Biochimica et Biophysica Acta 1818 (2012) 896–906

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbamem

Predicting the effect of steroids on membrane biophysical properties based on the molecular structure

Jorge J. Wenz *

Instituto de Investigaciones Bioquímicas de Bahía Blanca, Camino "La Carrindanga" Km 7, B8000FWB Bahía Blanca, Argentina

ARTICLE INFO

Article history:
Received 18 July 2011
Received in revised form 19 December 2011
Accepted 22 December 2011
Available online 31 December 2011

Keywords:
Sterol structure
Membrane property
Multivariate analysis

ABSTRACT

The relationship between sterol structure and the resulting effects on membrane physical properties is still unclear, owing to the conflicting results found in the current literature. This study presents a multivariate analysis describing the physical properties of 83 steroid membranes. This first structure–activity analysis supports the generally accepted physical effects of sterols in lipid bilayers. The sterol chemical substituents and the sterol/phospholipid membrane physical properties were encoded by defining binary variables for the presence/absence of those chemical substituents in the polycyclic ring system and physical parameters obtained from phospholipid mixtures containing those sterols. Utilizing Principal Coordinates Analysis, the steroid population was grouped into five well-defined clusters according to their chemical structures. An examination of the membrane activity of each sterol structural cluster revealed that a hydroxyl group at C3 and an 8–10 carbon isoalkyl side-chain at C17 are mainly present in membrane active sterols having rigidifying, molecular ordering/condensing effects and/or a raft promoting ability. In contrast, sterol chemical structures containing a keto group at C3, a C4–C5-double bond, and polar groups or a short alkyl side-chain at C17 (3 to 7 atoms) are mostly found in sterols having opposite effects. Using combined multivariate approaches, it was concluded that the most important structural determinants influencing the physical properties of sterol-containing mixtures were the presence of an 8–10 carbon C17 isoalkyl side-chain, followed by a hydroxyl group at C3 and a C5–C6 double bond. Finally, a simple Logistic Regression model predicting the dependence of membrane activity on sterol chemical structure is proposed.

1. Introduction

The physical state of biological membranes has received increasing attention as it has been linked to several biological functions, such as the sorting of membrane components, membrane signaling, viral budding, amyloid formation, biosynthetic and endocytotic trafficking, etc. [1–9]. Typically, the ordered gel phase bilayer has a tight phospholipid molecular packing in which the lipid molecules also have restricted lateral motion. In the liquid-crystalline phase, a more disordered structure exists with a faster lateral molecular motion. A third phase, the liquid-ordered phase exhibits a well-packed and ordered arrangement of lipids, together with a relatively fast lateral diffusion [10–13]. The initial picture of a homogenous phospholipid matrix in the fluid mosaic model of membranes [14] was g radually superceded by the raft hypothesis in the 90s, which proposed a laterally segregated distribution of lipid molecules [1,15,16]. Although the bilayer architecture depends primarily on the phospholipid physicochemical properties, on their differential interactions and consequently the miscibility of the components, additional molecules incorporated into the lipid bilayer, such as membrane active sterols, can modify the bilayer physical properties [17].

The current literature on the non-genomic effects of sterols is generally focused on their ability to influence membrane physical properties such as permeability, lateral diffusion, the ordering/packing of lipids, and formation/stabilization of lateral-segregated lipid domains [18–30]. Physical studies investigating the influence of sterols on lipid bilayers are typically compared to those containing cholesterol under the same experimental conditions. However, the results of such studies, frequently performed using different physical techniques, vary considerably and often disagree (see a comprehensive review in [30]). Most of those studies (see Table 1, Suppl. Data) evaluate a limited number of sterol molecules (less than ten), usually membrane-associated sterols with functional and/or structural similarities (phytosterols, steroidal hormones, the presence or absence of methyl groups in the ring and isoalkyl side-chain and the number and position of double bonds, particularly in sterols in the biosynthetic pathway). Thus, inferences and conclusions arising from the analysis of a small number of sterol/lipid mixtures cannot easily be extended to sterol/lipid systems differing in their chemical structure and composition. Moreover, the application of a range of physical techniques to those sterol/lipid mixtures increases the variability of the results obtained and promotes misinterpretations of the influence

Abbreviations: PCoA, Principal Coordinate Analysis; LR, Logistic Regression; PC, principal coordinate; OR, odds ratio; CI, confidence interval

* Tel.: +54 291 486 1201; fax: +54 291 486 1200.
E-mail address: jwenz@criba.edu.ar.

0005-2736/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2011.12.021

of sterol chemical structure on the membrane phase properties. To address this issue, a wider set of sterols must be analyzed concurrently, in order to achieve a comprehensive perspective and establish a broadly applicable structure-activity relationship.

In the present study, the influence of sterol molecular structure on the physical properties of lipid membranes is investigated by analyzing a library of structurally diverse sterol molecules whose influence on the physical properties of different lipid mixtures have been documented in the literature. The effects of the different sterol molecular structures are quantified by defining binary variables that encode both the presence/absence of each difference in chemical structure and the physical properties associated with membrane activity. After the construction of a data matrix of sterol molecular structure versus experimentally-derived physical variables, Principal Coordinate Analysis (PCoA) and Logistic Regression (RL) were applied to assess the influence of each change in sterol chemical structure on the physical properties of the phospholipid bilayers. This approach permitted the construction of models for the prediction of sterol membrane activity as a function of their molecular structure. Since no distinctions were done regarding factors other than the chemical structure of sterols and their reported membrane activity, this work is addressed to ascertain a general picture of the structure–activity relationship of sterols on membranes, without concerning of specific phospholipid matrix, methods and/or experimental conditions associated with the measurements.

2. Methods

2.1. Defining the variables for the quantification of sterol membrane activity and sterol molecular structure

Sterol membrane activity is commonly considered in qualitative and comparative terms relative to cholesterol-containing or sterol-free lipid mixtures. In the present study, a quantitative statistical analysis was applied in which changes in sterol chemical structure were correlated with corresponding changes in membrane physical properties. Sterol membrane activity and molecular structure were converted into quantitative data by coding the information into categorical values representing the presence/absence of a given characteristic chemical or physical property. A dependent variable, termed "activity", was defined to sum all the measured effects of sterols on membrane physical properties. Accordingly, membrane "activity" was 1 for sterols reported as having rigidifying, molecular ordering, condensing effect, and/or raft promoting/stabilizing ability on membranes relative to that of free-sterol membranes; on the contrary, "activity" was 0 for those molecules documented as having fluidifying, disordering, and/or raft disrupting/destabilizing effect on membranes. It is noteworthy that the effect of sterol chemical structure on the lipid bilayer physical properties was not calculated relative to a cholesterol-containing lipid mixture (as is commonly found in the literature), but relative to a control bilayer having no sterol/steroid. As always occur after a categorization process, it should be noted that some information was missing after the discretization of the original data into categories, as different magnitudes of activity (but to the same direction) were considered as equals.

Sterols having no effect on the membrane-associated physical properties, "activity" was assigned a value of 0.5 in the classification analysis (see Section 2.2). In the study of the relationship between sterol chemical structure and membrane activity (see Section 2.3.3.), the variables were 1 for half of neutral sterols and 0 for the other half.

Independent variables (Table 1, Suppl. Data) were defined to mirror the molecular structure of sterols (Fig. 1, Suppl. Data) by different combinations of 1s and 0s that inform on the presence/absence of each chemical substituent in the fused-polycyclic ring structure. A variable was created for each different chemical substituent present in the sterol to be incorporated in the analysis. The presence of a given chemical substituent (hydroxyl, methyl, keto, double bond, etc.) at a specific position in the ring system was assigned the value of 1, whereas 0 indicates the absence of the same chemical substituent. Accordingly, 68 independent variables were required to identify the 83 sterols/steroids examined in this study with specific combinations of 1s and 0s (Table 1, Suppl. Data). It is worth noting that sterols may differ in their chemical structure by having the same chemical substituents in different arrays, since the positions of one or more of those groups may change on the ring system between two different sterols. Thus, the number of variables is smaller than the number of molecules.

2.2. Grouping sterols according to their molecular structure by Principal Coordinates Analysis

The dimension reduction method known as Principal Coordinate Analysis is a type of Multidimensional Scaling that explores similarities between observations [31,32]. In contrast to the popular Principal Component Analysis (PCA), which requires continuous variables, PCoA is a multivariate method useful to explore and visualize similarities (and dissimilarities) of categorical data. PCoA permits the analysis of the interdependence between variables and, through the construction of a similarity matrix, a graphical representation of the distances between samples. PCoA assigns to each sample a location in a low-dimensional space (usually as a 2D or 3D graphic), where individual and/or inter-group differences can be visualized. The major aim is to explain as much variability as possible by employing a reduced number of dimensions.

In this study, PCoA was employed to group sterols according to their similarities in molecular structure which were then correlated with sterol membrane activity. PCoA was applied over the matrix of sterols and the independent variables (83×68), which was constructed by calculating the Euclidean distance (d) between each pair of sterol data points given by:

$$d = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2 + \ldots + (a_{68}-b_{68})^2} \quad (1)$$

where a and b are two given sterols that may possess (1) or lack (0) one or more of the 68 traits (variables). Accordingly, a total of 3403 distances [$3403 = ((83*83)-83)/2$] were computed to encompass all pairs of molecules. Prior to the analysis, the software [33] converts distances into similarities between sterols by means of the expression $S = 0.5*(d^2)$, it computes some statistical parameters and finally shows two- or three-dimension scatter plots with the samples (molecules) distributed according to their similarities.

2.3. Logistic Regression for studying the structure–activity relationship and model building

2.3.1. Theoretical background

Logistic regression is an explanatory and predictive tool which analyzes the relationship between a dependent binary variable (0 or 1) and the independent variables, which may be of any type, categorical or continuous [34,35]. It may be used to determine the importance or the weight of the independents over the dependent variable and to determine a dependent variable as a function of one or more independent variables.

The general logistic regression equation is:

$$p = \frac{1}{1 + \exp-(b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n)} \quad (2)$$

where p is the probability of an event occurring, b_0 is the constant of the model, and b_n is the regression coefficient of the n independent variables (X). The event to be predicted in the present work is that a sterol displays membrane promoting activity. Thus, the probability p will range within the 0–1 interval, and it is expected to tend to 1 for membrane promoting sterols, and to 0 for membrane disrupting sterols. The cut value was established at 0.5 and sterols with a p

value higher (or lower) than this value were then classified as promoters (or disrupters).

The weight of each independent variable can be explained in terms of the regression coefficient b and/or the odds ratio (OR), related to b and to the probability of the event occurring, as:

$$OR = \exp(b) = \frac{\left[\frac{P}{(1-p)}\right]_{X=1}}{\left[\frac{P}{(1-p)}\right]_{X=0}}. \quad (3)$$

In this study, Eq. (3) represents the increase in the odds of a sterol to have membrane promoting activity if the trait is present ($X=1$) compared with its absence ($X=0$), as long as independent variables remain equal. Variables that report on chemical substituents frequently present in membrane promoting sterols typically exhibit regression coefficients $b>0$, and $ORs>1$. In contrast, variables associated with substituents present in membrane neutral sterols exhibit regression coefficients $b\sim0$ and $OR\sim1$. Finally, variables associated with membrane disrupting sterols typically have negative regression coefficients b and ORs close to 0 (i.e., $b<0$, and $0<OR<1$).

The outcome of a LR includes several statistics parameters informing on the model fit. The Wald chi-square test proves the null hypothesis that the parameter (b coefficient or the constant) equals 0. This hypothesis is rejected if the associated p-value is smaller than a given critical value (usually 0.05). Hence, the conclusion is that the parameter is not 0. The 95% confidence interval for OR is calculated as $\exp(b \pm 1.96 * S.E.)$; because of the mode of calculation, the interval is asymmetrical and the mean value of b is closest to its inferior limit. If 1 is contained in the interval, it should be concluded that there are no significant differences ($p<0.05$) between sterols having and lacking the respective trait. The statistic -2 Log Likelihood reports on the goodness of fit. It is not particularly informative by itself, but it can be used to compare different models. The lower the value, the better the fit. The Nagelkerke R square is a pseudo R square also informing on the goodness of fit. It is analogous to the R^2 in standard multiple regression, but it is not representative of the amount of variance in the dependent variable accounted for by the independent variables. It ranges between 0 and 1 and it is commonly used as an indicator of model fit. The better the model the closer to 1 the value.

2.3.2. Selection and grouping of variables

Although selection of variables is among the goals of the LR, a preselection (if possible) facilitates handling and interpretation of the regression output. Given the low amount of studies relating some specific substituents with sterol membrane activity (see these infrequent substituents in Table 1, Suppl. Data) some variables were grouped in order to increase the proportion of substituent-containing sterols relative to substituent-free sterols. For instance, if there is only one sterol with a specific substituent, the associated variable will be 1 for such sterol and 0 for the rest of the 82 substituent-free sterols. Thus, the low variance of this variable will not provide substantial information, but only confusion when interpreting the outcome of the repression. An unchanging variable likely introduces noise, but not an enhanced precision and/or accuracy to the model. Even though an increasing number of variables may turn the model more flexible in its fitting, an increasing amount of noise (whatever the source) is also modeled. An optimum number of variables are therefore usually recommended to include in the model build. For example, several studies report on the relation between membrane activity and the presence/absence of the C4,5 and C5,6 double bonds. However, few studies report on double bonds at ring positions different from those (e.g. 1,2–3,4–6,7–7,8–8,9–9,10–9,11–8,14–14,15 and 16,17). Accordingly, the original variables reporting on such double bonds were grouped into a new variable, named "DblBndOther" (see Table 1), which has an acceptable variance as it collects the information from the originals. Furthermore, the minimum of ten samples per independent variable suggested for a reliable LR and the actual number of cases (201) versus variables (68), justify a rationale reduction of the number of dimensions [36]. The initial variables (Table 1, Supplementary Data) were condensed into nine new variables according to: i) the location of the chemical substituent in the molecular structure, ii) the chemical nature of the substituent group, and iii) the number of cases possessing the chemical substituent (Table 1). The new variables were arranged into three categories according to the region in the sterol molecule (head, core and tail) in which the chemical substituent occurred, the location of the sterol chemical substituent in the bilayer (shallow, intermediate or deep). Within these three categories, variables were grouped according to the chemical nature of the substituent, but also according to the number of sterols with a given substituent, due to the low frequency of some chemical structures (i.e. epoxy, doxyl, 3 to 7 atoms acyl chain, etc.). Information about sterols having no group at C3 or having groups different from hydroxyl or keto (e.g. SO_3, SO_4, COOH, O–CH_3, hemisuccinate, acetate, O–C_2H_5 or doxyl) are provided by variables C3OH and C3CO, since they are zero in such circumstances. Hence, a third and redundant variable is avoided. The variable PlrGrpRings indicates the presence/absence of the followings groups and positions: 5α-OH, 6β-OH, 6-C=O, 5α,6α-epoxy, 5β,6β-epoxy, 7α-OH, 7β-OH, 7-C=O, 11-OH, 11-C=0 and 19-OH. The variable DblBndOther indicates the presence/absence of double bonds at positions: (1,2), (3,4), (6,7), (7,8), (8,9), (9,10), (9,11), (8,14), (14,15) and (16,17). The variable C17Chain3_7or_PlrGrp informs on the presence/absence at C17 of alkyl chains of 3 to 7 atoms and any of the following polar groups: OH, =O, C=O(CH_3), C=O(CH_2)OH, C=O(C_2H_5). Variable PlrGrpTail informs on the presence/absence of the following groups at different positions in the side alkyl chain: 20α-OH, 22R-OH, 22S-OH, 24α -OH, 25-OH, 27-OH, 22-NBD, 25-NBD and 25-doxyl.

Table 1
Selection and grouping of the variables for the Logistic Regression.

Region	Variable name	Description
Head	C3OH	Hydroxyl group at C3 (α-, β- or planar)
	C3CO	Keto group at C3
Core	PlrGrpRings	Polar group/s (OH; =O, epoxy) at rings, from C5 to C19
	DblBnd4_5	Double bond between C4 and C5
	DblBnd5_6	Double bond between C5 and C6
	DblBndOther	Double bond/s at ring positions different from C4–C5 and C5–C6.
Tail	C17Chain8_10	Long alkyl chains (8 to 10 atoms) at C17
	C17Chain3_7or_PlrGrp	Short alkyl chains (3 to 7 atoms) or polar groups at C17
	PlrGrpTail	Polar groups at tail

The first column denotes the region of the sterol backbone where the structural trait is present (or absent). The third column summarizes the criterion employed for grouping structural traits into each of the nine variables. Original variables are shown in Table 1, Supplementary Data.

As before, if the sterol contains a substituent group the new variables have a value of 1 or a value of 0 if it does not. At this stage, it is noteworthy that unequivocal identification of sterols is no longer valid using nine variables, and related molecules may then have identical combinations of 1s and 0s in the data matrix. Some of the original variables defined for the classification analysis (PCoA) were not included in the LR analysis, given that their information is implicit in other variables and then these initially defined variables are not helpful in the regression process. For example, the presence of polar groups different from OH or =O at C3 (i.e. SO_3, SO_4, COOH, O–CH_3, hemisuccinate, acetate, O–C_2H_5 or doxyl) is reported by "C3OH" and "C3CO", given that both variables are 0 in such situation; on the other hand, the absence of such polar groups is registered when "C3OH" or "C3CO" is 1. Incorporation of redundant variables is without merit and furthermore may hamper the interpretation of the LR outcome.

2.3.3. Selection of cases

The published works dealing with the effects of steroids on membrane physical properties have been carried out by using one or more methods within a wide range of availables, as electron differential scanning calorimetry, nuclear magnetic resonance, electron paramagnetic resonance, fluorescence intensity, fluorescence anisotropy and polarization, fluorescence resonance energy transfer, fluorescence recovery after photobleaching, infra-red spectroscopy, Fourier transform infrared spectroscopy, X-ray diffraction, small-angle X-ray scattering, differential scanning densitometry, dilatometry and ultrasound velocimetry, optical and electron microscopy, freeze-fracture electron microscopy, detergent solubility, atomic molecular dynamics simulations, etc. The use of such a great diversity of techniques together with varying experimental conditions (type and concentration of phospholipids, steroid concentration, etc.) increases the inter-study variability and explains, at least partially, why certain sterols have been reported as promoters, neutral or disrupters in different studies. Such inter-study variance was contemplated and modeled in the present report, as all cases were included in the analysis regardless of the reported membrane activity. It is worth noticing here that the term "case" does not refer only for a sterol, but also regard on each time an activity (equal or different) of a given sterol was found in the literature. Thus, the activity of some of the 83 sterols (see Table 1, Suppl. Data) is documented more than once. The mentioned variance was subsequently revealed by the width of the confidence interval obtained for some statistical parameters. From the 19 cases reported as having no effect on membrane physical properties (i.e. neutral), 10 of them were randomly assumed as promoters of membrane activity ("activity" = 1) in the present study, whereas the remaining 9 cases as disrupters ("activity" = 0). Finally, a matrix of 201 cases, regarding activity of 83 sterols, and 10 dichotomous variables ("activity" and nine independents) was constructed and subsequently employed in the LR analysis [37]. It is noteworthy that the term "case" accounts here for any time the membrane activity (the same or distinct) of a sterol is reported in different studies.

3. Results

3.1. Grouping of sterols according to similarities in their chemical structure and correlation with their membrane activity

PCoA was performed to reduce the dimensionality of the data and to provide a graphical representation of similarities/dissimilarities between the sterol molecular structures. Since only independent variables were included in this process, no information concerning the sterol membrane activity was initially included. From the 68 independent variables describing the variation in chemical structure of the entire sterol data set, the first two orthogonal coordinates were able to explain 35% of such variance. The coordinates of each molecule in the new axes (PC1 and PC2) are shown in Fig. 1A, where distance (or closeness) between points is proportional to the structural differences (or similarities) between molecules. An obvious clustering pattern was found, with most of the sterols arranged into one of five clusters. Fig. 1B shows a three-coordinate scattering plot after the incorporation of a further coordinate (PC3), which increased the percentage of the variance explained to 43%. To help in the view this three-coordinate space, values of PC3 are highlighted as projections of the points onto the plane PC1/PC2, e.g. at PC3 = 0 (dotted lines). This third coordinate improved the discriminating power of the analysis by increasing the inter-cluster separation in the third dimension. Clusters 1 and 5 exhibit negative values in PC3 and are then located below the plane, whereas clusters 2 and 4, having positive values, are located above the plane. The most isolated cluster in the two-coordinate plot, cluster 3, is situated near the plane PC1/PC2 and it remains well separated from the rest. Finally, fifteen molecules are dispersed in the center of the graph, whose structures are shown in Fig. 1 Supplementary Data, panel "dispersed".

After plotting the entire data set reporting differences/similarities among sterol molecular structure, the membrane activity of the sterols within each cluster was examined to determine the correlation with specific differences in sterol molecular structure. Since different activities for the same sterol molecule have been documented in the literature, an average value was calculated and employed in the subsequent analysis. Accordingly, sterols were classified as disrupter, neutral or promoting of membrane activity if the mean activity (codified initially with 0, 0.5 or 1, respectively) was in the 00.33, 0.34–0.66 or 0.67–1 interval, respectively. Table 2 summarizes this categorization based on the percentage of cluster members having a particular chemical substituent. Sterols structures are shown in Fig. 1, Supplementary Data.

All sterols in cluster 1 possess a C3-beta hydroxyl group, a C5,6 double bond and an iso-octyl side-chain at C17 as distinctive traits, and are the most similar to cholesterol in their structure, which also belongs to this cluster. Sterols in cluster 2 mainly differ from those in cluster 1 by the lack of the C5,6 double bond. The extremely high percentages of membrane promoting sterols in clusters 1 and 2 (100 and 94%, respectively) and the prevalence of hydroxyl groups

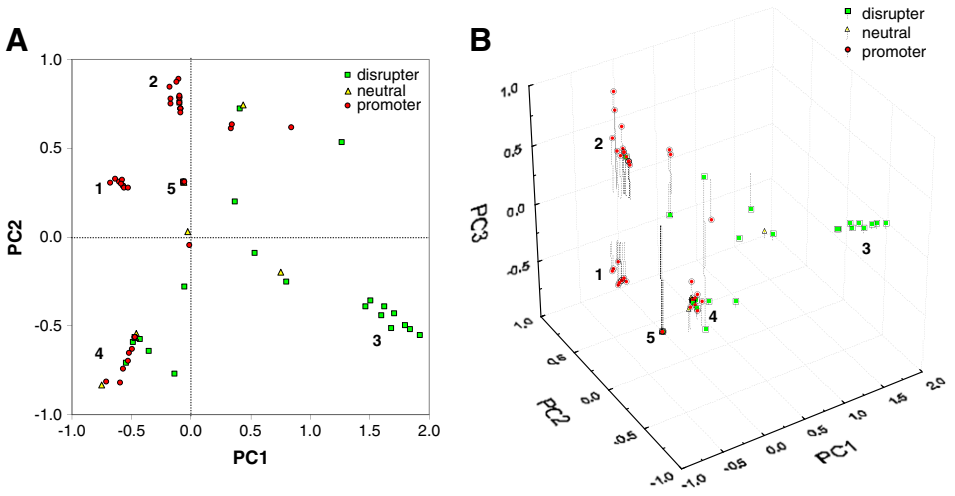

Fig. 1. The grouping of sterols according to their similarities in chemical structure, as deduced from the first two (A) and three (B) coordinates of the Principal Coordinate Analysis. The numbers inside graphs designate the five main clusters. The dotted lines in panel B are projections of PC3 values onto the plane PC1/PC2 (e.g. at PC3 = 0).

Table 2
Grouping of sterols according to structural similarities and activity.

	CLUSTER N°	1	2	3	4	5	Disp.
	N° of sterols/cluster	10	16	9	26	7	15
		% of sterols in clusters					
Activity	Disrupters (0–0.33)[a]	0	6	100	27	29	47
	Neutrals (0.34–0.66)	0	0	0	23	29	20
	Promoters (0.67–1)	100	94	0	50	43	33
Structural trait	C3-beta-hydroxyl	100	100	0	100	0	40
	C3-keto	0	0	100	0	0	20
	C3-other polar groups	0	0	0	0	100	33
	4,5-double bond	0	0	100	0	0	20
	5,6-double bond	100	0	0	100	100	20
	C17-alkyl chain (8 atoms)	100	100	0	38	100	73
	C17-polar groups	0	0	100	8	0	20
	C17-alkyl chain (3 to 7 atoms)	0	0	100	23	0	0
	C17-alkyl chain (9 to 10 atoms)	0	0	0	27	0	7

Except for the first row (sterols per cluster), numbers are percentages of cluster's members that possess the corresponding trait. Total number of sterols examined: 83.
[a] Numbers in parenthesis indicate the interval for the average "activity" employed for the classification of those sterols with different reported activities (see text for details).

at C3, C5,6 double bonds and eight carbon chains, strongly suggest that the presence of these common chemical substituents is associated with membrane promoting sterols.

Reinforcing this finding, all sterols in cluster 3 lack these chemical substituents and all 100% are known to display membrane disrupting activity. Instead, they all possess a keto at C3, a double bond C4,5 and substituents other than the iso-octyl side-chain at C17, suggesting that this change in the nature of the substituent, its size or position in the molecule confer disrupting activity to sterols. However, it is not possible to discriminate between the individual contributions of these traits to the overall disrupting activity, since all molecules in cluster 3 have these traits and lack any of the chemical substituents found in the membrane promoting sterols in clusters 1 and 2 (see Table 2).

All sterols in cluster 4 possess a C3–OH group and a C5,6 double bond, as was evident in clusters 1 and 2. However and in comparison with these clusters, only 38% of the sterols in cluster 4 have an iso-octyl side-chain and a lower fraction (50%) are membrane promoters, emphasizing that the C17 iso-octyl side-chain is an essential substructure necessary for a membrane promoting activity.

As was evident in cluster 1, all sterols in cluster 5 possess a C5,6 double bond and a C17 iso-octyl side-chain, but they have a polar group at C3 which is not an OH or keto group. Only 43% of the members in this cluster were reported to be membrane promoter, supporting previous finding that a hydroxyl group at C3 is also important for a promoting membrane activity. Finally, fifteen sterols displaying few structural similarities could neither be grouped into a cluster, nor associated with any of the preceding five clusters. A slight prevalence of disrupting activity (47%) is observed among these molecules.

In summary, on the basis of the analysis of the available literature measurements these findings suggest that the distinctive traits of cluster 1, namely C3β-OH, C5,6 double bond and a C17 iso-octyl side-chain are typical of membrane promoting sterols. On the other hand, the simultaneous presence of the distinctive traits seen in cluster 3, namely the C3-keto, a C4,5 double bond, the C17-polar groups and a shorter 3–7 carbon atom alkyl side-chain at C17 is usual in membrane disrupting sterols, as has been suggested from the interpretation of a range of experimental measurements.

3.2. Effects of the structural traits on sterol activity in membranes

3.2.1. Univariate logistic regression

A logistic regression was initially performed between the variable denoting "activity" and each of the nine independent variables in order to outline their effects alone, disregarding the interaction with the other variables. The importance or weigh of each substituent in promoting membrane activity in sterols is shown in Fig. 2 (gray bars). A negative *b* coefficient indicates that the presence of the substituent (i.e. a change in the independent variable from 0 to 1) decreases the probability (*p*), and thus it confers membrane disrupting activity to sterols. Conversely, a positive value of *b* increases *p*, signifying that the presence of the trait increases the promoting activity of the molecule. In other words, a value of *b* above and below 0 refers to traits found in promoting or disrupting sterols, respectively. The closer the coefficient *b* is to zero, the less the influence of the trait in the sterol membrane activity (i.e. neutral). When the 95% confidence interval (CI) for *b* ($b \pm 1.96 * \text{S.E.}$) includes the zero, it is assumed that there is no significant difference (at $p<0.05$) between the membrane activity of sterols with or without the chemical substituent.

The fact that zero is contained in the 95% CI for "PlrGrpRings" and "PlrGrpTail" indicates that each of those traits (see Table 1 for details of the implicated traits) has no significant effect in determining the sterol membrane activity. Based on the positive sign and magnitude of coefficients *b* of variables "C17Chain8_10" > "C3OH" > "DblBnd5_6" > "DblBndOther", it can be concluded that the corresponding traits increase *p* and thus are associated with membrane promoting activity. Conversely, coefficient *b* was negative for variables "C17Chain3_7or_PlrGrp", "C3CO" and "DblBnd4_5", indicating that the related traits decrease *p* and thus they are associated with disrupting activity on the sterols in phospholipid mixtures.

The statistical significance of coefficient *b* and the constant is estimated by the Wald statistics, is equal to zero (Table 3). Except for "PlrGrpRings" ($p = 0.328$) and "PlrGrpTail" ($p = 0.208$), all variables were highly significant ($p<0.05$), indicating that, independent of the presence of the other traits, they have an important disrupting or promoting effect on the sterol membrane activity. It is worth noting that the absence of statistic significance (e.g. $p>0.05$) accounts for both a poor correlation between the presence (or absence) of the traits and the reported membrane activity (variability in the published results), and/or a small number of physical studies of lipid mixtures containing sterols with the specific chemical substituent. For example, in the 201 cases, only 23 contain polar groups on the rings ("PlrGrpRings") and only 15 exhibit polar groups at the tail ("PlrGrpTail").

OR is a function of coefficient *b* (Eq. (3)) and represents the increment in the odds of a sterol to display membrane promoting activity

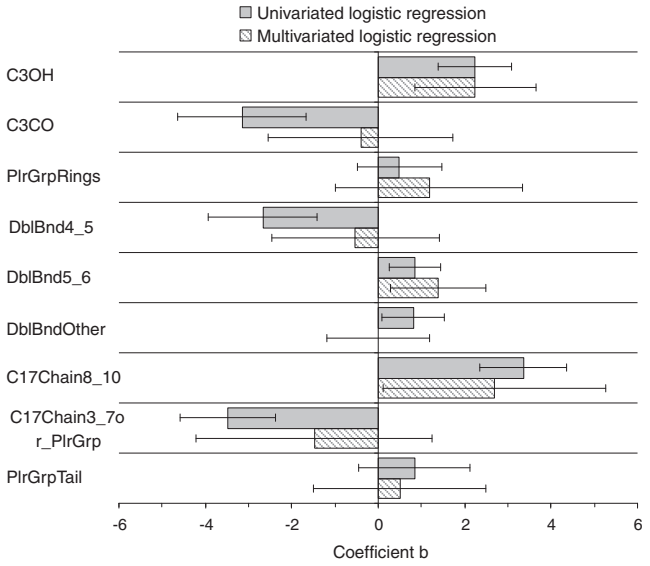

Fig. 2. The importance of substituent groups in determining sterol activity on membranes, based on the logistic regression coefficients. The horizontal bars indicate a 95% confidence interval ($b \pm 1.96 * \text{S.E.}$).

Table 3
Univariate logistic regression models.

Dependent variable: "activity"									Predicted			
Independent variables	Wald	Sig.(p)	exp(*b*) [OR]	95% CI Lower	95% CI Upper	−2 Log Likelihood	Nagelkerke R Sqr	Obs.	D	P	% Corr.	Overall
C3OH	25.9	0.000	9.4	3.9	22.1	229.6	0.20	D	27	44	38.0	74.1
Constant = −1.22	9.1	0.003						P	8	122	93.8	
C3CO	17.1	0.000	0.04	0.01	0.2	229.6	0.20	D	19	52	26.8	73.1
Constant = 0.90	30.0	0.000						P	2	128	98.5	
PlrGrpRings	1.0	0.328	1.6	0.6	4.3	260.1	0.01	D	0	71	0	64.7
Constant = 0.55	12.6	0.000						P	0	130	100	
DblBnd4_5	17.1	0.000	0.07	0.02	0.3	235.4	0.16	D	18	53	25.4	72.1
Constant = 0.87	28.6	0.000						P	3	127	97.7	
DblBnd5_6	7.8	0.005	2.4	1.3	4.3	253.2	0.05	D	0	71	0	64.7
Constant = 0.08	0.1	0.726						P	0	130	100	
DblBndOther	4.9	0.027	2.3	1.1	4.7	255.7	0.04	D	0	71	0	64.7
Constant = 0.41	6.0	0.014						P	0	130	100	
C17Chain8_10	42.7	0.000	28.8	10.5	78.9	192.8	0.40	D	38	33	53.5	81.1
Constant = −2.03	18.2	0.000						P	5	125	96.2	
C17Chain3_7or_PlrGrp	38.5	0.000	0.03	0.01	0.09	194.6	0.39	D	36	35	50.7	80.6
Constant = 1.28	44.9	0.000						P	4	126	96.9	
PlrGrpTail	1.6	0.208	2.3	0.6	8.5	259.3	0.01	D	0	71	0	64.7
Constant = 0.55	13.1	0.000						P	0	130	100	

The meaning of the statistical parameters is given in the text (Section 2.3.1). The last four columns in the table account for the goodness of fit regarding on how well the model predicts cases. "Obs.": observed number of disrupting (D) and promoting (P) cases in the data matrix. "Predicted": number of cases predicted as disrupting (D) or promoting (P) by the model. "% Corr.": percentage of cases correctly classified into each category (D or P). "Overall": overall percentage of cases for which the membrane activity was correctly predicted by the model.

($p \approx 1$) when the trait is present ($X = 1$) with respect to its absence ($X = 0$). An OR higher than unity is expected for traits conferring promoting membrane activity to sterols, whereas values between 0 and 1 for represent disrupting membrane activity. As can be seen in Table 3, sterols having traits associated with variables "C3OH", "DblBnd5_6", "DblBndOther" and "C17Chain8_10" are more likely to be promoters (9.36, 2.35, 2.26 and 28.8 more times, respectively), whereas sterols having traits associated with variables such as "C3CO", "DblBnd4_5" and "C17Chain3_7or_PlrGrp" are less likely to exhibit such activity (0.04, 0.07 and 0.03, respectively). Inclusion of the 1 in the 95% CI for OR indicates that there is no significant effect of variables "PlrGrpRings" and "PlrGrpTail", as deduced previously from the coefficients *b*. It should be pointed out that interpretation of OR must be done cautiously, since extremely high or low ORs may arise from small number of cases having the trait in relation to the number of independent variables, or from employing independent variables with low variance (e.g. with a low proportion of 1s or 0s).

Measurements of goodness of fit such as −2 Log Likelihood and the Nagelkerke R square (Table 3) are usually employed to compare statistical models. Good models have low values of −2 Log Likelihood and high values of the Nagelkerke R square (over the interval 0–1). Accordingly, models with the variables "C17Chain8_10" or "C17Chain3_7or_PlrGrp" offer the best predictions among the univariate models.

The last five columns in Table 3 indicate the model's ability to correctly classify cases. A case is correctly classified if the predicted activity coincides with the reported (i.e. promoter or disrupter). A model with no predicting capability at all should correctly classify approximately 50% of the cases because of simple rules of probability, whereas a satisfactory model should correctly classify more than 75% of cases [34]. As shown, models depending on the variables "C17Chain8_10" or "C17Chain3_7or_PlrGrp" display the best prediction capability, with 81.1 and 80.6% of cases classified into the correct class, respectively (see Table 3). It is worth noting that all of the models outlined here have poorer performance in classifying disrupting compared to promoting sterols; furthermore, four models (those associated with "PlrGrpRings", "DblBnd5_6", "DblBndOther" and "PlrGrpTail") misclassified all of the disrupting sterols (i.e. 0% were correct). This difference in the correct classification of the two possible categories depends on both the degree of correlation between each independent variable and "membrane activity", and also on the relative proportions of disrupting compared to promoting cases. By contrasting columns "D" and "P" in Table 3, it is apparent that there are 130 promoting cases compared to 71 of disrupting cases.

These findings derived from the logistic regression between "membrane activity" and each of the independent variables alone by means of the univariate logistic regression provide an initial overview of the relative importance of structural traits in determining the effect of a sterol on the physical properties of membranes. By means of Eq. (2) and the coefficients in Table 3, univariate logistic regression model as a function of the presence (or absence) of the most important traits (those involved in structural variables "C17Chain8_10" or "C17Chain3_7or_PlrGrp") can be constructed. Any of these models classifies more than 80% of the entries into the correct class, and it can be employed to calculate the probability (p) that a sterol display a given membrane activity. However, a more satisfactory model is required for the prediction of the disrupting activities of sterols.

3.2.2. Multivariate logistic regression

When performing multivariate regression, the inclusion of other variables may modify the statistical parameters of a variable in comparison to its values when examined isolated with the univariate regression, given that the effect of each variable is adjusted and controlled by the others. If two independent variables are correlated, the parameters of each one will differ in relation to those obtained from a simple regression. Each variable is expected to have fewer "weight" when both are present, as some fraction of the variability is concurrently explained by both. In the present context, correlated variables would be those reporting on substituents with a high concurrent occurrence. The magnitude of the changes in the parameters will depend on the degree of correlation between the independent variables, and on how much the presence of a trait modifies the effect of the others on the sterol behavior. This latter phenomenon, usually termed interaction between variables, refers to situations where the effect of a substituent is modified (increased or decreased) by the presence of other substituent. This adjustment is seen in Fig. 2 when

comparing the weight or importance of structural traits by means of the univariate (gray bars) and the multivariate (striped bars) regression. Although all the coefficients *b* retained their sign, suggesting that the overall effect remains the same, they differ in magnitude. The 95% CI for *b* in all variables was wider than in the univariate step, and the zero was now included in variables "C3CO", "PlgGrpRings", "DblBnd4_5", "DblBndOther", "C17Chain3_7or_PlrGrp" and "PlgGrpTail". This implies that these variables do not provide significant information in comparison with that provided by the remaining variables ("C3OH", "DblBnd5_6" and "C17Chain8_10") and they should not be included in the model. For example, the presence of keto group at C3 (i.e. "C3CO" = 1) is reported, up to certain extent, if a hydroxyl group is absent in such position (i.e. "C3OH" = 0). A similar situation, where the simultaneous presence of two traits is not possible in a specific position, arise among pairs "DblBnd4_5"/"DblBnd5_6r" and "C17Chain8_10"/"C17Chain3_7or_PlrGrp". `It is worth mentioning that the exclusion of these variables does not disagree with the preceding findings regarding their statistical significance when considered in isolation. Their exclusion does not necessarily imply the absence of an effect on sterol activity, but only that the model does not get extra information from them, nor improve its performance.

In view of the significant Wald statistic (Table 4), variables "C3OH", "DblBnd5_6" and "C17Chain8_10" were found as the most important ($p<0.05$) in determining sterol membrane activity. The lower limit of 95% CI for OR is greater than 1 in these variables, which indicates a promoting membrane activity in those sterols having such substituents. It should be recalled that an OR > 1 (or a coefficient $b>0$) is an indicator of a promoting membrane activity in the substituent-containing sterols, and not necessarily of significance of the variable. This is mainly reflected by the Wald statistic, where the higher the statistic (and hence the lower the associated p-value), the higher the significance of the variable. Thus, the fact that the most relevant variables have positive coefficients should be presumed as a coincidence, as the sign of the regression coefficient *b* is independent of the significance of the variable. The interaction among variables is mirrored by the changes in their relative weights computed with univariate and multivariate regressions (contrast Tables 3 and 4). This is particularly noticeable for "C17Chain8-10" as the OR decreases from 28.8 to 14.7, which suggests that a sterol having an 8 to 10 carbon chain at C17 is almost fifteen times more likely to be membrane promoter than a sterol with no such side-chain.

As expected, incorporation of all variables improves model performance, as is evident from the increase in the Nagelkerke R square (0.58) and the overall classification ability (84.1%), together with the decrease in the −2 Log Likelihood (151.1) in relation to those in any of the univariate models. This nine-variable model also has an enhanced performance in classifying disrupting cases (66.2% correct) in comparison with the preceding univariate methods.

3.3. Prediction of sterol activity as a function of their molecular structure

3.3.1. The parsimonious model

After an initial approach to ascertain the effect and the relative importance of the structural traits by the univariate and multivariate models, forward and backward variable selection methods were used in order to select the variables that suffice for the most simple, parsimonious model. These methods determine which variables to add or remove according to the significance of their inclusion or exclusion from the model, respectively. The analysis attempts to find the optimal number of variables in the model, looking for the best combination between simplicity and predicting capability. The above mentioned variables "C3OH", "DblBnd5_6" and "C17Chain8-10" were selected and found to be statistically significant ($p<0.001$) by means of these two procedures (Table 5; model A), which is in agreement with findings from the univariate regression (Section 3.2.1) and from the multivariate regression without variable selection (Section 3.2.2).

This three-variable model predicts that the likelihood that a sterol will be membrane promoting is augmented by around 13, 5 and 61 times if it possesses a trait involved in "C3OH", "DblBnd5_6" and "C17Chain8-10", respectively, than if not (i.e. if the independent variable changes from 0 to 1). Note that coefficients (and other parameters) differ from those of the full multivariate model, since this time they were adjusted to the presence of only the retained variables. No worthwhile differences were observed in the −2 Log Likelihood or in the Nagelkerke R square in relation to the complete model (154.6 and 0.57 in relation to 151.1 and 0.58, respectively), indicating that a reduction in the number of variables from 9 to 3 did not worsen the fit. Model A exhibits the same overall classifying ability as the previous full-variable model (84.1% correct), and an improved performance in classifying disrupter sterols (73.2% correct).

Finally, it is possible to calculate the probability of sterols to possess membrane promoting ($p\approx1$) or disrupting ($p\approx0$) activity based on their molecular structure, as a function of the presence ($X=1$) or absence ($X=0$) of the traits associated with variables "C3OH", "DblBnd5_6" and "C17Chain8_10". By replacing coefficients *b* of model A (Table 5) in Eq. (2) a three-variable model is obtained,

$$p=\frac{1}{1+\exp-[-5.67+(2.60*C3OH)+(1.51*DblBnd5\_6)+(4.12*C17Chain8\_10)]} \quad (4)$$

As an example, the model calculates a $p=0.93$ for the well-known membrane promoting sterol, cholesterol, which has a hydroxyl group at C3 ("C3OH" = 1), a C5,6 double bond ("DblBnd5_6" = 1) and a C17 iso-octyl side-chain ("C17Chain8_10" = 1). On the other hand, it calculates a $p=0.0034$ for 11α-hydroxyprogesterone, a membrane

Table 4
Multivariate logistic regression model including all independent variables.

Independent variables	Wald	Sig.(p)	exp(*b*) [OR]	95% CI Lower	95% CI Upper
C3OH	10.0	0.002	9.4	2.3	37.9
C3CO	0.1	0.708	0.7	0.08	5.6
PlrGrpRings	1.1	0.285	3.2	0.4	28.1
DblBnd4_5	0.3	0.595	0.6	0.09	4.1
DblBnd5_6	6.2	0.013	4.0	1.3	12.0
DblBndOther	0.0	0.997	1.0	0.3	3.3
C17Chain8_10	4.2	0.042	14.7	1.1	195
C17Chain3_7or_PlrGrp	1.1	0.287	0.2	0.01	3.5
PlrGrpTail	0.2	0.626	1.6	0.2	12.1
Constant	6.5	0.011			
−2 Log Likelihood	151.1	Nagelkerke R Sqr.			0.58

Predicted			
Obs	D	P	% Corr.
D	47	24	66.2
P	8	122	93.8
Overall			84.1

Table 5
Multivariate logistic regression models with selected variables.

Independent variables	*b*	95% CI Lower	95% CI Upper	Wald	Sig.(p)	exp(*b*) [OR]	95% CI Lower	95% CI Upper	Predicted			
MODEL A												
C3OH	2.60	1.5	3.7	22.2	0.000	13.4	4.6	39.5	Obs	D	P	% Corr.
DblBnd5_6	1.51	0.6	2.4	10.7	0.001	4.5	1.8	11.2	D	52	19	73.2
C17Chain8_10	4.12	2.9	5.3	45.0	0.000	61.5	18.5	205	P	13	117	90.0
Constant	−5.67	−7.5	−3.8	37.0	0.000				Overall			84.1
−2 Log Likelihood	154.6	Nagelkerke R Sqr.			0.57							
MODEL B												
C3CO (*)	−2.66	−4.3	−1.0	9.9	0.002	0.07	0.01	0.4	Obs	D	P	% Corr.
DblBnd5_6	1.18	0.4	2.0	7.8	0.005	3.3	1.4	7.4	D	45	26	63.4
C17Chain8_10	3.66	2.5	4.8	40.9	0.000	38.7	12.6	118.7	P	7	123	94.6
Constant	−2.76	−4.0	−1.5	19.1	0.000				Overall			83.6
−2 Log Likelihood	167.1	Nagelkerke R Sqr.			0.51							
MODEL C												
C3OH	2.19	1.2	3.2	17.6	0.000	8.9	3.2	24.8	Obs	D	P	% Corr.
DblBnd4_5 (*)	−1.60	−3.2	0.02	3.8	0.053	0.2	0.04	1.0	D	52	19	73.2
C17Chain8_10	3.52	2.5	4.6	41.3	0.000	33.9	11.6	99.4	P	13	117	90.0
Constant	−3.82	−5.2	−2.4	29.0	0.000				Overall			84.1
−2 Log likelihood	162.5	Nagelkerke R Sqr.			0.53							
MODEL D												
C3OH	2.48	1.4	3.5	21.1	0.000	11.9	4.1	34.4	Obs	D	P	% Corr.
DblBnd5_6	1.45	0.6	2.3	10.6	0.001	4.3	1.8	10.2	D	50	21	70.4
C17Chain3_7or_PlrGrp (*)	−4.15	−5.4	−2.9	41.5	0.000	0.02	0.00	0.06	P	12	118	90.8
Constant	−1.51	−2.6	−0.4	7.3	0.007				Overall			83.6
−2 Log Likelihood	158.5	Nagelkerke R Sqr.			0.55							
MODEL E												
C3OH	2.43	1.5	3.4	23.9	0.000	11.4	4.3	30.2	Obs	D	P	% Corr.
C17Chain8_10	3.49	2.4	4.6	41.8	0.000	32.9	11.4	94.8	D	52	19	73.2
Constant	−4.09	−5.4	−2.8	36.4	0.000				P	13	117	90.0
−2 Log likelihood	166.3	Nagelkerke R Sqr.			0.52				Overall			84.1
MODEL F												
C3OH	2.10	1.2	2.9	22.0	0.000	8.2	3.4	19.6	Obs	D	P	% Corr.
DblBnd5_6	0.52	−0.1	1.2	2.4	0.120	1.7	0.9	3.3	D	27	44	38.0
Constant	−1.43	−2.27	−0.6	11.0	0.001				P	8	122	93.8
−2 Log Likelihood	227.2	Nagelkerke R Sqr.			0.20				Overall			74.1
MODEL G												
DblBnd5_6	1.37	0.6	2.1	12.2	0.000	3.9	1.8	8.5	Obs	D	P	% Corr.
C17Chain8_10	3.66	2.6	4.7	44.8	0.000	38.8	13.3	113.3	D	38	33	53.5
Constant	−3.07	−4.2	−1.9	27.4	0.000				P	5	125	96.2
−2 Log Likelihood	179.8	Nagelkerke R Sqr.			0.46				Overall			81.1

Model A: independent variables retained after forward and backward variable selection procedures. Models B, C and D: one independent variable was replaced by one of the excluded in the variable selection procedures (marked with *), one at a time. Models E, F and G: two of the three most relevant variables.

disrupting sterol [28] having a C3-keto group instead of an OH group ("C3OH" = 0), a C4,5 double bond replacing that at C5,6 ("DblBnd5_6" = 0) and a C = O(CH3) group at C17 instead of the iso-octyl side-chain ("C17Chain8_10" = 0). According to this, possible values for *p* will range between these two extreme values of *p* (0.0034 and 0.93) and a sterol will be classified as membrane disrupter or promoter if the calculated *p* is lower or higher than the cut value of 0.5, respectively.

3.3.2. Alternative models

As mentioned above, a variable is excluded if it does not substantially improve model performance, although this not always implies the absence of effect of the trait on sterol behavior. In this sense, three of the variables excluded during the forward and backward variable selection procedure were next incorporated in place of each of the selected ones, one at a time (e.g. "C3CO" instead of "C3OH", "DblBnd4_5" instead of "DblBnd5_6" and "C17Chain3_7or_PlrGrp" instead of "C17Chain8_10") generating three new models (Table 5; models B, C and D). Note that every pair of interchanged variables reports on the same portion of the sterol molecule and the traits cannot be present concurrently. The interchanged variables (marked with asterisk in Table 5) have highly significant and negative coefficients, indicating a disrupting effect of traits associated with variables "C3CO", "DblBnd4_5" and "C17Chain3_7or_PlrGrp", in agreement with preceding findings from both the univariate and multivariate regression. Bearing in mind the −2 Log Likelihood, the Nagelkerke R square and the overall classification performance, the models do not fit data as well as model A, although the differences are small. For instance,

replacing "DblBnd5_6" in model A by "DblBnd4_5" yields model (C), which exhibits the same overall classification ability (84.1%) and a slightly inferior goodness of fit.

Alternative models (Table 5; models E, F and G) with two of the three variables of model A were also tested. The model as a function of "C3OH" and "C17Chain8_10" (model E) displayed the best performance (−2 Log Likelihood = 166.3; Nagelkerke R square = 0.52; % overall correct classification = 84.1%), not far from model A; on the other hand, the one depending on "C3OH" and "DblBnd5_6" (model F) exhibited the poorest performance. Combined, these findings point out that "C17Chain8_10" is the most important trait in determining sterol activity on membranes, followed by "C3OH" and "DblBnd5_6".

Considering the parameters reporting on goodness of fit, model A (Eq. (4)) possesses the best balance between simplicity and accuracy for the prediction of sterol activity on membranes. In view of their satisfactory performance, several of the two- or three-variable proposed models (Table 5) could also be employed.

4. Discussion

A systematic study analyzing the relationship between the sterol molecular structure and the effects of sterols on membrane physical properties have been presented. The present multivariate analysis has contemplated the discrepancies found in the literature by including all cases in the analysis in spite of their reported membrane activity, in order to compute the inter-study variability. Accordingly, findings and conclusions are averaged tendencies in the complex structure-activity relationship of sterols in membranes, and may not agree with some reported cases, as cited. The sterol membrane activity predicted in the present report will depend on the array of disrupting and/or promoting traits in the cyclo pentaneperhydrophenanthrene and on their relative weights. The prediction of the activity of sterols which have infrequent traits (i.e. very few molecules having the trait) may lead, however, to classifications in disagreement with some published cases, as was found in around 16% (Table 5) of the sterols examined. For instance, coprostanol possesses the two upmost promoting traits (the 8–10 carbon chain at C17 and the OH at C3) and differs from the archetypal promoting cholesterol only in its lack of the less relevant promoting trait, the C5–C6 double bond (Fig. 1, Suppl. Data). The probability p (Eq. (4)) for coprostanol is 0.74 ($p = 1/(1+\exp-(5{,}67+0+2{,}6+4{,}12)$), which is higher than the cut value of 0.5 and it is then classified as a promoter molecule, in contrast with some published results [38–40]. Even though assumptions emerging from the analysis of a single molecule should be avoided, a likely explanation could be the fact that coprostanol has the hydrogen at C5 in the beta configuration and thus the bulky A-ring is oriented toward the α-face of the molecule. Accordingly, the C3–OH is located away from the water–lipid interface and the "smoothness" of the molecular structure required for a proper fit in a phospholipid bilayer is absent [41]. As well, the α-OH would not be able to establish H-bonding with the carbonyl ester and phosphate ester groups of phospholipids or with the interfacial water molecules, as it does in the β configuration.

By means of the concurrent examination of the majority of the sterols with reported activity, the most relevant structural traits that govern the sterol effects on membranes have been ascertained. It should be noted, however, that other specific but infrequent molecular modifications and/or substituents in the sterol molecule may have an effect on its behavior when examined alone, demanding further studies. For example, the less common configurations of C3 and C5 have been considered in this analysis by including androstenol and epicholesterol (having a 3α-OH), and coprostanol and 5β,6β-epoxycholesterol (having a C5β configuration). Although these subtle variations were found to be not significant in comparison with other traits in determining sterol activity, they have been reported to affect the thermotropic phase behavior of DPPC mixtures to some extent [42,43]. It should be also noted that, besides the sterol molecular structure, there are additional factors that can modify the physical properties of bilayers, as sterol concentration in the mixture, solubility, sterol depth and tilt in the bilayer, phospholipid composition, etc. The variation that these factors may introduce in the evaluation of the sterol activity is clearly captured in the experimental measurement of the activity. Given that the reported activities were included regardless of such factors, the associated variation is implicitly modeled in this study, and further estimated and expressed by means of the parameters of goodness of fit.

For the current purposes, sterols reported in the literature as having rigidifying, molecular ordering, condensing effect, and/or raft promoting/stabilizing effects on membranes were named as "promoters" in this analysis, whereas sterols reported as having fluidifying, molecular disordering, and/or raft disrupting/destabilizing effects were named as "disrupters". Since these various membrane aspects are commonly associated to different lipid phases not dissected in this study, the present findings should be assumed as a general overview of the phenomenon. Since no discrimination has been made regarding the lipid matrix or phase, which actually contributes to the great variability found in the literature, conclusions are not addressed to a particular bilayer of lipid phase. In view of the multivariate analysis between the current "activity" and the molecular structure of sterols, the picture that emerges from this study is that a C17 8–10 carbon atom side-chain is the most important structural trait in determining the effect of sterols in phospholipids bilayers, followed by a C3–OH group and a C5,6 double bond. It is noteworthy that the fact that these traits confer membrane promoting activity to sterols is not the reason why they have been found to be the most relevant, given that the importance of a trait is dictated by its relative influence in the regression process, regardless of the type of activity that they confer to sterols.

As mentioned, exclusion of a trait (variable) from the model does not necessarily imply the absence of effect of the trait on sterol behavior. In this sense, structural traits as a keto group at C3, a double bond between C4,5, and short alkyl chains (3 to 7 atoms) or polar groups at C17 were excluded from the model when analyzed concurrently with the rest of the variables. However, these traits displayed significant and negative coefficients when examined in isolation by means of the univariate regression, indicating that they certainly confer disrupting activity to sterols. Unfortunately, the fact that these traits are present concurrently in most of the disrupter sterols (See Table 2) becomes a hindrance to discriminate among individual effects and to establish a rank of importance among such traits.

It is worth noticing that a molecular structure quite different from that of a typical promoter sterol does not imply an opposite membrane activity. For instance, lanosterol seems to be fairly different from cholesterol (Fig. 1, Suppl. Data): it has 2 methyl groups at C4, a methyl group at C14, a C8,9 double bond, and a C24,25 double bond. In fact, this structural difference was clearly detected in the PCoA (see Section 3.1), as it is one of the fifteen sterols excluded from any of the five structurally related clusters, and it remained among the dissimilar and ungrouped sterols. Nevertheless, lanosterol is not so different when considering the traits that govern sterol effects on the membrane properties: it differs from the typical promoting sterol only in the lack of the C4,5 double bond. Thus, by means of Eq. (4) (C3OH = 1; DblBnd5_6 = 0; C17Chain8_10 = 1) a $p = 0.74$ is obtained. Since $p > 0.5$, lanosterol was classified as a promoter sterol, although not as strong as cholesterol ($p = 0.93$).

It is interesting to compare the phase diagram of binary lipid mixtures containing sterols with opposite predicted membrane activity, to test the model performance. From the phase diagram of DPPC/androsterol mixture [44] it was found that this sterol reduces the main transition temperature of the bilayer more effectively than cholesterol [45], ergosterol and stigmasterol [46]. The three-phase line in the phase diagram was also 3 to 5° lower than that of the other binary

mixtures, indicating a less compact packing of the androsterol/DPPC self-assembled structure than is the case with the other sterols. It was also found that the two end-points of the three-phase line in androsterol/DPPC mixtures were higher (11.1 and 30.9 mol%, respectively) than that in the phase diagram of cholesterol, ergosterol, or stigmasterol (6–8 and 20–25 mol%), implying that a greater proportion of androsterol in DPPC is required to achieve an ordered phase. Finally, the authors conclude that androsterol is less effective in promoting the formation of an ordered phase and that this ordered phase is less compact than the normal liquid-ordered phase. The model proposed here (Eq. (4)) predicts a p value of 0.17 for androsterol and 0.93 for the other sterols, and so they are classified as disrupter and promoters, respectively. On the basis of the membrane physical properties that were associated with disrupters and promoters (see Section 2.1), this prediction agrees with the published data.

In addition to the contribution to the understanding of the structure–activity relationship of sterols, the recognition of the effects of the structural traits presented in this study is expected to be useful in those research fields where the activity of a number of sterols must be examined. By means of the proposed model/s and molecular structures, a first picture of the activities could be achieved and then the amount of confirmatory experiments can be minimized to that of the potential candidates. Furthermore, the present findings can also be helpful in experimental- and computer-assisted drug design studies aimed at obtaining a particular activity on a prototype, or to increase/decrease an already existing one, since the knowledge of the effect of the structural traits would serve as a guide for the structural modifications to be done.

The interaction between sterols and phospholipids in membranes is a very complex problem, seeing that the behavior of both lipid and sterol molecules is not only determined by steric impediment, but also by interactions as H-bonding, van der Waals forces, etc. Although the present work attempts to contribute to the understanding of the structural requirements of sterols to affect the thermotropic phase behavior of bilayers, much work remains to be done in order to provide a detailed insight into sterol/lipid interactions, and to address currently open questions regarding the structure–activity relationship, as well as some of the discrepant results found in the literature.

5. Conclusions

The present study has shown that a C17 8–10 carbon atom sidechain is the most important structural trait in determining the effect of sterols in phospholipids bilayers, followed by a C3–OH group and a C5,6 double bond. On the basis of the current definition of membrane activity, these traits confer promoting activity to sterols, implying that an increased rigidity, molecular ordering, packing, and/or raft formation/stabilization are expected in lipid mixtures containing sterols with such traits. On the other hand and having less importance in governing the activity on membranes, a keto group at C3, a double bond between C4,5, and short alkyl chains (3 to 7 atoms) or polar groups attached to C17 were found to confer disrupting activity to sterols, decreasing the aforementioned physical properties. Finally, by employing a logistic regression model as a function of the encoded chemical structure, the membrane activity of sterols can be predicted.

Supplementary materials related to this article can be found online at doi:10.1016/j.bbamem.2011.12.021.

Financial and competing interests disclosure

No financial support, grants or writing assistance was utilized in the production of this work.

References

[1] A. Rietveld, K. Simons, The differential miscibility of lipids as the basis for the formation of functional membrane rafts, BBA 1376 (1998) 467–479.
[2] D.A. Brown, E. London, Functions of lipids rafts in biological membranes, Annu. Rev. Cell Dev. Biol. 14 (1998) 111–136.
[3] K. Simons, E. Ikonen, How cells handle cholesterol, Science 290 (2000) 1721–1726.
[4] S. Heino, S. Lusa, P. Somerharju, C. Ehnholm, V.M. Olkkonen, E. Ikonen, Dissecting the role of the golgi complex and lipid rafts in biosynthetic transport of cholesterol to the cell surface, Proc. Natl. Acad. Sci. 97 (2000) 8375–8380.
[5] J. Herreros, T. Ng, G. Schiavo, Lipid rafts act as specialized domains for tetanus toxin binding and internalization into neurons, Mol. Biol. Cell. 12 (2001) 2947–2960.
[6] D.H. Nguyen, D. Taub, CXCR4 function requires membrane cholesterol: implications of HIV infection, J. Immunol. 168 (2002) 4121–4126.
[7] W. Popik, T.M. Alce, W.C. Au, Human immunodeficiency virus type 1 uses lipid raft-colocalized CD4 and chemokine receptors for productive entry into CD(+) T cells, J. Virol. 76 (2002) 4709–4722.
[8] F.R. Maxfield, I. Tabas, Role of cholesterol and lipid organization in disease, Nature 438 (2005) 612–621.
[9] T. Róg, M. Pasenkiewicz-Gierula, I. Vattulainen, M. Karttunen, Ordering effects of cholesterol and its analogues, Biochim. Biophys. Acta 1788 (2009) 97–121.
[10] D.J. Recktenwald, H.M. McConnell, Phase equilibria in binary mixtures of phosphatidylcholine and cholesterol, Biochemistry 20 (1981) 4505–4510.
[11] J.H. Ipsen, G. Karlstrom, O.G. Mouritsen, H. Wennerstron, M.J. Zuckermann, Phase equilibria in the phosphatidylcholine–cholesterol system, Biochem. Biophys. Acta. 905 (1987) 162–172.
[12] T.H. Huang, C.W. Lee, S.K. Das Gupta, A. Blume, R.G. Griffin, A 13C and 2H nuclear magnetic resonance study of phosphatidylcholine/cholesterol interactions: characterization of liquid-gel phases, Biochemistry 32 (1993) 13277–13287.
[13] M.B. Sankaram, T.E. Thompson, Interaction of cholesterol with various glycerophospholipids and sphingomyelin, Biochemistry 29 (1990) 10670–10675.
[14] S.J. Singer, G.L. Nicolson, The fluid mosaic model of the structure of cell membranes, Science 175 (1972) 720–731.
[15] D. Brown, J.K. Rose, Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface, Cell 68 (1992) 533–544.
[16] K. Simons, E. Ikonen, Functional rafts in cell membranes, Nature 387 (1997) 569–572.
[17] Y. Barenholz, Cholesterol and other membranes active sterols: from membrane evolution to "rafts", J. Lipid Res. 41 (2002) 1–5.
[18] R.A. Demel, K.R. Bruckdorfer, L.L.M. Van Deenen, Structural requirements of sterols for the interactions with lecithin at the air–water interface, Biochim. Biophys. Acta 255 (1972) 311–320.
[19] P.L. Yeagle, R.B. Martin, A.K. Lala, H.K. Lin, K. Bloch, Differential effects of cholesterol and lanosterol on artificial membranes, Proc. Natl. Acad. Sci. 74 (1977) 4924–4926.
[20] K.W. Butler, I.C. Smith, Sterol ordering effects and permeability regulation in phosphatidylcholine bilayers. A comparison of ESR spin-probe data from oriented multilamellae and dispersions, Can. J. Biochem. 56 (1978) 117–122.
[21] J. Rogers, A.G. Lee, D.C. Wilton, The organisation of cholesterol and ergosterol in lipid bilayers based on studies using non-perturbing fluorescent sterol probes, Biochim. Biophys. Acta 552 (1979) 23–37.
[22] K.E. Bloch, Sterol structure and membrane function, CRC Crit. Rev. Biochem. 14 (1983) 47–92.
[23] J.A. Urbina, S. Pekerar, H.B. Le, J. Patterson, B. Montez, E. Oldfield, Molecular order and dynamics of phosphatidylcholine bilayer membranes in the presence of cholesterol, ergosterol and lanosterol: a comparative study using 2H-, 13C- and 31P-NMR spectroscopy, Biochim. Biophys. Acta 1238 (1995) 163–176.
[24] X. Xu, E. London, The effect of sterol structure on membrane lipid domain reveals how cholesterol can induce lipid domain formation, Biochemistry 39 (2000) 843–849.
[25] A.B. Serfis, S. Brancato, S.J. Fliesler, Comparative behavior of sterols in phosphatidylcholine–sterol monolayer films, Biochem. Biophys. Acta. 1511 (2001) 341–348.
[26] G.V. Martinez, E.M. Dykstra, S. Lope-Piedrafita, M.F. Brown, Lanosterol and cholesterol-induced variations in bilayer elasticity probed by 2H NMR relaxation, Langmuir 20 (2004) 1043–1046.
[27] J. Wang, Megha, E. London, Relationship between sterol/steroid structure and participation in ordered lipid domain (lipid rafts): implications for lipid rafts structure and function, Biochemistry 43 (2004) 1010–1018.
[28] J. Wenz, F. Barrantes, Steroid structural requirements for stabilizing or disrupting lipid domains, Biochemistry 42 (2003) 14267–14276.
[29] G. Orädd, V. Shahedi, G. Lindblom, Effect of sterol structure on the bending rigidity of lipid membranes: a 2H NMR transverse relaxation study, Biochim. Biophys. Acta 1788 (2009) 1762–1771.
[30] D.A. Mannock, R.N. Lewis, T.P. McMullen, R.N. McElhaney, The effect of variations in phospholipid and sterol structure on the nature of lipid–sterol interactions in lipid bilayer model membranes, Chem. Phys. Lipids 163 (2010) 403–448.
[31] A. Mead, Review of the development of multidimensional scaling methods, The Statistician 41 (1992) 27–39.
[32] I. Borg, P. Groenen, Modern multidimensional scaling, Theory and Applications, Springer, New York, 1997.
[33] InfoStat v, J.A. Di Rienzo, F. Casanoves, M.G. Balzarini, L. Gonzalez, M. Tablada, C.W. Robledo, Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina, http://www.infostat.com.ar2010.
[34] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, John Wiley Sons, Inc., New York, 1989.

[35] D.G. Kleinbaum, Logistic Regression: A Self-learning Text, Springer-Verlag, New York, 1994.
[36] C.Y.J. Peng, K.L. Lee, G.M. Ingersoll, An introduction to logistic regression analysis and reporting, J. Educ. Res. 96 (2002) 3–13.
[37] SPSS Statistic v. 17.0, IBM Corporation, Somers, NY 10589, , 2008.
[38] R.K. Keller, T.P. Arnold, S.J. Fliesler, Formation of 7-dehydrocholesterol-containing membrane rafts in vitro and in vivo, with relevance to the Smith-Lemli-Opitz syndrome, J. Lipid Res. 45 (2004) 347–355.
[39] M.E. Beattie, S.L. Veatch, B.L. Stottrup, S.L. Keller, Sterol structure determines miscibility versus melting transitions in lipid vesicles, Biophys. J. 89 (2005) 1760–1768.
[40] B.L. Stottrup, S.L. Keller, Phase behavior of lipid monolayers containing DPPC and cholesterol analogs, Biophys. J. 90 (2006) 3176–3183.
[41] T. Róg, M. Pasenkiewicz-Gierula, I. Vattulainen, M. Karttunen, What happens if cholesterol is made smoother: importance of methyl substituents in cholesterol ring structure on phosphatidylcholine–sterol interaction, Biophys. J. 92 (2007) 3346–3357.
[42] M.G.K. Benesch, D.A. Mannock, R.N. McElhaney, Sterol chemical configuration influences the thermotropic phase behaviour of dipalmitoylphosphatidylcholine bilayers containing 5alpha-cholestan-3beta- and 3alpha-ol, Chem. Phys. Lipids 164 (2011) 62–69.
[43] M.G.K. Benesch, D.A. Mannock, R.N. McElhaney, Sterol chemical configuration and conformation influence the thermotropic phase behaviour of dipalmitoylphosphatidylcholine mixtures containing 5beta-cholestan-3beta- and -3alpha-ol, Chem. Phys. Lipids 164 (2011) 70–77.
[44] W. Gao, L. Chen, R. Wu, Z. Yu, P.J. Quinn, Phase diagram of androsterol–dipalmitoylphosphatidylcholine mixtures dispersed in excess water, J. Phys. Chem. B. 112 (2008) 8375–8382.
[45] M.R. Vist, J.H. Davis, Phase equilibria of cholesterol/dipalmitoylphosphatidylcholine mixture: 2H nuclear magnetic resonance and differential scanning calorimetry, Biochemistry 29 (1990) 451–464.
[46] R. Wu, L. Chen, Z. Yu, P.J. Quinn, Phase diagram of stigmasterol–dipalmitoylphosphatidylcholine mixtures dispersed in excess water, Biochim. Biophys. Acta 1758 (2006) 764–771.