Pergamon

J Steroid Biochem. Molec. Biol. Vol 48, No 1, pp 31–46, 1994

0960-0760/94 $6 00 + 0.00

Affiliations Among Steroid Receptors as Revealed by Multivariate Analysis of Steroid Binding Data

T. Ojasoo,[1]* J. P. Raynaud[2] and J. C. Doré[3]

[1]*Groupe Cristallographie et Simulations Interactives des Macromolécules Biologiques, Université Pierre et Marie Curie (Paris VI),* [2]*ARIBIO and* [3]*CNRS URA 401 & Muséum National d'Histoire Naturelle, Paris, France*

To illustrate the informative value of descriptive multivariate analysis in biochemical screening, we have analyzed several data matrices relating to the binding of steroids to the estrogen, progestin, androgen, glucocorticoid and mineralocorticoid receptors in different organs and species. We first compared dendrograms of steroid hormone receptors, that were obtained by an automatic hierarchical classification analysis of the binding data, to published phylogenetic trees of nuclear receptors based on amino-acid sequence analysis. The former classification describes the affiliations among the receptors as given by the binding specificity of a population of 187 steroids in a traditional cytosol binding assay (an indirect comparison of ligand binding sites); the latter describes the affiliations among the receptors as given by a comparison of selected primary sequences involved in ligand-dependent regulation of transactivation and dimerization. A similar hierarchical classification was also performed on the binding data of 62 steroids to myometrium cytosol from different species in order to show to what extent the progesterone-binding proteins in these species are affiliated. Hierarchical clustering methods classify each type of variable (receptor or steroid) independently. In order to be able to correlate both types of variable (receptors and steroids) on single-display graphs, it is necessary to resort to correspondence factorial analysis (CFA). CFA ranks the information content within the experimental system, highlighting major correlations and disclosing secondary correlations by eliminating redundant information and background noise. This multivariate method, applied to the analysis of published data, illustrated the particular specificity of estrogen binding in human vagina and raised the question of the nature of the binding protein in this tissue. Our examples are based on small data tables that can and have been analyzed *de visu*. However, it is certain that such descriptive multivariate techniques are indispensable for the analysis of large data banks not only to define structure–activity relationships but to estimate the degrees of affiliation among the biological variables being measured. Knowledge of such affiliations will help to organize available information in a context where the complexity of the biological systems under study is becoming increasingly apparent.

J. Steroid Biochem. Molec. Biol., Vol. 48, No. 1, pp. 31–46, 1994

INTRODUCTION

There are two fundamentally different but complementary approaches to drug design, on the one hand, a divergent approach based on suitably selected screening tests and, on the other, a convergent approach based on original laboratory models that attempt to identify cause-to-effect relationships in cellular and molecular events. In the first case, the only hypothesis lies in the selection of the battery of tests and emphasis is on the systematic and comparative nature of the experiments. In the second, a specific hypothesis that needs to be verified is formulated at the very start of the study and emphasis is on the appropriate control of a maximum number of confounding variables that could interfere with the expected results.

It is our contention that the very large data matrices obtained in screening programs contain more information than is disclosed and that this information can be extracted from the data by appropriate multivariate

*Correspondence to T. Ojasoo at 4bis rue René Barthélémy, Montrouge 92120, France.
Received 15 July 1993, accepted 6 Sep. 1993.
Abbreviations: ER, estrogen receptor; PR, progesterone receptor, AR, androgen receptor; GR, glucocorticoid receptor; MR, mineralocorticoid receptor; TAT, tyrosine aminotransferase; RBA, relative binding affinity; CFA, correspondence factorial analysis; AC, absolute contribution; RC, relative contribution.

analysis. Screening programs involve the analysis of standard matrices comprised of rows of chemically-related compounds that have been tested on columns of biological variables. The usual question is what particular structural features of the compounds are related to which tests in order to be able to design more specific tools or drugs. However, the way in which the compounds react in the tests not only tells us something about the compounds but also yields precious information on the tests themselves. For instance, how the tests are related to each other, whether their information content is redundant. A multivariate analysis of such data would mean that industrial and service companies, for instance, could not only offer the scientific community new ligands of novel structure but knowledge of relationships among biological variables (without necessarily disclosing full structure–activity relationships). Such knowledge is necessary since, as investigations are pursued, the biological systems we study appear increasingly complex and since new information obtained by new techniques needs to be integrated into the existing patterns of our understanding.

The present review therefore illustrates how simple descriptive multivariate methods applied to the analysis of data matrices published by us and others can be used to classify biological variables into coherent meaningful pictures with visual impact such as dendograms or factorial maps. The examples we have chosen relate to our understanding of the nature of steroid binding to different classes of steroid hormone receptor in different organs and species. This work begun in the 1970s with an emphasis on steroid-structure and specificity [1, 2] will now focus on the information that can be obtained on the interrelationships among the receptors.

MATERIALS AND METHODS

The materials of the study are published data tables that give the relative binding affinities (RBAs) of populations of chemical variables (i.e. steroids) for binding to a number of biochemical variables (i.e. different cytosol steroid receptors). These tables of rows of steroids by columns of receptors were first converted into χ^2-distances and then analyzed either by an automatic classification procedure (hierarchical clustering) [3] or by correspondence factorial analysis (CFA) [4–8]. The former method treats steroids and receptors separately and gives independent diagrams representing the relationships among the rows and among the columns whereas CFA represents both steroids and receptors on a single-display map.

In each column, the RBA of the natural hormone for its corresponding receptor was taken as 100 (except for the glucocorticoid receptor where, unless otherwise specified, dexamethasone was used as a reference). The data therefore did not need to undergo any transformation prior to conversion into χ^2-distances.

We shall not dwell on the mathematics of the multivariate methods which have been published previously [3–8]. To understand the diagrams, it is simply necessary to remember that any multivariate method is just a means of comparing distances among items in an n-dimensional space. Proximity between items in a given plane implies affiliation between these items whereas distance implies diversity.

Agglomerative hierarchical clustering

Hierarchical trees [3], in which correlated variables are clustered beneath interconnected nodes of different heights, were constructed by applying to the χ^2-distance tables an algorithm for agglomerative hierarchical clustering that uses the aggregation criterion of Lance and Williams [9] with standard coefficients of $\alpha = 0.625$ and $\beta = -0.25$. In brief, the matrix formed by the RBA values of 'i' test-compounds for 'j' receptors is converted into a matrix of the distances that separate the receptors, taken two-by-two, when these are projected into the multidimensional space defined by the i test-compounds. These distances are arrayed into a symmetrical $j \times j$ semi-matrix. Within this distance matrix, the two closest receptors are united into a single group and the dissimilarity of this newly-formed group with each of the other receptors is calculated. The two closest receptors or groups of receptors are again united and the process is iterated a total of $j - 1$ times. The resultant hierarchical tree is a sequence of partitions of the receptors. The search for closest-members explains the terms 'single-link clustering' and 'nearest-neighbour method' used for this agglomerative clustering approach. In the present paper, we shall only describe the mergers operated among receptors, but the test-compounds can be clustered in a similar fashion.

Correspondence factorial analysis (CFA)

CFA [8] determines multiple relationships (correlations) among the steroids, among the receptors, and between steroids and receptors. It is based on the assumption that the information within the system must be partially redundant if some of the variables are highly correlated and reduces the n-dimensional space into more convenient optimal 1, 2 or 3D-spaces that best account for the total variance (information content) of the system.

In brief, the Eigenvalues and Eigenvectors of the χ^2-distance tables are calculated and yield the rank order of a series of factorial axes ($\phi_1, \phi_2, \phi_3 \ldots \phi_{n-1}$) accounting for the total variance. Plotting any two of these orthogonal axes gives factorial maps that display the projections of the cloud of points of the n-dimensional system. Clustering of points within the factorial maps reflects closeness of correlation. The first factorial map ($\phi_1\phi_2$), which is a plot of the two principal factorial axes ϕ_1 and ϕ_2, is defined by the most

discriminatory steroids and receptors and incorporates the major portion of the total variance. However, since these axes do not account for all the variance, i.e. the true points are not superimposed upon their projections on the $\phi_1\phi_2$ map, additional information is necessary to interpret the map correctly and is given by the absolute (AC) and relative (RC) contributions of each variable to the factorial axes, i.e. respectively, the extent to which an axis is representative of the variance of the system ($\Sigma AC = 100\%$) and the dispersion of a variable over all the axes (ΣRCs of each variable to all factorial axes = 1). Advantages of CFA are: (a) all types of variable can be licitly and simultaneously displayed on the same factorial maps (unlike in principal component analysis which uses covariance instead of χ^2-metrics), (b) whereas the principal axes highlight the most important correlations, the lower order axes disclose lesser correlations, that are no longer concealed by the main correlations and that are as meaningful, (c) small data tables can be analyzed.

Calculations and program availibility

Calculations were performed on a microcomputer (16–32 bits of 655K of central memory, Hewlett-Packard 9836) with a program adapted for BASIC (Microsoft Language) from FORTRAN Anacor software.

A simplified version of the CFA program for running on an IBM PC compatible computer is available upon request from J. C. Doré (Muséum National d'Histoire Naturelle, 63 rue de Buffon, 75005 Paris, France). CFA programs are commercially available from several sources including ADDAD (Association pour le Développement et la Diffusion des Données), Laboratoire de Statistique, Tour 45-55, 4 Place Jussieu, 75005 Paris; Professor M. J. Greenacre (SimCa version 2), P.O. Box 567, Irene, 1675 South Africa; BMDP Statistical Software Inc (PC-90 User's Guide 1990), Los Angeles; SPSS Inc (Categories Reference Guide 1990), Chicago; SAS Institute Inc (SAS/STAT User's Guide, Vol 1: ANOVA-FREQ, Version 6, 1990), Cary, NC.

RESULTS

It is not materially feasible to reproduce all the data tables we analyze in the present paper. For the crude data, the reader should refer to the original publications that are cited in reference.

Comparison of hierarchical classifications of receptors with evolutionary trees

The first published data table we shall consider [10] gave the RBAs of 39 steroids, mainly glucocorticoids and progestins, for the following cytosol receptors: progesterone receptor (PR) of rabbit uterus, androgen receptor (AR) of rat prostate, mineralocorticoid receptor (MR) of rat kidney and glucocorticoid receptor (GR) of rat thymus, rat liver, and rat hepatoma tissue culture (HTC) cells. A biological response variable, the induction of tyrosine aminotransferase (TAT) activity in HTC cells, was also included in the analysis. The hierarchical tree of these biochemical variables, receptors and TAT induction, which was obtained as described under Materials and Methods, is shown in Fig. 1(a). The ordinate represents the distance between groups in the same units as those of the distance matrix. The order of the variables along the abcissa is to some extent arbitrary since the tree has the degrees of freedom of a mobile, the only restriction imposed being that no branches may cross over. The method partitions the variables into a nested family of clusters and reveals that, *for this population of steroids that were primarily selected for their dual progestin/glucocorticoid specificity*, the GR from three different sources are very closely correlated among themselves and to TAT induction, which is therefore to be considered a glucocorticoid response; GR is most closely affiliated to PR; the group GR–PR is affiliated to MR. The most dissimilar receptor in its response to these steroids is AR; AR is closest to MR. This classification, first published in 1988 [10], summarized our perception of the steroid receptor specificity of progestins and glucocorticoids at that time.

The greater the number of randomly selected items (steroids) with respect to the number of biochemical variables (receptors), the more accurate and broadly applicable the classification. We therefore analyzed all steroid receptor binding data obtained under identical experimental conditions that we have published in the past (cumulated data on 187 steroids from Refs [1, 2, 8, 11–13]). The steroid population was primarily, but not uniquely, characterized by differences in ring saturation, alkylation, hydroxylation, and 17α-ethynylation. There were no very bulky substituents nor lengthy side-chains in strategic positions that could easily interact with amino-acid sequences outside the traditional hormone binding site. The analysis included several steroids with phenolic A-rings and also a measurement of RBA to the cytosol estrogen receptor (ER) of mouse uterus. As expected ER has the most different binding specificity [Fig. 1(b)]. MR and GR are affiliated as are PR and AR but, as indicated by the lower node height, the relationship between the second pair is closer than between the first for this population of molecules. However, since the stem above the node that groups MR and GR is very short, there is no great leap between the two pairs. This objective assessment by multivariate analysis confirms our past intuitive interpretations.

The relative proportions to which the individual receptors contribute to the overall system are as follows: ER (5%), PR (47%), AR (23%), GR (16%), MR (9%) indicating that there was a distinct bias toward PR due either to steroid availibility governed by ease of synthesis or to instinctive selection of molecules with a dominant progestin component. Increasing the steroid population even further to numbers such as found in industrial data banks and/or the use of techniques of random sampling can help to minimize this bias.

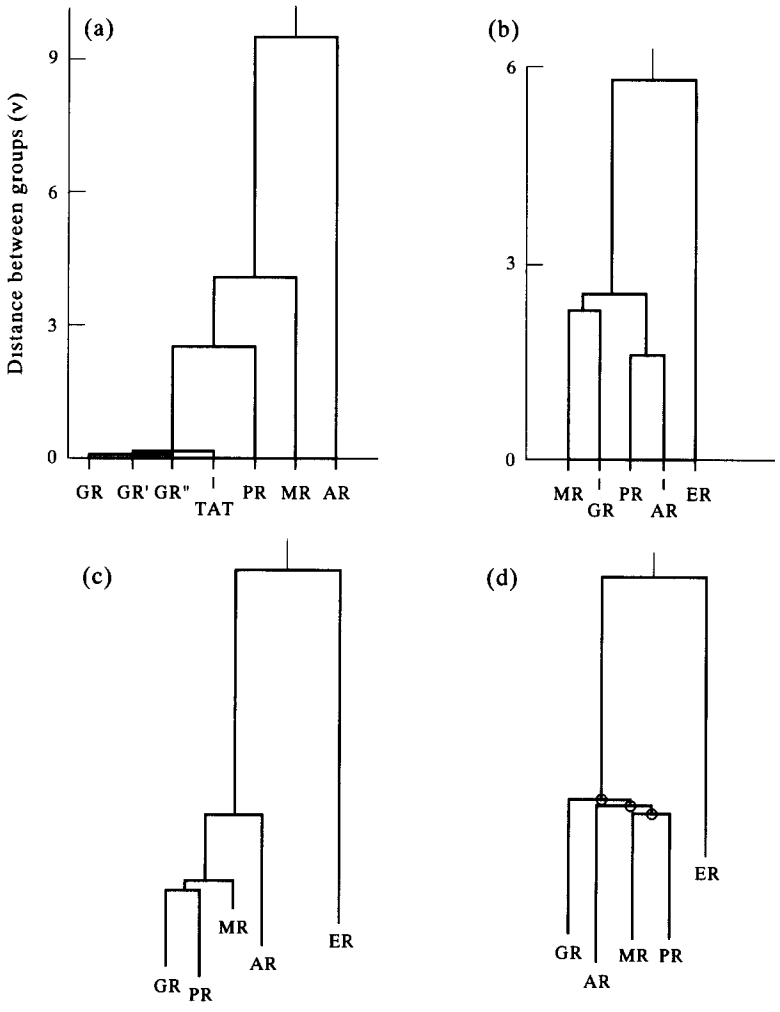

Fig. 1. (a) Hierarchical ascending classification of cytosol steroid receptors as given by an analysis of the binding profiles of a selected population of 39 glucocorticoids and progestins. Induction of TAT activity in HTC cells was also measured. The sources of the cytosol receptors were as follows (from left to right): GR (HTC), GR′ (rat thymus), GR″ (rat liver), PR (rabbit uterus), MR (rat kidney), AR (rat prostate) (classification reproduced from Ref. [10]). (b) Analysis of the binding profiles of 187 steroids of related chemical structure. Receptor sources: MR (rat kidney), GR (rat liver), PR (rabbit uterus), AR (rat prostate), ER (mouse uterus) (data from Refs [1, 2, 8, 11–13]). (c and d) Phylogenetic trees based on Fitch least squares analyses of the amino-acid sequences of the DNA-binding domains (c) and of the *trans*-activation and dimerization regions of the ligand-binding domain (d) of nuclear receptors according to Laudet *et al.* A neighbour-joining analysis gave a different classification (see encircled nodes) (adapted from Ref. [15]).

Despite these restrictions concerning the nature of the steroid population, the above classifications are nevertheless highly instructive since they illustrate how these steroids perceive different steroid receptors and can be compared to recent classifications of receptors on the basis of amino-acid sequence [14, 15]. The phylogenetic trees obtained by a Fitch analysis either of the DNA-binding domain or of part of the ligand-binding domain of nuclear receptors of different sources are reproduced in Fig. 1(c and d) (adapted from Ref. [15]). In the phylogenetic tree, the height of the branches is proportional to the number of mutations that have accumulated in the genes that code these receptors. In all classifications, the position of ER is remote from those of the receptors that bind 3-keto-4-ene steroids. Moreover, in all classifications, the internodal segments linking these receptors are very short. Indeed, on using a different analytical approach (neighbour-joining analysis), Laudet *et al.* [15] found permutations at the level of the encircled nodes and conclude that "the relationships between PR, MR and GR cannot be resolved unambiguously (trichotomy)".

In the absence of a mechanistic explanation, it is probably fortuitous that our classification, based on the receptor specificity of progestins and glucocorticoids and which emphasises the proximity between PR/GR/MR and the relative distance of AR [Fig. 1(a)], is so similar to the phylogenetic tree obtained by Fitch analysis of the DNA-binding domain [Fig. 1(c)]. On the other hand, the confrontation between Fig. 1(b and d) is more meaningful since there is some, although incomplete, complementarity in their information content. In one instance, we have a comparison of primary protein structures considered to have functional importance in ligand-dependent regulation

Table 1. RBAs for cytosol receptors (PR of rabbit uterus, AR of rat prostate, and GR of rat thymus) after incubation at 0°C for short and long incubation times

	PR		AR		GR	
	2 h	24 h	30 min	2 h	1 h	24 h
Testosterone (T)	1	1	100	100	3	0.5
5α-DHT	1.5	1	95	120	1	1
Danazol	9	3	35	8	<0.1	<0.1
17α-MethylT	3	2	90	45	2	1
19-NorT	20	15	230	155	4	0.5
Trenbolone	75	15	250	190	9	2.5
Norgestrienone (R2010)	65	45	95	70	55	15
Gestrinone (R2323)	75	50	95	85	265	150
Metribolone (R1881)	210	190	160	205	60	20
Ethynylnortestosterone (ENT)	155	265	75	45	20	5.5
17α-EthynylT (ET)	35	15	7	0.1	2	0.5
Cyproterone ac. (CPA)	80	60	50	15	9[a]	6
Megestrol ac. (MA)	150	120	65	20	90[b]	45
Chlormadinone ac (CA)	175	320	80	20	215	35
Progesterone (P)	100	100	20	5 5	115	75
RU25051	330	870	40	15	155	80
RU23747	225	380	3.5	2	55	25
RU22779	205	335	7.5	3.5	35	15
Demegestone (R2453)	230	420	7.5	1	40	10
Promegestone (R5020)	220	535	10	1.5	100	30
Nomegestrol ac. (NA)	170	450	20	1	20[a]	15
16α-MethylP	65	60	10	1	160	55
6α,16α-DimethylP	50	155	5	8	395	220
RU25253	180	530	40	6	770	330
Medroxyprogesterone ac. (MPA)	125	305	40	50	470	215
RU4841	230	675	60	135	380	245
Norgestrel	170	905	110	85	210	40
RU2999	260	305	205	160	85	55
RU2420	280	330	175	180	85	25
RU25593	40	35	6	5	390	180
RU25055	70	85	20	6	550	245
Corticosterone (B)	5	3	1.2	0.5	100	100
Cortisone (E)	<0.1	—	—	<0.1	1.5	1 0
16α-Methylprednisolone	<0.1	—	—	<0.1	135	150
21-Desoxydexamethasone	<0.1	—	—	<0.1	225	190
Cortexolone (S)	0.8	0.5	<0.1	0.3	40	15
Hydrocortisone (F)	<0.1	—	—	<0.1	45	60
Cortivazol	<0.1	—	—	<0 1	300	475
Cortivazol-21OH	<0.1	—	—	<0 1	310	530
Prednisolone	<0.1	—	—	<0.1	60	115
Betametasone	<0.1	—	—	<0.1	185	380
Dexamethasone (DXM)	0.4	<0.1	0.3	<0.1	165	455
Desoximethasone	25	15	2.5	1	345	990
RU23739	35	65	1.5	0.3	450	1410
Triamcinolone acetonide	15	12	0.2	<0.1	210	790
Triamcinolone	<0 1	—	—	<0.1	55	290
Fluocinolone acetonide	38	50	0.8	0.2	275	1250
RU38140	45	85	35	15[b]	360[a]	1300
RU38486	80	530	10	25[b]	465[a]	1365
RU28289	230	440	—	10[b]	450[a]	1365
Deoxycorticosterone (DOC)	26	18	10	2.5	65	110

[a]Incubation for 4 h; [b]for 24 h.
—: Not determined but considered equal to 0.1 in the multivariate analyses as were the <0.1 values.
Reproduced from Ref. [20].

of transcription (τ_1-domain) and dimerization; in the other, a comparison of the ways in which a substantial number of ligands perceive these receptors within a ligand binding assay. In the absence of the influence of confounding variables due, for instance, to environment, our approach could be interpreted as a comparison of the hormone-binding domains of these receptors under ligand-activating low-temperature conditions.

Influence of experimental conditions on receptor classifications

Although the initial screening system included only one set of incubation conditions, this was increased to

two on the basis of the observation that ligands that dissociate slowly from the cytosol receptor have RBAs that increase on prolonging incubation time and those that dissociate fast have decreasing RBAs [16]. We therefore performed a hierarchical classification on the basis of the RBAs obtained after short (1) and long (2) incubation times for an extended series of 51 progestins and glucocorticoids (Table 1). The hierarchical classification in Fig. 2 confirms the conclusions of Fig. 1(a), namely, that the affiliation between GR and PR is close and that AR is more remote and furthermore indicates that the two incubation times induce rather similar effects on steroid behavior toward these receptors since the heights of the nodes linking receptor pairs (1 and 2) are more or less the same. Incubation time is most influential in the case of GR. It is probable, but to be proven, that some, but no drastic, conformational change in receptor conformation occurs on ligand binding that is, however, less marked than that which occurs on exposure to temperature increases that induce activation [17].

Visualization by CFA of the steroids that enable receptor classification

Until now, we have only considered the hierarchical classification of receptors on the basis of steroid binding specificity. The steroids themselves can be similarly classified by hierarchical or non-hierarchical methods (e.g. minimum spanning trees) to highlight chemical affiliations (not shown) (for examples, refer to Refs [18–20]). However it is only a method such as CFA that can reveal the relationships among different types of variable (among receptors, among steroids, and between receptors and steroids) in single-display factorial maps. Unlike principal component analysis, CFA treats the rows and columns of the data matrix in a symmetrical manner so that it becomes totally licit to superimpose the steroids on the receptor map. This strategy enables correlations to be drawn directly among all types of variable.

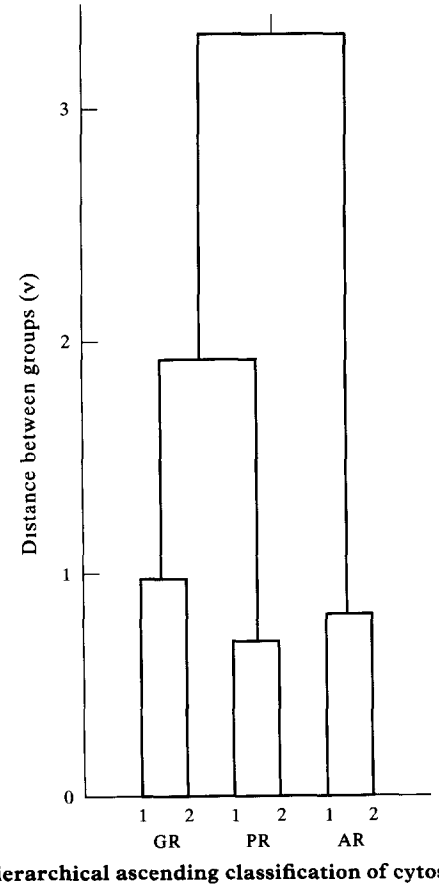

Fig. 2. Hierarchical ascending classification of cytosol steroid receptors on the basis of binding data obtained under two incubation conditions. Receptor sources: GR (rat thymus), PR (rabbit uterus), AR (rat prostate) (see Table 1).

CFA was thus used to convert the multidimensional RBA data table (Table 1, Fig. 3) into a ranked series of factorial axes of decreasing variance (information content) representing both steroids and receptors [20]. A very high proportion of the total variance (87.2%) is embodied in the first two factorial axes (56.7% for ϕ_1 and 30.5% for ϕ_2) which are plotted in the $\phi_1\phi_2$ factorial map of Fig. 4. This map, in which proximity reflects a correlation and distance reflects diversity, describes the main organization of the system and highlights its most discriminatory features.

In Fig. 4, the points representing steroids (stars) and receptors (open circles at short incubation times, solid circles at longer incubation times) are given by the projections onto the ϕ_1 and ϕ_2 axes of the true values in the multidimensional system. To grasp the meaning of these axes, it is helpful to refer to the absolute contributions (ACs) of the receptors to the making of the axes (Table 2). The ACs of the steroids are not given. The plus or minus sign in square brackets that precedes the AC indicates whether the coordinate of the projection of this variable onto the axis is positive or negative with respect to the origin. The variable PR ($PR_1 + PR_2$) contributes little (AC = 2.5%) to the main (ϕ_1) factorial axis which represents 56.7% of the information content of the system. The ACs of GR and AR are substantial and in opposition: [+]57.7%(18.5 + 39.2) for GR vs [−]39.8%(19.2 + 20.6) for AR. The ACs of AR are similar for both incubation times (about 20%) but the AC of GR doubles on increasing incubation time (from 18.5 to 39.2%). The ϕ_2 factorial axis, which represents 30.5% of the information content of the system, is characterized by an opposition of PR binding ([−]57.7% = 27.0 + 30.7) with AR binding ([+]31.0% = 13.1 + 17.9) and also with GR binding at long ([+]10.5%) but not short ([+]0.7%) incubation times. In Fig. 4, these oppositions are pictorially reflected in the triangular distribution of the receptors on the map. AR and GR are located on either side of the origin when considering the ϕ_1 axis whereas PR is situated in the bottom quadrants in contrast to AR and GR (ϕ_2 axis) and is more firmly anchored in the bottom left-hand than right-hand quadrant (ϕ_1 axis). The location of the receptors does not differ greatly with incubation time except in the case of GR, GR_1 being much closer to the ϕ_1 axis than GR_2.

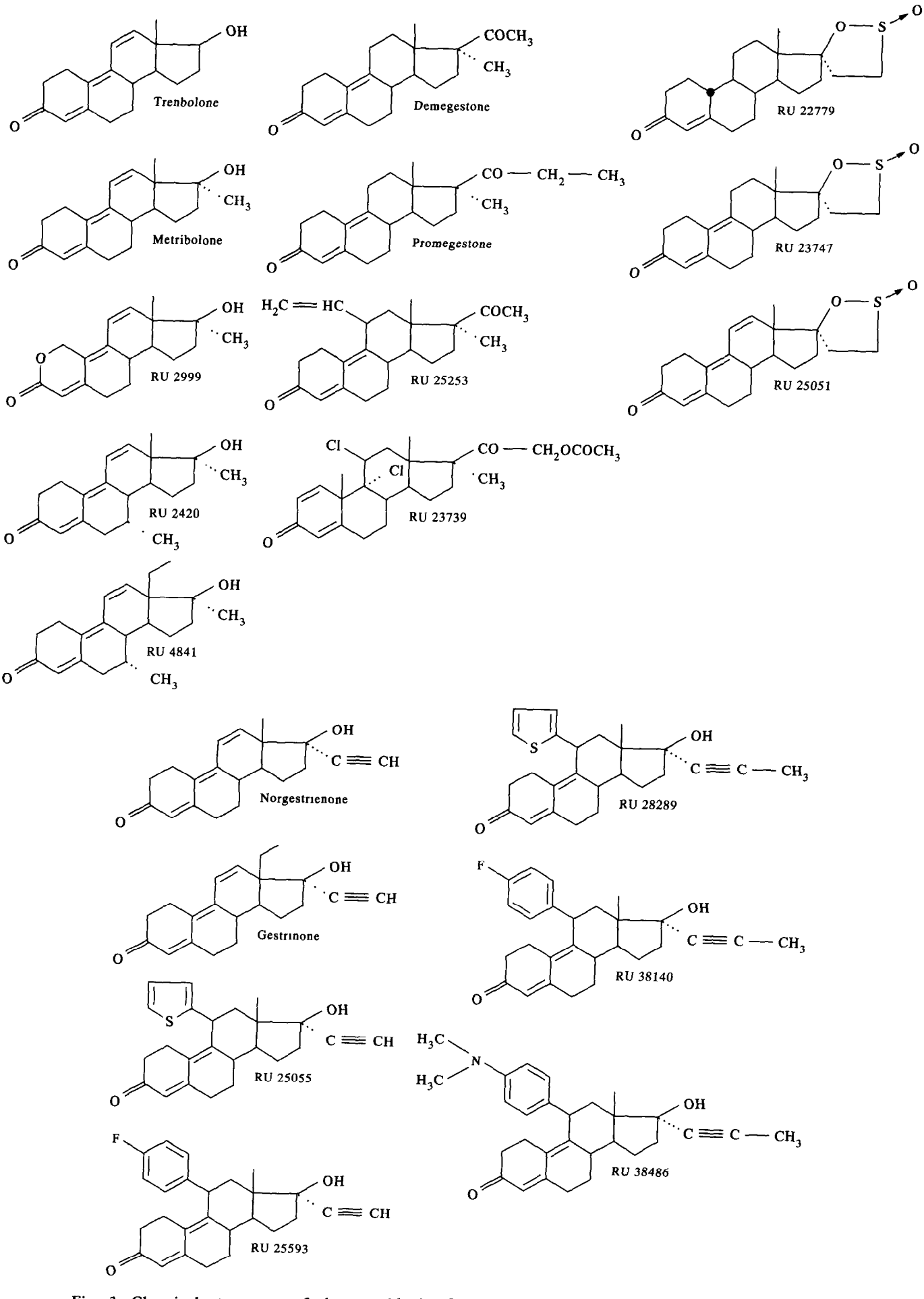

Fig. 3. Chemical structures of the steroids in Table 1 synthesized by the Roussel-Uclaf company. The structures of the other steroids are to be found in the Merck Index.

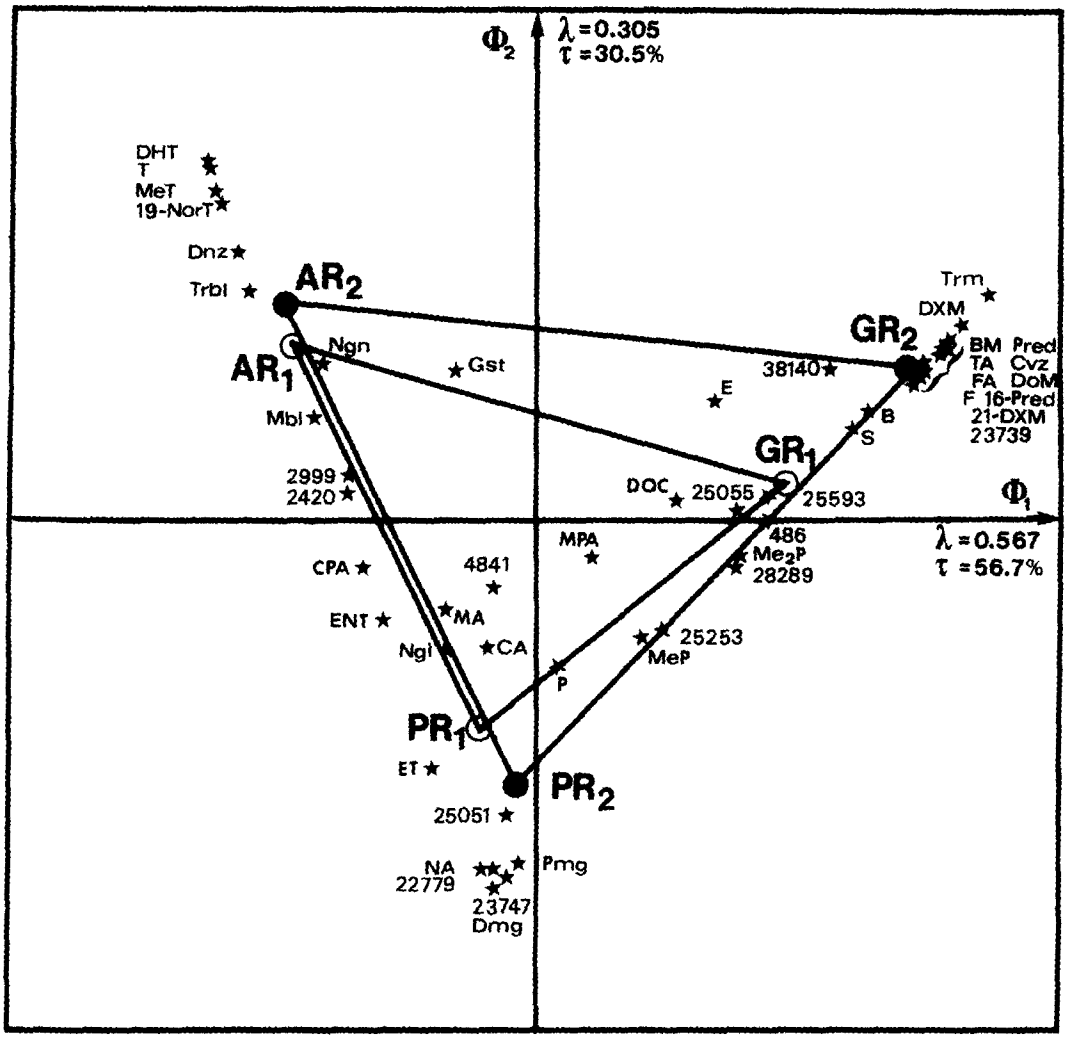

Fig. 4. $\phi_1\phi_2$ map derived by CFA of the data in Table 1 showing the relative locations of the test-steroids ★ with respect to the receptors (AR from rat prostate, PR from rabbit uterus, GR from rat thymus) at short (○) and long (●) incubation times. The abbreviations used are as in Table 1 and as follows: MeT = 17α-methyltestosterone, Dnz = danazol, Trbl = trenbolone, Ngn = norgestrienone, Mbl = metribolone, Gst = gestrinone, Ngl = norgestrel, Pmg = promegestone, Dmg = demegestone, MeP = 16α-methylprogesterone, Me_2P = 6α,16α-dimethylprogesterone, 21-DXM = 21-deoxydexamethasone, 16-Pred = 16α-methylprednisolone, FA = fluocinolone acetonide, DoM = desoximethasone, Cvz = cortivazol, TA = triamcinolone acetonide, Pred = prednisolone, BM = betamethasone, Trm = triamcinolone. (Reproduced from Ref. [20].)

Table 2. ACs and RCs of three receptors (AR, PR and GR), as defined under two sets of experimental conditions (1 and 2), to the factorial axes obtained by CFA of the data in Table 1

% Variance (τ) =	ϕ_1 Axis 56.7%	ϕ_2 Axis 30.5%	ϕ_3 Axis 7.9%	ϕ_4 Axis 3.1%	ϕ_5 Axis 1.9%
ACs					
GR_2	[+]39.2	[+]10.5	[+]35.0	0.0	0.0
AR_2	[−]20.6	[+]17.9	0.0	[−]16.0	[−]31.5
AR_1	[−]19.2	[+]13.1	0.0	[+]10.5	[+]42.5
GR_1	[+]18.5	[+] 0.7	[−]62.1	0.0	0.0
PR_1	[−] 2.1	[−]27.0	[+] 1.6	[+]33.5	[−]13.9
PR_2	[−] 0.4	[−]30.7	[+] 1.3	[−]40.0	[+]12.0
RCs					
GR_2	0.79	0.11	0.10	0.0	0.0
AR_2	0.64	0.30	0.0	0.03	0.03
AR_1	0.68	0.25	0.0	0.02	0.05
GR_1	0.67	0.01	0.31	0.0	0.0
PR_1	0.11	0.76	0.12	0.10	0.02
PR_2	0.02	0.84	0.01	0.11	0.02

The $\phi_1\phi_2$ factorial map thus highlights the most fundamental differences among the molecules which result from major differences in binding specificity rather than from lesser differences in interaction kinetics (incubation times). As indicated in Table 2, differences in interaction kinetics are revealed by the lower order axes (ϕ^3 to ϕ_5) that occlude specificity differences and oppose the ACs for long and short incubation times. The ϕ_3 axis reveals the opposition for GR ([−]62.1% vs [+]35.0%), the ϕ_4 axis for PR rather than AR and the ϕ_5 axis for AR rather than PR. It is clear that, for this population of molecules, the increase in incubation time has the greatest influence on the results obtained for GR.

The positions of the receptors in Fig. 4 are determined by the specificity of binding of the steroids which are clustered around or between them. Highly specific progestins [e.g. demegestone, promegestone, nomegestrol acetate, and the spirosultines (RU 22779, RU 23747, RU 25051)] are located around PR; specific androgens (e.g. 5α-dihydrotestosterone, testosterone, 17α-methyltestosterone, danazol and trenbolone) around AR, and glucocorticoids around GR. The steroids near the origin with the lowest variance are medroxyprogesteone acetate and RU4841 which interact appreciably with all three receptors. Androgenic 3-keto-trienes with increasing PR binding (e.g. metribolone, gestrinone, RU 2999, RU 2420 . . .) are in the top left-hand quadrant whereas progestins with decreasing AR binding (e.g. cyproterone acetate, megestrol acetate, chlormadinone acetate . . .) are in the lower left-hand quadrant.

The position of norgestrel, closer to the acetoxy derivatives of hydroxyprogesterone than to RU4841, is somewhat unexpected in view of its known competition for AR binding. Closer inspection reveals that its ACs (not shown) to the ϕ_1 and ϕ_2 axes are very low compared to its AC to the ϕ_4 axis which, as indicated in Table 2, discriminates between PR binding after long and short incubation times. The crude data in Table 1 confirm that the RBA of norgestrel for PR increases more than 5-fold with incubation time. Norgestrel thus forms an extremely stable complex with PR and this is its main characteristic compared to the overall population of molecules.

16α-Methylprogesterone and RU25253 are intermediate between PR and GR and have proportionally more important PR than GR binding components. Somewhat closer to the GR poles than to the PR pole are steroids with a greater GR than PR comp-onent (RU 28289, 6α,16α-dimethylprogesterone, RU25055, and RU25593) and also steroids with artificially high RBA values for GR at short incubation times (RU 38486, RU 28289, RU 38140) because the RBAs were measured at 4 and not 1 h. By grouping these steroids around the GR_1 pole instead of closer to PR as expected, the CFA method has highlighted the importance and value of standard conditions for meaningful screening tests. Well-known glucocorticoids (triamcinolone acetonide, dexamethasone etc . . .) are grouped around GR_2.

Analysis of receptors from different species and organs by hierarchical clustering and CFA

The above analyses were performed on data that were obtained in a screening system using cytosol from different species and different organs. The initial choice of material was governed by the species and organs used in traditional biological activity tests in an attempt to correlate RBAs with potency [1]. Data on species and organ differences are rather more scarce but we have selected two tables from the literature for analysis, one on progesterone-, the other on estrogen-binding proteins.

Progesterone-binding proteins. A data table remarkable for its minimal number of missing values and published in 1975 [21] concerns the RBAs of 63 steroids for progesterone-binding proteins in myometrium cytosol from four different mammalian species, namely, estrogen-primed postmenopausal women, sheep, rabbits and guinea-pigs. Competition for progesterone binding to myometrium and plasma from pregnant guinea-pigs was also measured.

The hierarchical classification derived from this data table [Fig. 5(a)] depicts the differences in binding specificity according to species. There is a marked contrast between the progesterone-binding proteins in the guinea-pig (rodent) compared to the three other species—rabbit (lagomorph), sheep (ruminant), and human (primate). There is also a contrast between the proteins in the pregnant and non-pregnant guinea-pig suggesting the appearance of a protein with different properties during pregnancy. The behavior of this myometrium pregnancy protein towards this population of steroids is close to that of pregnancy plasma suggesting that the tissue protein may be a plasma contaminant although the height of the node separating the two binding proteins is great enough to imply an intrinsic difference in binding properties. The progesterone-binding protein in the human myometrium binding protein is closest to that of the sheep and rabbit. The primary amino-acid sequences of both the human and rabbit progesterone receptors are known [22].

In the CFA analysis (Table 3), the ϕ_1, ϕ_2 and ϕ_3 factorial axes account for 96% of the information content of the data table. The ϕ_1 factorial axis [Fig. 5(b)], representing 83.8% of the information, confirms that for this population of molecules (a) the behavior of the guinea-pig binding proteins stands opposed to that of the binding proteins in the other species; (b) the binding specificity differs in the pregnant and non-pregnant guinea-pig; myometrium from pregnant guinea-pigs behaves like plasma; and (c) the sheep, rabbit and human binding proteins react in a very similar fashion. The steroids grouped around the proteins are at the source of the oppositions. The top group are rejected by the guinea-pig binding proteins

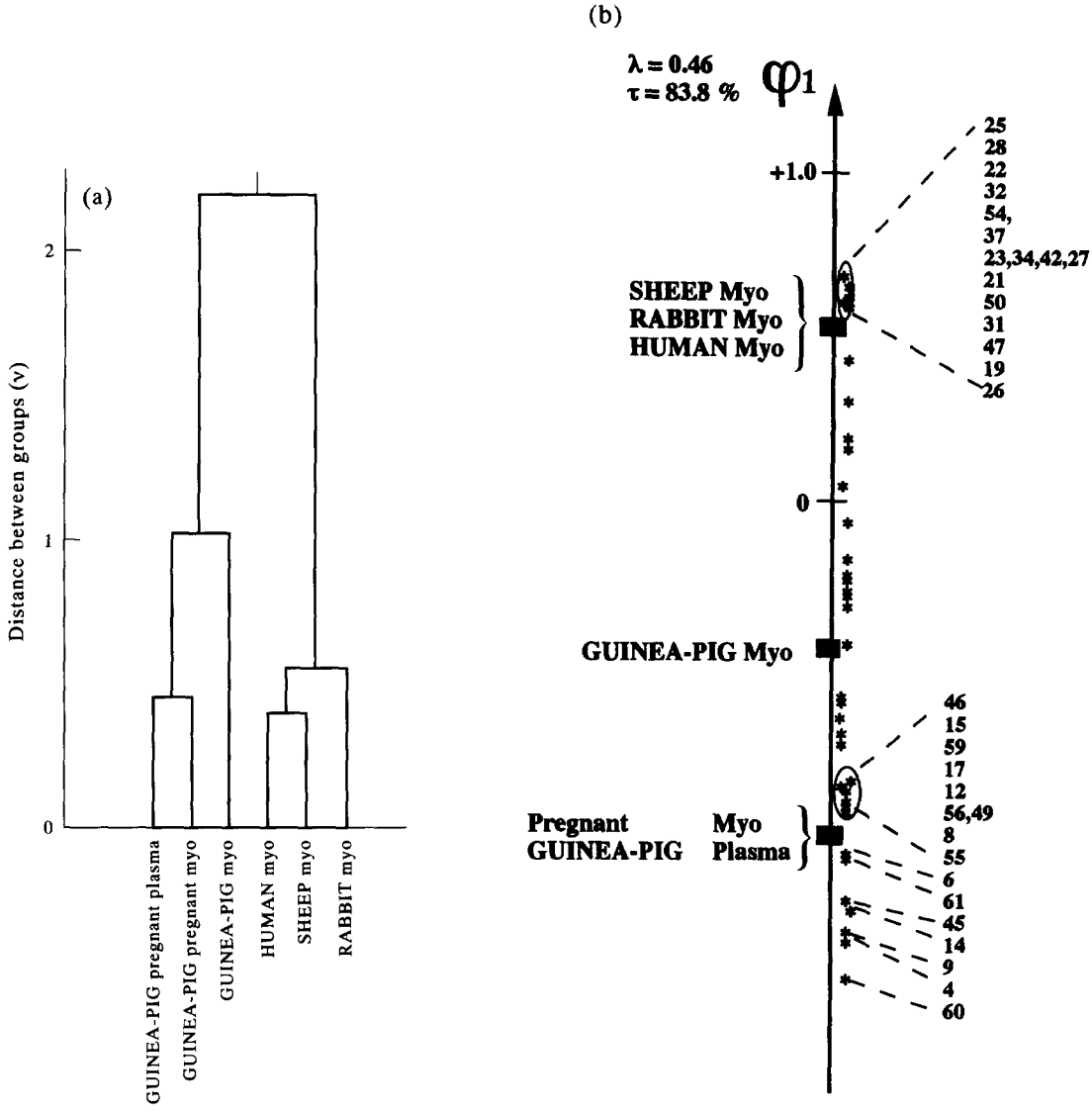

Fig. 5. (a) Hierarchical ascending classification of progesterone-binding proteins on the basis of specificity data for 62 steroids published by Kontula *et al.*[21]. (b) ϕ_1 factorial axis obtained by CFA analysis of the same data indicating the positions of the steroids (*) in relation to the receptors and to each other. The numbers (1–62) refer to the steroids in Table 1 of Ref. [21]. (17β-hydroxy-5α-androstan-3-one was excluded from the analysis because of two missing values. A single missing value was extrapolated for steroids Nos 6 and 60.) The most discriminatory steroids are as follows: *From the top*: 25: 4,17α-dihydroxy-4-pregnene-3,20-dione-17α-acetate; 28: chlormadinone acetate; 22: 17α-hydroxy-4-pregnene-3,20-dione acetate; 32: 16α-ethyl-21-hydroxy-19-nor-4-pregnene-3,20-dione phenylpropionate; 54: 17α-allyl-4-estren-17β-ol; 37: 17α-ethynyl-4-estren-3β,17β-diol; 23: 17α-hydroxy-4-pregnene-3,20-dione caproate; 34: 16α-ethyl-21-fluoro-19-nor-4,6-pregnadiene-3,20-dione; 42: lynestrenol; 27: megestrol acetate; 21: 16α,17α (1,1-dimethyl-methylene)-4-pregnene-3,20-dione; 50: d-norgestrel; 31: 16α-ethyl-21-hydroxy-19-nor-4-pregnene-3,20-dione; 47: 11β-chloro-17α-ethynyl-17β-hydroxy-4-estren-3-one; 19: 16α-ethyl-4-pregnene-3,20-dione; 26: medroxyprogesterone acetate. *From the bottom*: 60: 4-androstene-3,17-dione; 4: 20α-hydroxy-4-pregnen-3-one; 9: 19-hydroxy-4-pregnene-3,20-dione; 14: 4-pregnene-3,12,20-trione; 45: 17α-ethynyl-5-estren-17β-ol; 61: testosterone; 6: 3β-hydroxy-5-pregnen-20-one; 55: retroprogesterone; 8: 17α-hydroxy-4-pregnene-3,20-dione; 56: dydrogesterone; 49: 7α-methyl-17α-ethynyl-17β-hydroxy-5(10)-estren-3-one; 12: cortisol; 17: 4,16-pregnadiene-3,20-dione; 59: A-nor-3-pregnene-2,20-dione; 15: 16α,17α-epoxy-4-pregnene-3,20-dione; 46: 17α-ethynyl-6-estrene-5α,17β-diol.

in particular by the pregnancy proteins, the bottom group are rejected by the sheep, rabbit and human proteins.

The $\phi_2\phi_3$ display (Fig. 6), which describes 12.4% (6.8 + 5.6%) of the variance, analyzes lower-order relationships without any interference from the overpowering anti-correlation between the pregnant guinea-pig and other species. As shown by its central position and low ACs to these axes (Table 3), the binding protein of the pregnant guinea-pig myometrium contributes virtually no information to this map (ACs = [−]1.5 and [−]0.3% to the ϕ_2 and ϕ_3 axes, respectively). The map shows that the stereochemistry of rabbit 'receptor' binding (AC = [−]49.6% to the ϕ_2 axis) differs from that of human (AC = [+]19.0%), sheep (AC = [+]16.9%) and guinea-pig (AC = [+]12.5%) 'receptor' to very similar extents as reflected in the positions of these receptors

Table 3. ACs and RCs of the progesterone-binding proteins in the myometrium of different species and in guinea-pig plasma to the factorial axes obtained by CFA of the data in Ref. [21]

	ϕ_1 Axis	ϕ_2 Axis	ϕ_3 Axis	ϕ_4 Axis	ϕ_5 Axis
τ =	83.8%	6 8%	5.6%	2.5%	1 4%
ACs					
Pregnant guinea-pig	[−]31.4	[−] 1 5	[−] 0.3	[−]24.4	[−]29.5
Plasma	[−]31 0	[−] 0.3	[−]13 6	[+]27.2	[+]15.0
Rabbit	[+]15.4	[−]49.6	[+] 3.6	[+] 0.2	[+] 0.5
Human	[+]10.4	[+]19.0	[−]16.8	[−]21.4	[+]11.2
Sheep	[+] 7.2	[+]16.9	0.0	[+]26.7	[−]35.5
Guinea-pig	[−] 4.7	[+]12.5	[+]65.7	[−] 0.1	[+] 8.4
RCs					
Pregnant guinea-pig	0.96	0 00	0.00	0.02	0.01
Plasma	0.94	0.00	0.03	0.02	0.01
Rabbit	0.78	0.20	0.01	0.00	0.00
Human	0.75	0.11	0 08	0.05	0.01
Sheep	0.72	0.14	0.00	0.08	0.06
Guinea-pig	0.46	0.10	0.43	0.00	0.01

τ: Percent variance embodied by the factorial axis.

along the ϕ_2 axis and on either side of the ϕ_3 axis. Sheep myometrium 'receptor' does not contribute to the ϕ_3 axis which opposes the guinea-pig myometrium 'receptor' with the human 'receptor' and the plasma pregnancy protein. The steroids with the greatest information content (contributions not given) are numbered.

Estrogen-binding proteins. The data are extracted from work published by Bergink *et al.* [23–25] on the RBAs of C7 and C11-substituted estrogens for the cytosol ER of rabbit myometrium, pituitary, thymus and vagina, rat myometrium, endometrium and vagina, human myometrium, vagina, breast tumor (solid) and

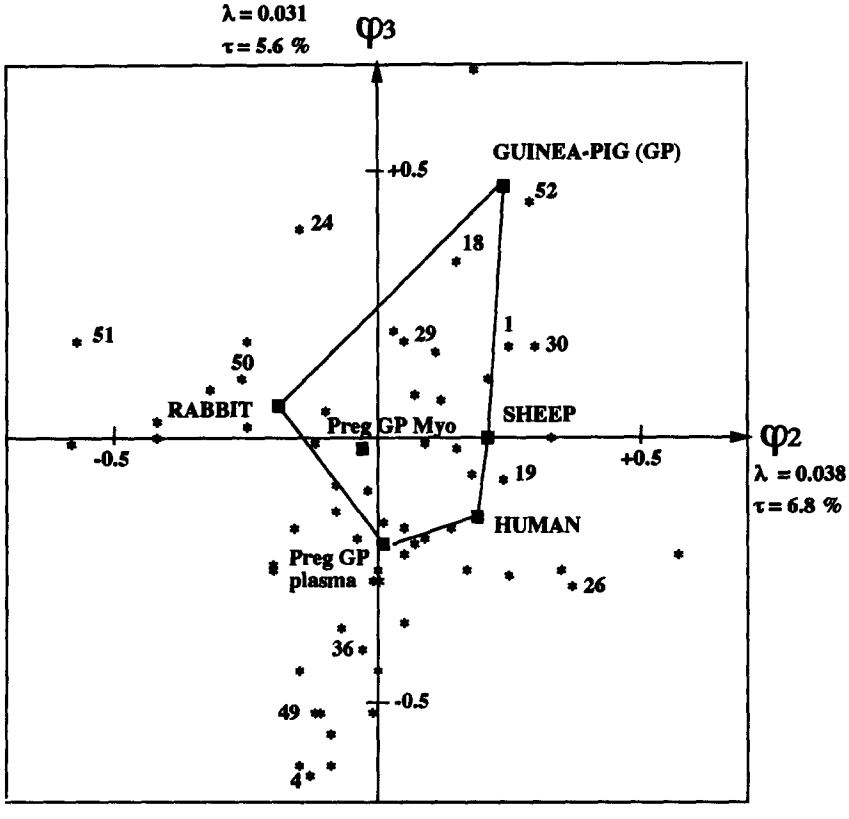

Fig. 6. $\phi_2\phi_3$ factorial map corresponding to the CFA analysis of Fig. 5b. Steroids (*) with RCs > 0.2 and ACs > 12% are indicated. (1: progesterone; 4: 20α-hydroxy-4-pregnen-3-one; 18: 7α-methyl-4-pregnene-3,20-dione; 19: 16α-ethyl-4-pregnene-3,20-dione; 24: 21-hydroxy-4-pregnene-3,20-dione acetate; 26: medroxyprogesterone acetate; 29; 19-nor-4-pregnene-20-dione; 30: 20β-hydroxy-19-nor-4-pregnen-3-one; 36: 17α-ethynyl-17β-hydroxy-5α-estran-3-one; 49: 7α-methyl-17α-ethynyl-17β-hydroxy-5(10)-estren-3-one; 50: *d*-norgestrel; 51: norethisterone acetate; 52: 19-nortestosterone.

Table 4. RBAs of estrogens in the cytosol of several tissues from different species

	Rat			Rabbit				Human			
	MYO	ENDO	VAG	MYO	PIT	THY	VAG	MYO	BK	MCF_7	VAG
1 Estradiol (E2)	100	100	100	100	100	100	100	100	100	100	100
2 Ethinylestradiol (EE)	126	87	116	100	101	56	68	103	98	102	58
3 11β-Methyl-EE	84	107	95	90	87	150	84	100	105	184	80
4. 7α-Methyl-E2	123	115	110	104	98	123	80	85	125	160	42
5. 7α,18-Dimethyl-EE	33	30	46	62	41	68	33	50	41	96	68
6. 11β-Methoxymethyl-EE	26	50	45	34	31	20	45	44	37	47	39
7 7α-Methoxymethyl-EE	26	39	22	37	37	38	36	41	35	28	190
8 3-Desoxy-EE	0 5	0.4	1 0	3.6	2.6	0.5	3 1	0.5	0.5	3	0.5
9. 17α-Estradiol	3	2	9	28	23	13	22	14	4	6	140
10. Estriol (E3)	7	4	17	17	13	6	28	18	7	7	100
11 Diethylstilbestrol (DES)	ND	36	34	29	14	19	51	32	42	82	4

ND: not determined (arbitrary values of 0 to 125 were assigned in the CFA when introducing DES as a supplementary variable). MYO = myometrium; ENDO = endometrium, VAG = vagina, PIT = pituitary, THY = thymus, BK = solid breast cancer tumor, MCF_7 = human malignant breast cancer cell-line Data from Ref [23]

MCF_7 cells (Table 4). Seven steroids in the original data table were excluded from the analysis because of missing values. To our knowledge, only the human (wild-type and MCF_7) [26–28] and rat [29] ER amino-acid sequences are known. Rabbit uterine ER has been characterized by partial peptide mapping [30].

The variance (information content) of the system is distributed as follows: ϕ_1 (84.2%), ϕ_2 (8.8%), ϕ_3 (2.6%), ϕ_4 (2.4%), ϕ_5 (1.0%). The ϕ_1 factorial axis represents the most discriminatory information (Table 5) which is the marked contrast between the specificity of this population of steroids toward the estrogen-binding protein in human vagina (AC = [+]81.6%) compared to ER from all other species and organs (ΣACs = [−]16.7%) except for rabbit vagina (AC = [+]0.2%).

The $\phi_1\phi_2$ factorial map (Fig. 7) represents 93% (84.2 + 8.8) of the variance. In other words, lower-order correlations and background noise accounting for 7% of the information content do not intervene in the representation. The most striking feature of Fig. 7 is the eccentric position of the binding protein in human vagina situated way out along the ϕ_1 axis and illustrating the atypical behavior of this protein toward the test-molecules. This specificity difference noted in the human vagina could reflect an intrinsic difference in the nature of the binding protein, which may not be a classic ER [31], a difference in the relative

Table 5. ACs and RCs of the estrogen-binding proteins in the organs of different species to the factorial axes obtained by CFA of the data in Table 4

	τ =	ϕ_1 Axis 84 2%	ϕ_2 Axis 8.8%	ϕ_3 Axis 2.6%	ϕ_4 Axis 2.4%	ϕ_5 Axis 1 0%
ACs						
Human vagina		[+]81 6	[+] 0.3	[+] 0.1	[−] 2 0	[+] 1 8
Human MCF_7		[−] 5 2	[+]29.6	[−] 3 2	[+] 6.5	[+]14.3
Rat myometrium		[−] 3.8	[−]13.6	[−] 4.8	[−]27 0	0 0
Human BK		[−] 2 8	[−] 0.6	[+] 4 4	[−] 7.2	0.0
Rat endometrium		[−] 2.7	[−] 1 6	[+]44.3	[−] 0 8	[+] 0.3
Rat vagina		[−] 1.9	[−] 7.1	[−] 1.8	[+] 5 0	[+] 7 8
Rabbit thymus		[−] 1 5	[+]39.3	[+] 1 2	[−] 8.4	[−]12 1
Human myometrium		[−] 0 2	[+] 2.0	[+] 0 4	[+] 3.2	[+]20 7
Rabbit vagina		[+] 0.2	[−] 2.7	[+] 6.6	[+]36.9	[−]26.0
Rabbit pituitary		[−] 0 1	[−] 3 2	[−] 6.2	[−] 0 9	[−]11.7
Rabbit myometrium		0 0	[−] 0 1	[−]27 1	[−] 2 2	[+] 5 2
RCs						
Human vagina		0.99	0.00	0.00	0 00	0.00
Human MCF_7		0.59	0.35	0 01	0 02	0.01
Rat myometrium		0 61	0 23	0.02	0.12	0 00
Human BK		0 86	0 02	0.04	0 06	0 00
Rat endometrium		0 62	0.04	0.31	0 01	0 00
Rat vagina		0 62	0 24	0.02	0.05	0 03
Rabbit thymus		0.24	0 67	0 01	0.04	0.02
Human myometrium		0.21	0.20	0.04	0.06	0.00
Rabbit vagina		0.09	0.14	0.10	0 50	0 15
Rabbit pituitary		0 14	0 36	0 20	0.03	0 15
Rabbit myometrium		0 02	0.01	0.75	0 05	0 06

τ Percent variance embodied by the factorial axis
BK: Solid breast cancer tumor

proportions of ER isoforms, a difference in ER hormone-binding sites, or a difference in the cellular (cytosol) environment. It is however clear that, whatever the species, the vagina is always displaced towards the right with respect to the other tissues. Although similar steroid specificities have been previously reported for rabbit uterus and vagina [32], the drug centrochroman was found to compete more effectively for rabbit vagina than uterus cytosol ER [33]. Differences in the proportion of ER isoforms have been recorded in the vagina and endometrium of the Australian marsupial *Trichosurus vulpecula* [34].

The ERs (except for the human vagina binding protein) are clustered according to species with hardly any overlap (see shaded areas) even if one takes into consideration the ER from malignant human sources.

In a published study [35], affinity ranking for ring B unsaturated estrogens was found to be similar for ER of human endometrium and rat uterus. However, in Fig. 7, human vagina ER is somewhat closer to ER from rabbit than rat vagina just as human myometrium ER is closer to ER from rabbit than rat myometrium thus indicating that the lagomorph might be a better model for the human than the rodent. In conformity with conclusions derived from proteolytic digest patterns of affinity-labeled ER, which indicate extensive structural relatedness between ER from MCF_7 cells and rat uterus and also GH_4 rat pituitary tumor cells [36], one observes an affiliation between ER from rat endometrium and from a human breast cancer but the proximity between these two variables is less than apparent since the high absolute contribution of rat endometrium ER to the ϕ_3 axis (Table 5) indicates that this variable is actually situated above the plane of the $\phi_1\phi_2$ factorial plot of Fig. 7.

Most ERs are relatively close to the origin indicating that differences are small. Those furthest removed from the central pool are rabbit thymus ER and MCF_7 ER. The distance separating ER from MCF_7 cells and from a solid breast tumor, and also the distance between these two ERs and human myometrium ER, may be of biochemical significance and reflect possible receptor polymorphism (see Discussion).

The positions of the steroids (*) explain the distribution pattern of the estrogen-binding proteins. 17β-estradiol (1) is in a relatively central position together with 3-desoxy-ethynyl estradiol (8) and 11β-methoxymethyl-ethynyl estradiol (6) implying an absence of preferential binding for any particular tissue. The steroids in the vicinity of human vagina ER, and far remote from uterine and malignant breast tissues, account for the eccentric position of this estrogen-binding protein. These are 17α-estradiol (9), estriol (10), and 7α-methoxymethyl-ethynyl estradiol (7). Estriol is known for its trophic action on the vagina and limited proliferative action on the endometrium [24] and it has been suggested by certain authors that estriol and estradiol may possess separate binding sites [37]. The steroids closest to human breast malignancy ERs are the C7 or C11-alkylated steroids, 7α-methylestradiol (4) and 11β-

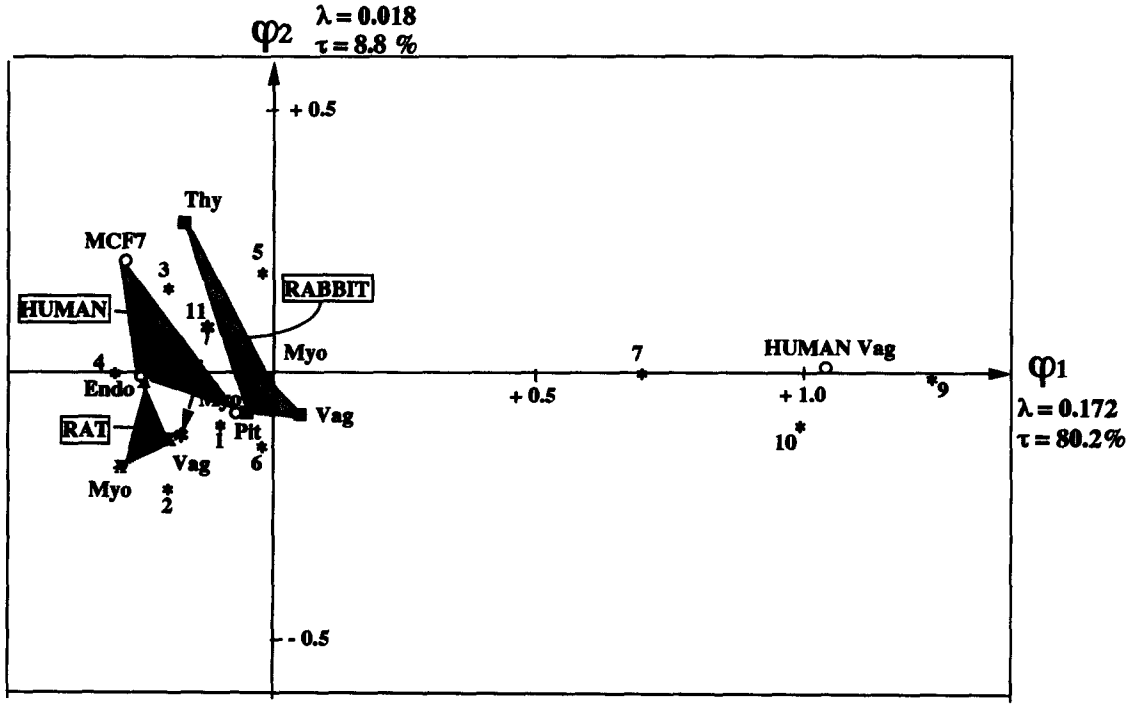

Fig. 7. $\phi_1\phi_2$ factorial map showing the relationships among steroids (*) and cytosol estrogen binding proteins from human (○), rat (x) and rabbit (■) tissues as derived from a CFA analysis of the data in Table 5. The theoretical range of positions of diethylstilbestrol (11), introduced into the CFA analysis as a supplementary variable after ascribing 8 arbitrary values between 0 and 125 to its RBA for ER in rat myometrium, are indicated by the vector (origin: RBA = 0; tip of vector: RBA = 125).

methyl-ethynylestradiol (3). 17α-Ethynyl estradiol (2) would have a propensity toward rat rather than rabbit or human tissues.

It is possible to introduce supplementary variables into a CFA analysis (either steroids or receptors). The RBA for diethylstilbestrol (DES) competition for binding to rat myometrium ER was missing, so we assigned 8 arbitrary RBA values ranging from 0 to 125. The positions taken up by DES within the factorial map, if these had been the true RBA values, are shown by the vector in Fig. 7 and, whatever the chosen RBA value, are within the zones governed by ER from rabbit and human myometrium and extending toward rat vagina at higher values.

The above analysis, which suggests that not only is there a species difference in estrogen-binding specificity but also a tissue difference due to either environment, presence of other binding proteins [38] and/or receptor polymorphism, is an illustration of the power of CFA in describing affiliations within the data matrix. Nevertheless, any generalization needs to be confirmed by a much larger study on a more diverse population of molecules.

DISCUSSION

Descriptive multivariate methods are not only extremely useful tools in studies of structure–activity relationships as already described in different fields [8, 10, 18–20, 39–43] but in order to assess correlations among biological parameters whether in biochemical experimentation [e.g. 44] or medical research [e.g. 45]. In the above examples, we have endeavored to illustrate how such methods can order steroid receptors on the basis of their response to populations of steroids in classic binding studies performed over the last two decades and have interpreted the results of the analyses in the light of present day literature.

The intention of these analyses is to be descriptive, i.e. to organize the information within the data tables in such a way as to discard redundant information and bring to light affiliations (correlations) among the variables, whatever their significance. No specific initial hypothesis is formulated other than that inherent in the choice of the variables (biological parameters or items), thus circumventing the bias of cause-to-effect studies. The aim of the approach is not to be original at all cost but informative and to incite questions based on an objective overall visual description of the system rather than on scientific intuition. The reasons for the correlations that are observed may need to be addressed in appropriate mechanism of action studies.

The hormone–receptor interaction has long been considered as a lock-and-key fit. Multivariate methods arrange the keys in an orderly fashion (chemical relationships among ligands as published elsewhere [18, 19, 43]), arrange the locks as illustrated above in the hierarchical classification of receptors, and also indicate which sets of keys open which types of lock (CFA analyses above and elsewhere on ligand–receptor interactions [20]). The present analyses have all been based entirely on the notion of specificity and not amplitude of response, i.e. whether a particular key can fit a lock but not the ease with which it does so. Amplitude of response can be taken into account in CFA analyses in which case the variables (e.g. receptors) are no longer represented by points in the factorial maps but by vectors [18] or by discrete functions [46].

Our analyses have shown that, for a population of 187 steroids tested within a pharmaceutical screening program, binding to GR and MR were affiliated but less closely than binding to AR and PR. Binding to ER was, as expected, most dissimilar. These affiliations describe the perception of the hormone binding site *by this population of steroids*. If this steroid population is representative of a larger population of all types of molecules that can bind to these receptors (which it is probably not since it was largely governed by the synthetic procedures in use by the pharmaceutical company), it could be considered a true picture of the affiliations among the hormone-binding sites.

Analyses of further tables highlighted the species specificity of binding. It was shown, on the basis of a population of 62 steroids, to what extent the specificity of the progesterone binding protein in human myometrium differed from that in the sheep, rabbit, and guinea-pig. The closest species to the human was the sheep, then the rabbit, emphasizing the choice of relevant experimental animal models for the study of steroids destined for human use. It is known that there are differences in the amino-acid sequences of PR according to species [22]. The analysis also highlighted the difference between the protein the authors detected in the myometrium of the pregnant and non-pregnant guinea-pig.

The analysis of estrogen-binding proteins in several tissues of several species was particularly revealing. Species-specific binding was recorded for a population of primarily C7- and C11-substituted estrogens but with a high degree of proximity verging upon overlap between the species, particularly in the case of certain tissues. There was one exception, namely, the estrogen-binding protein in the vagina which differed greatly from all other human tissues in its response to three steroids including estriol. The tissue differences could have one or more origins, for instance either differences in the cytosolic environment, the presence of interfering binding proteins, the accessibility to more than one hormone-binding site on the protein and/or receptor polymorphism. Since, in the factorial map with the highest variance, CFA exudes the most fundamental information and eliminates lesser order correlations and background noise, the correlations observed are liable to reflect an intrinsic property of the biological variables rather than extraneous influences. The presence of distinct heterogeneous forms of ER [e.g. 47] and also of multiple monomeric ER isoforms [e.g. 48]

has been described in different target tissues. The origin of this heterogeneity could lie in different ER mRNAs. Although genetic variants of the ER gene exist normally outside tumorigenesis [49], so far, with the exception of an ER variant in meningioma [50] and in the uterine tissue of women with frequent spontaneous abortions [51], most variant ER proteins have been observed in malignant breast tissue [49, 52, 53]. Polymorphism of ER in normal tissues may merit further study.

In contemporary articles on tissue specificity of steroid action, the emphasis is very much on the formation of a receptor–DNA complex which is modulated by tissue-specific factors that account for differences in biological response [e.g. 54]. The present analysis is a reminder that not only species differences but more minor tissue differences may exist at the level of the stereochemical recognition of the ligand by ER [55].

The data tables analyzed in the present paper are small and have been discussed in detail by the investigators [21, 23–25] including ourselves without the use of multivariate analysis. Our multivariate analysis confirms and extends published conclusions but without relying solely upon the experimentators' personal knowledge. Whereas personal analysis of small data tables is frequent, it becomes impossible when dealing with large data banks. In view of the surprising amount of specificity and activity data on steroid receptors gathered over the years, much information that may be relevant to present-day research, which is strongly influenced by the availability of new techniques and ligands, may lie in company and university archives. Multivariate analysis of these tables would tell us to what relative extents a property of a population of molecules is or is not accounted for by its other properties.

The present paper is therefore an incitement to all those dealing in ligand screening (a relevant field would be the screening of ligands to mutant receptors) to make maximum use of their data not only to design novel compounds but to further our understanding of relationships in complex biological systems. It also contains a word of caution. As the screening systems become more and more 'artificial' (e.g. use of diverse constructs), it becomes increasingly necessary to be able to correlate their information content with the *in vivo* situation. A new generation of drugs could be increasingly tissue-specific steroids such as estrogens active on the vagina, bone, cardiovascular and central nervous systems but inactive on the endometrium. If scientific experiments are not to be endlessly repeated, a simple statistical tool such as multivariate analysis may become indispensable to design the compounds with the appropriate mix of activities.

Acknowledgements—Complete data tables tend to be scarce in the literature at a time when the spotlight is on lead compounds and the authors would like to express their deep appreciation of the data published by Drs F. Zeelen, E W. Bergink and their coworkers (ORGANON)

REFERENCES

1. Ojasoo T. and Raynaud J. P.: Unique steroid congeners for receptor studies. *Cancer Res.* 38 (1978) 4186–4198.
2. Raynaud J. P., Ojasoo T., Bouton M. M. and Philibert D.: Receptor binding as a tool in the development of new bioactive steroids. In *Drug Design* (Edited by E. J. Ariëns). Academic Press, New York, Vol. 8 (1979) pp 169–214.
3 Chatfield C. and Collins A J.: Multidimensional scaling and cluster analysis. In *Introduction to Multivariate Analysis.* Chapman and Hall, London (1989) p. 189.
4. Benzécri J. P.: *L'Analyse des Données: II. L'Analyse des Correspondances.* Dunod, Paris, Vol. 2 (1983).
5. Greenacre M. J : *Theory and Application of Correspondence Analysis.* Academic Press, New York (1983).
6 Greenacre M J.· Correspondence analysis in medical research. *Stat. Meth Med. Res.* 1 (1992) 97–117
7. Jambu M.: *Exploratory and Multivariate Data Analysis.* Academic Press, New York (1991).
8 Doré J. C., Gilbert J., Ojasoo T. and Raynaud J. P.: Correspondence analysis applied to steroid receptor binding. *J Med. Chem.* 29 (1986) 54–60.
9. Lance G. N. and Williams W. T.. Computer programs for hierarchical polythetic classification ("similarity analyses"). *Comput. J.* 9 (1966) 60–64.
10. Ojasoo T., Doré J. C., Gilbert J. and Raynaud J. P.: Binding of steroids to the progestin and glucocorticoid receptors analyzed by correspondence analysis. *J. Med. Chem.* 31 (1988) 1160–1169.
11. Raynaud J P., Ojasoo T. and Vaché V.: Unusual steroids in measuring steroid receptors. In *Steroid Receptors and the Management of Cancer* (Edited by E. B. Thompson and M. E. Lippman) CRC Press, Boca Raton, FL (1979) pp. 215–232.
12. Delettré J., Mornon J P., Lepicard G., Ojasoo T. and Raynaud J. P.: Steroid flexibility and receptor specificity. *J. Steroid Biochem.* 13 (1980) 45–59
13. Delettré J.: Interaction "Hormone-stéroïde-protéine cytoplasmique" au travers d'une approche structurale des stéroides estrogènes, progestogènes, androgènes. Le stéroïde à "l'état isolé", corrélations paramètres structuraux-réponse biologique Ph. D. Thesis. Paris VI (1978).
14. Amero S. A., Kretsinger R. H., Moncrief N. D., Yamamoto K. R. and Pearson W R.: The origin of nuclear receptor proteins· A single precursor distinct from other transcriptional factors. *Molec. Endocr.* 6 (1992) 3–7.
15. Laudet V., Hanni C , Coll J., Catzeflis F. and Stéhelin D : Evolution of the nuclear receptor gene superfamily. *EMBO Jl* 11 (1992) 1003–1013
16 Bouton M. M. and Raynaud J. P.: The relevance of interaction kinetics in the determination of specific binding to the estrogen receptor. *J Steroid Biochem.* 9 (1978) 9–15.
17. Fritsch M , Leary C. M., Furlow J D., Ahrens H., Schuh J., Mueller G C. and Gorski J.: A ligand-induced conformational change in the estrogen receptor is localized in the steroid binding domain. *Biochemistry* 31 (1992) 5303–5311
18. Bignon E., Pons M , Doré J. C , Gilbert J., Ojasoo T., Miquel J. F., Raynaud J. P. and Crastes de Paulet A.: Influence of di- and tri-phenylethylene estrogen/antiestrogen structure on the mechanisms of protein kinase C inhibition and activation as revealed by a multivariate analysis. *Biochem. Pharmac.* 42 (1991) 1373–1383.
19. Doré J. C., Gilbert J., Bignon E., Crastes de Paulet A., Ojasoo T , Pons M., Raynaud J. P. and Miquel J. F.: Multivariate analysis of the combinations of structural determinants of diphenylethylenes and triphenylacrylonitriles implicated in estrogen receptor binding, modulation of protein kinase C and MCF_7 cell proliferation. *J. Med. Chem.* 35 (1992) 573–583.
20. Ojasoo T , Doré J. C., Mornon J. P. and Raynaud J. P : Two approaches to structure-activity relationships in the field of sex-steroids and their analogs. In *Molecular Structure and Biological Activity of Steroids* (Edited by M. Bohl and W. L. Duax). CRC Press, Boca Raton, FL (1992) pp. 157–207.
21. Kontula K., Jänne O., Vihko R., de Jager E , de Visser J. and Zeelen F.: Progesterone-binding proteins· *In vitro* binding and biological activity of different steroidal ligands. *Acta Endocr.* 78 (1975) 574–592
22. Misrahi M , Atger M., d'Auriol L., Loosfelt H., Meriel C , Fridlansky F , Guiochon-Mantel A., Galibert F. and Milgrom E · Human and rabbit progesterone receptors. Comparison of amino acid sequences deduced from cloned cDNAs. *Steroid Hormone Action* (1988) 127–139

23 Bergink, E. W., Kloosterboer H. J., van der Velden W. H. M., van der Vies J and de Winter M. S.. Specificity of an estrogen binding protein in the human vagina compared with that of estrogen receptors in different tissues from different species. In *Steroids and Endometrial Cancer* (Edited by J. V. Jasonni, I. Nenci and C. Flamigni) Raven Press, New York (1983) pp. 77–84.

24. Bergink E W. Oestriol receptor interactions· their biological importance and therapeutic implications. *Acta Endocr.* 233 (Suppl.) (1980) 9–16.

25 Bergink, E. W., Kloosterboer H. J. and van der Vies J.· Oestrogen binding proteins in the female genital tract *J. Steroid Biochem.* 20 (1984) 1057–1060

26. Green S., Walter P., Kumar V., Krust A., Bornert J. M, Argos P. and Chambon P.: Human oestrogen receptor cDNA sequence, expression and homology to V-erb-A. *Nature* 320 (1986) 134–139.

27. Greene G. L., Gilna P., Waterfield M., Baker A., Hort Y. and Shine J.: Sequence and expression of human estrogen receptor complementary DNA. *Science* 231 (1986) 1150–1154

28 Tora L , Mullick A., Metzger D., Ponglikitmongkol M., Park I. and Chambon P · The cloned human oestrogen receptor contains a mutation which alters its hormone binding properties. *EMBO Jl* 8 (1989) 1981–1986.

29 Koike S., Sakai M. and Muramatsu M Molecular cloning and characterization of rat estrogen receptor cDNA *Nucleic Acids Res.* 15 (1987) 2499–2513.

30. Olsen M. R , van Oosbree T. R. and Mueller G C . Characterization of rabbit uterine estrogen receptor proteins by radioiodination and partial peptide mapping. *Endocrinology* 119 (1986) 2177–2185.

31. Wiegerinck M A. H. M., Poortman J , Agema A R. and Thijssen J. H. H.: Estrogen receptors in human vaginal tissue *Maturitas* 2 (1980) 59–67

32. Payne D. W. and Katzenellenbogen J. A.: Differential effects of estrogens in tissues: A comparison of estrogen receptor in rabbit uterus and vagina. *Endocrinology* 106 (1980) 1345–1352

33. Joshi U. M., Naik V K. and Susheela P S · Effect of an anti-estrogenic contraceptive drug, Centchroman, on the binding of estrogen to rabbit reproductive tract tissues. In *Proceedings, III International Congress on Hormonal Steroids*, Hamburg 1970, *Excerpta Med.*, Amsterdam ICS 219 (1971)

34. Young C. E. and McDonald I. R.. Estrogen receptors in the genital tract of the Australian marsupial *Trichosurus vulpecula. Gen. Comp. Endocr* 46 (1982) 417–427.

35 Bhavnani B R and Woolever C. A.. Interaction of ring B unsaturated estrogens with estrogen receptors of human endometrium and rat uterus *Steroids* 56 (1991) 201–209.

36 Elliston J. F. and Katzenellenbogen B. S.: Comparative analysis of estrogen receptors covalently labeled with an estrogen and an antiestrogen in several estrogen target cells as studied by limited proteolysis. *J Steroid Biochem.* 29 (1988) 559–569.

37 Tamaya T , Kawabata I , Iida K and Imai A Effects of estradiol-17β and estriol on their binding sites in the rabbit uterus *Comp. Biochem Physiol.* **95B** (1990) 415–418.

38. Chae K., Johnston S. H and Korach K. S. Multiple estrogen binding sites in the uterus: Stereochemistry of receptor and non-receptor binding of diethystilbestrol and its metabolites *J Steroid Biochem. Molec Biol* 38 (1991) 35–42

39 Doré J C and Miquel J. F. Approche de l'étude des relations structure/activité par analyse des correspondances. *C. R Acad. Sci. Paris* 293 (1981) 1061–1064

40. Raynaud J. P., Fortin M., Hunt P., Ojasoo T., Doré J. C., Surcouf E. and Mornon J. P.: Approaches to drug development using receptors. In *The Role of Receptors in Biology and Medicine* (Edited by A M. Gotto Jr and B. O'Malley) Raven Press, New York (1986) pp. 63–77.

41. Renou M., Lalanne-Cassou B., Michelot D., Gordon G. and Doré J. C.: Multivariate analysis of the correlation between Noctuidae subfamilies and the chemical structure of their sex pheromones or male attractants. *J. Chem. Ecol.* 14 (1988) 1187–1215

42. Doré J. C., Lacroix J., Lacroix R. and Viel C.: De la manipulation informatique des formules développées à l'étude des relations structure-activité. Application aux médicaments antiparasitaires, *J. Pharmac. Belg.* 45 (1990) 375–384

43 Limasset B , Le Doucen C., Doré J. C., Ojasoo T , Damon M and Crastes de Paulet A · Effects of flavonoids on the release of reactive oxygen species by stimulated human neutrophils—Multivariate analysis of structure-activity relationships (SAR) *Biochem. Pharmac.* 46 (1993) 1257–1271.

44 Ojasoo T , Bignon E , Crastes de Paulet A , Doré J C , Gilbert J., Miquel J. F., Pons M. and Raynaud J. P.: Relative involvement of the estrogen receptor and protein kinase C in the action of a population of triphenylethylenes on MCF_7 cell proliferation as revealed by correspondence factorial analysis (CFA). *J. Steroid Biochem. Molec. Biol.* 44 (1993) 239–250

45. Ojasoo T , Fiet J , Raynaud J P. and Doré J. C.. A multivariate approach to the description of patient populations—Example of the analysis of the hormone profiles of patients with advanced prostate cancer. *J Steroid Biochem Molec. Biol.* 46 (1993) 183–193

46 Doré J. C., Gilbert J., Crastes de Paulet A , Michel F. and Miquel F. Analyse au moyen d'un traitement factoriel des relations structure-activité de di- et triphénylléthylènes sur deux tests biochimiques. *C. R. Acad. Sci. Paris* 294 (1982) 731–734.

47 Hyder S. M. and Wittliff J L. Detection of two high molecular weight hydrophobic forms of the human estrogen receptor. *J Steroid Biochem.* 33 (1989) 965–970

48 Geffroy-Roisne S , Duval J and Thieulant M.-L.: Multiple forms of affinity-labeled estrogen receptors in rat distinct pituitary cells *Endocrinology* 131 (1992) 1503–1510.

49. Murphy L C · Estrogen receptor variants in human breast cancer. *Molec. Cell Endocr* 74 (1990) C83–C86

50. Koehorst S. G. A., Jacobs H M., Tilanus M G J , Bouwens A. G M., Thijssen J H H. and Blankenstein M. A. Aberrant oestrogen receptor species in human meningioma tissue *J. Steroid Biochem Molec Biol.* 43 (1992) 57–61.

51. Lehrer S., Sanchez M , Song H K , Dalton J , Levine E., Savoretti P., Thung S. N and Schachter B Oestrogen receptor B-region polymorphism and spontaneous abortion in women with breast cancer. *Lancet* 335 (1990) 622–624.

52. Garcia T., Lehrer S., Bloomer W. D. and Schachter B. A variant estrogen receptor messenger ribonucleic acid is associated with reduced levels of estrogen binding in human mammary tumors. *Molec Endocr* 2 (1988) 785–791

53 McGuire W L., Chamness G. C. and Fuqua S A. W. Abnormal estrogen receptor in clinical breast cancer *J Steroid Biochem. Molec. Biol* 43 (1992) 243–247.

54 Korach K S , Chae K., Gibson M. and Curtis S Estrogen receptor stereochemistry· ligand binding and hormonal responsiveness *Steroids* 56 (1991) 263–270.

55 Martin L The biology of anti-oestrogens. *Biochem. Soc Trans* 7 (1979) 551–556