	φ_2	φ_3
PR	0.8582	0.1411	0.0005	34.71	10.99	0.13
AR	0.9784	0.0210	0.0004	64.51	2.67	0.15
MR	0.0246	0.6498	0.3254	0.74	37.67	55.57
GR	0.0007	0.7640	0.2351	0.02	48.65	44.12

Table III. Contributions to the Factorial Axes (Molecular Field)

	relative $\cos^2\theta$			absolute %		
no.	φ_1	φ_2	φ_3	φ_1	φ_2	φ_3
18	0.4209	0.5669	0.0120	0.64	1.68	0.10
19	0.1506	0.2441	0.6052	0.02	0.07	0.56
20	0.7734	0.0831	0.1434	0.46	0.09	0.49
21	0.0653	0.7068	0.2277	0.04	0.88	0.83
22	0.0024	0.0047	0.9928	0.00	0.00	2.41
23	0.0439	0.3951	0.5608	0.32	5.69	23.81
24	0.0077	0.3886	0.6036	0.00	0.18	0.84
25	0.0982	0.1868	0.7148	0.02	0.10	1.15
26	0.6491	0.2887	0.0621	0.44	0.37	0.23
1	0.9419	0.0538	0.0041	7.10	0.78	0.17
2	0.9465	0.0525	0.0008	6.23	0.66	0.03
3	0.8825	0.1165	0.0008	5.35	1.36	0.02
4	0.9389	0.0499	0.0111	4.11	0.42	0.27
5	0.9606	0.0383	0.0010	2.97	0.22	0.01
6	0.9631	0.0160	0.0207	3.18	0.10	0.38
7	0.9684	0.0310	0.0004	5.95	0.36	0.01
8	0.9737	0.0253	0.0009	1.96	0.09	0.01
9	0.9350	0.0611	0.0037	2.84	0.35	0.06
10	0.7011	0.2983	0.0004	1.96	1.61	0.00
11	0.4502	0.4805	0.0691	0.23	0.48	0.20
12	0.8882	0.1092	0.0025	0.86	0.20	0.01
14	0.0286	0.9713	0.0000	0.03	2.59	0.00
15	0.0129	0.9751	0.0118	0.01	1.98	0.07
16	0.0129	0.0003	0.9866	0.00	0.00	3.21

Table IV. Contributions to the Factorial Axes (Molecular Field)

	relative $\cos^2\theta$			absolute %		
no.	φ_1	φ_2	φ_3	φ_1	φ_2	φ_3
18	0.4209	0.5669	0.0120	0.64	1.68	0.10
21	0.0653	0.7068	0.2277	0.04	0.88	0.83
24	0.0077	0.3886	0.6036	0.00	0.18	0.84
19	0.1506	0.2441	0.6052	0.02	0.07	0.56
22	0.0024	0.0047	0.9928	0.00	0.00	2.41
25	0.0982	0.1868	0.7148	0.02	0.10	1.15
20	0.7734	0.0831	0.1434	0.46	0.09	0.49
23	0.0439	0.3951	0.5608	0.32	5.69	23.81
26	0.6491	0.2887	0.0621	0.44	0.37	0.23
1	0.9419	0.0538	0.0041	7.10	0.78	0.17
4	0.9389	0.0499	0.0111	4.11	0.42	0.27
7	0.9684	0.0310	0.0004	5.95	0.36	0.01
2	0.9465	0.0525	0.0008	6.23	0.66	0.03
5	0.9606	0.0383	0.0010	2.97	0.22	0.01
8	0.9737	0.0253	0.0009	1.96	0.09	0.01
3	0.8825	0.1165	0.0008	5.35	1.36	0.02
6	0.9631	0.0160	0.0207	3.18	0.10	0.38
9	0.9350	0.0611	0.0037	2.84	0.35	0.06
12	0.8882	0.1092	0.0025	0.86	0.20	0.01
13	0.0001	0.6146	0.3852	0.00	13.00	24.01
16	0.0129	0.0003	0.9866	0.00	0.00	3.21
17	0.0148	0.9747	0.0103	0.25	31.86	1.00

distribution map (Figure 3), the table of relative contributions ($\cos^2\varphi$ of the angle made by each item with each factorial axis) and absolute contibutions (Table III), and the experimental data (Table I).

17α-Ethynyl Steroids. With a C-13 Methyl Group (18–20). According to Figure 3, the monoene (**18**) has the most selective affinity for PR. The introduction of a double bond in position Δ^9 (**19**) leads to a displacement toward the origin of the axes indicating increasing androgenic and corticoid components, whereas the third double bond (**20**) gives rise to an even further displacement but, this time, uniquely toward AR. According to Table III, most of the information is given by the third factorial axis (φ_3) for steroid **19**.

With a C-13 Ethyl Group (21–23). The Δ^9 double bond (**22**) induces a marked propensity toward corticoid binding, in particular GR binding, compared to steroid **21** (Figure 2) explaining the high relative and absolute contributions (Table III). The third double bond (**23**) emphasizes this phenomenon, but the relative contribution decreases since the PR/AR ratio changes.

With a C-13 Propyl Group (24–26). Once again, the introduction of a Δ^9 double bond (**25**) induces corticoid (GR) binding (Figure 2; Table III). The experimental data (Table I) in fact show that this enhanced GR binding is due to an overall decrease in affinity for all receptors, which is however somewhat less pronounced for GR. On the other hand, the third double bond (**26**) leads to a marked deviation toward AR binding (Figure 2) and the relative contribution toward the φ_1 axis is correspondingly increased (Table II).

17α-Unsubstituted Steroids. With a C-13 Methyl (1–3), Ethyl (4–6), or Propyl (7–9) Group. The majority of the information is given by the φ_1 axis (Table II) since, in comparison, none of these compounds has appreciable affinity for any receptors other than AR and PR (Table I).

17α-Methyl Steroids. Without a 7α-Methyl Group (10–12). Steroid **12** tends toward GR binding (Figure 3), but the contributions to the φ_1 axis (AR/PR) nevertheless predominate (Table III). According to Table I, the RBA values for AR and PR decrease in the case of steroid **11** but increase for **12**.

With a C-7 Methyl Group (14–16). The relative and absolute contributions of steroid **14** are nil for the φ_3 axis, weak for φ_1, and most pronounced for φ_2 which represents the corticoid nature of the molecule (Table III). The second double bond (**15**) does not greatly affect this pattern, whereas the third (**16**) results in a marked transfer of information to the φ_3 axis. According to Figure 3, this would be due to the influence of MR binding.

(b) Effects of Lengthening the C-13 Side Chain. 17α-Ethynyl Steroids. Monoenes 18, 21, and 24. Steroid **18** is the most PR selective of all those studied. If the C-13 methyl is replaced by an ethyl (**21**) or by a propyl (**24**), a progressive displacement toward AR and GR is recorded (Figure 4). This is reflected in the low φ_3 contributions of steroids **21** and **24** (Table IV). However, as Table I shows, the effect of the increased GR binding of **24** is partially due to an overall decrease in affinity for the other receptors.

Dienes 19, 22, and 25. The displacement from methyl to propyl via ethyl is virtually linear (Figure 4) and in the same direction as for the monoenes. The high relative contributions and higher absolute contributions for φ_3 indicate that the variation concerns primarily the corticoid component. As above, this increase in GR binding is explained by a decrease in affinity for the other receptors (Table I).

Trienes 20, 23, and 26. An ethyl in the place of a methyl (**23**) leads to a very strong attraction for GR, whereas a propyl (**26**) has little influence on the initial position (Figure 4). Table IV confirms that the ethyl group, unlike the propyl group, contributes strongly to φ_3.

17α-Unsubstituted Steroids. Monoenes 1, 4, and 7, Dienes 2, 5, and 8, and Trienes 3, 6, and 9. The pattern recorded for these three sets of steroids is highly similar (Figure 4). In each case, the φ_1 axis, i.e., the AR/PR dualism, is predominant (Table IV) and, in each case, the effect of the ethyl group is to decrease specificity by engendering propensity for GR (see steroid **6**). The effect of the propyl group is negligible.

17α-Methyl Steroids. Without a 7α-Methyl Group (12, 13). As above, the ethyl group engenders increasing selectivity for corticoid binding.

With a 7α-Methyl Group (16, 17). This selectivity is further increased by the introduction of a 7α-methyl group

into the C-13 ethyl molecule. The relative contribution of steroid **16** toward the φ_3 axis is very high (Table IV) as illustrated by its location at the origin of the $\varphi_1\varphi_2$ factorial axes. The ethyl group of steroid **17** considerably increases the φ_2 contributions and thus reveals marked duality between PR and AR, MR, and GR. The MR component would seem to predominate over GR according to Figure 4.

(c) Conclusions on Structure–Selectivity Relationships. Any conclusions drawn from the present analysis are valid for the population under study only. In particular, it should not be forgotten that most of the steroids are derivatives of nortestosterone which is a compound with primarily androgen specificity, with some affinity for the progestin receptor, and with negligible affinity for the glucocorticoid and mineralocorticoid receptors. In relation to its derivatives (Figure 1), it is located in the zone of high androgen selectivity (steroid **1**). The diene **2** has virtually identical selectivity; only compound **7** is a "purer" androgen. The remaining compounds can be considered derivatives of norprogesterone (**33**), a compound with high specificity for the progestin receptor albeit some affinity for the glucocorticoid receptor. This progestin specificity is enhanced in several derivatives (**38**, **36**, **30**, **37**, **35**) and is nearly equivalent in sultine analogues.

Figure 1 illustrates the dualism between androgen and progestin binding and shows that derivatives of nortestosterone such as **18** (norethindrone) can encroach upon the high-PR selectivity zone. A large number of compounds possess dual PR/AR specificity, but in a more or less equivalent number, the introduction of certain substituents has generated affinity for the glucocorticoid and mineralocorticoid receptors. Steroid **17** is the least specific whereas steroids **16** and **22**, at the origin of the $\varphi_1\varphi_2$ factorial axes, have in relative terms the most marked affinity for GR and MR (Table III). The analysis of the effects of unsaturation and C-13 homologation above has shown that either the introduction of two double bonds (Figure 3) or the presence of a C-13 ethyl (Figure 4) decreases specificity. These two features are combined in steroid **17**, which according to Table I, has high affinity for all four receptors.

Steroids **40**–**48** are variations on steroid **39** (identical with **12**). The introduction of methyl groups on ring A improves androgen specificity (steroids **43**–**46**, all within the AR zone). On the other hand, an A-nor ring, a 2-oxo group, or elimination of the 3-keto group decreases specificity by introducing a corticoid component (**47**, **48**, and **40** respectively). In the case of steroid **40**, this decreased specificity is accompanied by a marked loss in affinity for all four receptors. Modifications of the D-ring are represented by steroids **41** and **42**. The former retains a PR/AR dualism that is virtually lost in steroid **42**, but according to Table I affinity is low.

Figure 1 has established that the majority of the molecules are decidely within the zone of influence of the androgen and progestin receptors rather than the corticoid receptors. Figure 2 delimits the relative importance of glucocorticoid and mineralocorticoid binding. The cluster of molecules along the ordinate is distinguished less by their different AR and PR binding than by their binding to GR and MR; those at the center have either equivalent binding affinities for GR and MR or no affinity at all. Thus, molecule **47** is markedly more mineralocoticoid- than glucocorticoid-like; this also applies to molecule **42**, but examination of the original RBA data (Table I) indicates that this observation is of little significance since the RBAs for all four receptors are negligible. On the other hand,

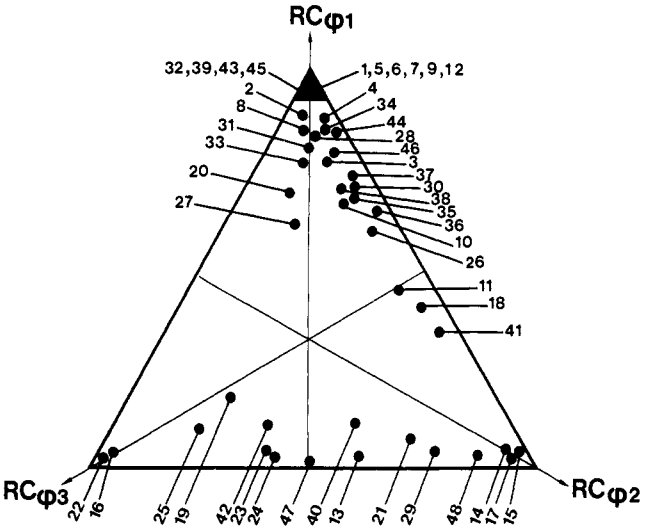


Figure 5. Distribution map of the 48 steroids in Table I. For each compound, the relative contribution to the φ_1, φ_2, and φ_3 axes is given by its perpendicular distance from the side of the triangle.

molecules **22**, **13**, and **23** are decidedly more glucocorticoid; inspection of the structures reveals that they are all C-13 ethyl substituted.

Another form of representation of the results is illustrated in Figure 5 where the position of each compound is determined by its relative contribution to each axis. This figure clearly shows that the majority of compounds are characterized by their AR–PR duality whereas only a couple, in particular compounds **16** and **22**, are distinguished, in a first instance, by their GR or MR binding.

(d) Relationships between Selectivity and Amplitude of Response. Up until now we have not taken into consideration the intrinsic levels of binding to the various receptors. In Figure 6, the absolute contributions have been plotted against the coordinates for the φ_1, φ_2, and φ_3 axes in order to illustrate the relationships between level of binding and specificity. The top panel of Figure 6 (φ_1 axis) reveals an interesting dissociation between high-specificity low-affinity (e.g., **43** and **8**) and high-specificity high-affinity (**1**, **2**, **7**) compounds. As regards the φ_2 axis (middle panel), **17** is the odd-man-out; it is the only compound that is preferentially a corticoid and also has a very high affinity. However, its dissociation between GR and MR binding is low (bottom panel φ_3), and **47** and **23** are more archetypal for appreciable MR and GR binding, respectively.

The same type of representation (Figure 7) has been used to highlight the level of response in the study of the effect of the C-13 alkyl substituent (Figure 4). This graph shows that only in the Δ^4 series (and not $\Delta^{4,9}$ or $\Delta^{4,9,11}$ series) does the propyl radical lead to enhanced affinity.

Discussion

The above example of the analysis of receptor binding data for 48 steroids illustrates some of the potential of correspondence analysis; all available information is taken into account; simple graphic representations of multiparametric data are possible (all possibilities have not been exploited in the present paper). The conclusions drawn are relevant to the population under study which, in this instance, has been restricted to molecules with a strong progestin and/or androgen bias. However, the method can be extended to a more vast population of molecules and battery of tests, on the condition that homogeneous comparable data are readily available and that the significance of the factorial axes can be ascertained. But increasing

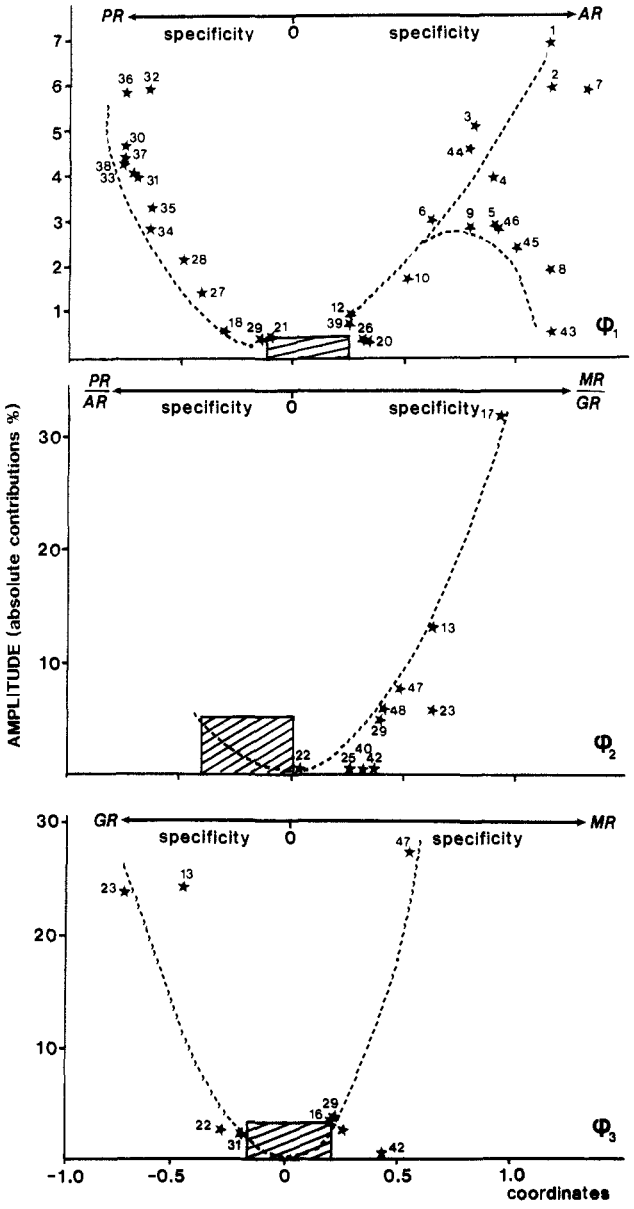


Figure 6. Illustration of the relationships between the amplitude of the response (as given by the percent absolute contribution) and the specificity of binding (as given by the coordinates for the φ_1, φ_2, and φ_3 factorial axes).

the number is not a prime objective since pertinent analyses of moderate populations should fulfil one of the objectives of this type of analysis, i.e., to avoid the systematic serial synthesis of analogues that yield redundant structure–activity information. Indeed, an important aim is to constitute a tool for the chemist in the interpretation of his results and thereby in his search for new lead compounds. For this, we are developing a program that will enable the chemist to converse directly with the computer and obtain immediate access to all the information that at present is given in the various tables and figures.

Before undertaking any mathematical analysis, several points have to be checked: (1) Are the data homogeneous? I.e., do they form a continuum of values? (2) Are they exhaustive? I.e. there should not be too many missing values. (3) Are they pertinent? I.e., does the analysis have a meaning? In the present example, the clear-cut nature of the observations is of particular interest since it reflects the high quality of the experimental results and confirms the aptness of the analysis. In terms of structure–affinity relationships, the analysis has elegantly confirmed and

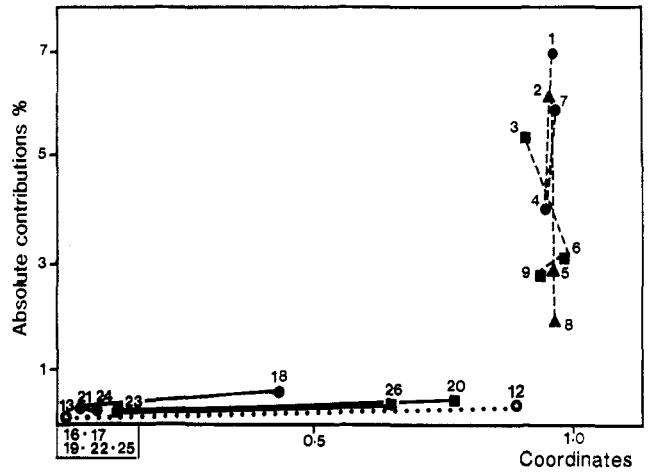


Figure 7. Relationship between amplitude of response and specificity for the C-13 alkylated steroids of Figure 4 (φ_1 axis).

simply illustrated the conclusions drawn previously by intuition and logic as regards the influence of unsaturation and of the C-13 alkyl substituent on binding.[4] However, in more sophisticated examples of available but partially unanalyzed data, it will be a useful means of addressing the following questions: (a) To what extent do experimental conditions influence the results and can data obtained under one set of conditions be extrapolated to another? (b) How similar are the receptors in different tissues and species? For instance, can the concept of a standard progesterone receptor be evolved? (c) What are the sensitive zones of a steroid in binding to the different receptors? (d) What is the geometry of the receptor binding sites? These questions have been touched upon on several occasions[6] but only a careful mathematical analysis will save us from obtaining redundant information in a large-scale screening system by indicating the new types of compounds that could be synthesized and tested.

Experimental Section

Relative Binding Affinity Determinations. The methodology has been described in detail elsewhere.[4] Briefly, each test steroid is incubated with a preparation containing the receptor under study ("cytosol") and a radioactively labeled marker known to bind selectively to this receptor. Bound radioactivity is separated by a dextran-coated charcoal adsorption method. The percentage of radioligand bound in the presence of test steroid compared to that bound in its absence is plotted against the concentration of competing test steroid. A standard curve for the competition of unlabeled radioligand is constructed with the use of 9–10 concentrations; five or six concentrations of each test steroid are used. From this plot, the molar concentrations of unlabeled radioligand or steroid competitor that reduce radioligand binding by 50% are determined. The effectiveness of the competitor is given by the ratio of the concentrations of unlabeled radioligand and of test steroid for 50% competition. This ratio multiplied by 100 is the relative binding affinity or RBA. The RBA of a reference hormone is arbitrarily taken as 100. In this study, the relative binding affinities were determined in at least 3 different experiments, and the mean RBA was calculated.

Registry No. 1, 434-22-0; 2, 6218-29-7; 3, 10161-33-8; 4, 793-55-5; 5, 4222-94-0; 6, 24957-63-9; 7, 898-62-4; 8, 19914-62-6; 9, 69065-05-0; 10, 514-61-4; 11, 14531-89-6; 12, 965-93-5; 13, 10109-62-3; 14, 3704-09-4; 15, 15506-10-2; 16, 10110-86-8; 17, 32516-32-8; 18, 68-22-4; 19, 14531-92-1; 20, 848-21-5; 21, 797-63-7; 22, 24881-79-6; 23, 16320-04-0; 24, 1044-96-8; 25, 852-78-8; 26, 10109-59-8; 27, 1722-54-9; 28, 6480-94-0; 29, 13885-21-7; 30, 74183-64-5; 31, 74183-65-6; 32, 74183-66-7; 33, 472-54-8; 34, 1231-97-6; 35, 2384-10-3; 36, 851-11-6; 37, 10116-22-0; 38, 2384-09-0; 40, 74135-05-0; 41, 74135-08-3; 42, 4642-95-9; 43, 23983-19-9; 44, 74135-07-2; 45, 98392-72-4; 46, 74135-06-1; 47, 58546-52-4; 48, 98304-58-6.