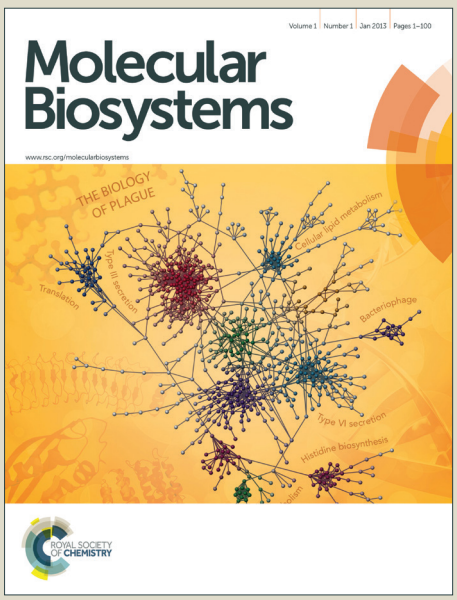
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

ARTICLE

Experimental-Computational Studies of Fatty Acids Distribution Networks

Received 00th January 2015,
Accepted 00th January 2015

DOI: 10.1039/x0xx00000x

www.rsc.org/

Yong Liu [a,b,c], Germán Buendía-Rodríguez [d,*], Claudia Giovanna Peñuelas-Rívas [a], Zhiliang Tan [b], María Rívas-Guevara [e], Esvieta Tenorio-Borroto [a], Cristian R. Munteanu [c,*], Alejandro Pazos [c], Humberto González-Díaz [f,g]

Unbalanced uptake of Omega 6 / Omega 3 (ω-6/ω-3) ratios could increase chronic disease occurrence, such as inflammation, atherosclerosis, or tumor proliferation *etc.*, and the methylation methods for measuring the ruminal microbiome fatty acids (FA) composition/distribution play a vital role in discovering the contribution of food components to ruminant products (*e.g.*, meat and milk) when pursuing a healthy diet. Hansch's models based on Linear Free Energy Relationships (LFER) using physicochemical parameters, such as partition coefficients, molar refractivity, polarizability, *etc.*, as input variables (V_k), are advocated. In this work, a new combined experimental-theoretical strategy was proposed to study the effect of ω-6/ω-3 ratios, FA chemical structure, and other factors over FA distribution networks in the ruminal microbiome. In step 1, experiments were carried out to measure long chain fatty acids (LCFA) profiles in rumen microbiome (bacteria and protozoan), and volatile fatty acids (VFA) in fermentation media. In step 2, the proportions and physicochemical parameter values of LCFA and VFA were calculated in different boundary conditions (c_j) like c_1 = acid and/or base methylation treatments, c_2 = with/without fermentation, c_3 = FA distribution phase (media, bacterial, or protozoan microbiome), *etc*. In step 3, Perturbation Theory (PT) and LFER ideas were combined to develop a PT-LFER model of a FA distribution network using physicochemical parameters (V_k), the corresponding Box-Jenkins (ΔV_{kj}) and PT Operators ($\Delta\Delta V_{kj}$) in statistical analysis. The best PT-LFER model found predicted the effects of perturbations over the FA distribution network with Sensitivity, Specificity, and Accuracy > 80% for 407,655 cases in training + external validation series. In step 4, alternative PT-LFER and PT-NLFER models were tested for training Linear and Non-Linear Artificial Neural Networks (ANN). PT-NLFER models based on ANNs presented better performance but are more complicated than the PT-LFER model. Last, in step 5, the PT-LFER model based on LDA was used to reconstruct the complex networks of perturbations in the FA distribution and compared the giant components of the observed and predicted networks with random Erdős–Rényi network models. In short, our new PT-LFER model is a useful tool for predicting a distribution network in terms of specific fatty acids distribution.

1. Introduction

The ω-6/ω-3 ratio plays an important role not only in the pathogenesis of cardiovascular diseases, but also in cancer, inflammatory and autoimmune diseases[1-3]. A high ω-6/ω-3 ratio is considered detrimental for human health, a value close to 1 is considered protective against the degenerative pathologies [4]. The inconsistent results [5-8] combined with meta-analysis methods reported the contributions of ω-3 fatty acids in cardio- and cerebrovascular diseases, inflammation, or tumor proliferation. Some researchers tend to explain metabolism mechanism not in terms of the absolute amounts of ω-6 and ω-3, but the balance of them.

Enrichment of ruminant meat or milk with ω-3 PUFAs, further to decrease ω-6/ ω-3 ratios uptake, is an efficient way to introduce these beneficial PUFAs into diet, but ruminal complex biohydrogenation process limits their bioavailability [9]. Petit *et al*. reported adding whole linseed riches in ALA to the rations of dairy cows, which resulted in the lowest ω-6/ω-3 ratio in milk compared to micronized soybeans or sunflower seeds [10]. Hess *et al.*[11] proved that the incorporation of ω-3 PUFAs into animal blood and muscle depends directly on the dietary supply of specific fatty acid. In addition, the long chain fatty acids (LCFA) have to be methylated by acid- and/or base- methylation before determining it with gas

[a] Faculty of Veterinary Medicine and Animal Science, Autonomous University of the State of Mexico, Toluca, 50090, México
[b] Key Laboratory of Subtropical Agro-ecological Engineering, Institute of Subtropical Agriculture, the Chinese Academy of Sciences, Changsha, Hunan, 410125, P. R. China
[c] Computer Science Faculty, University of A Coruna, Campus de Elviña s/n, 15071, A Coruña, Spain. E-mail: crm.publish@gmail.com; Tel: +34 981 167 000; Fax: +34 981 167 160.
[d] National Center for Disciplinary Research on Animal Physiology and Breeding, National Institute of Forestry, Agriculture and Livestock Research, Queretaro, 76280, Mexico. E-mail: gerarticulos@hotmail.com; Tel: +52 419 292 0036; Fax: +52 419 292 0033.
[e] Ethnobiology and Biodiversity Research Center, Chapingo Autonomous University, Texcoco, 56230, México.
[f] Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country UPV/EHU, 48940, Leioa, Spain
[g] IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain
† Electronic Supplementary Information (ESI) available:
http://dx.doi.org/10.6084/m9.figshare.1408852

chromatograph (GC). There are different methylation methods for measuring LCFA of milk, muscle or ruminal microbe membrane[12-14], accompany with generating different results. The structure properties of LCFA (especially the number, location or topology structure of double bonds) are highly related to the chronic disease. To address this problem, it was postulated that the LCFAs in ruminal microbial membranes change with the supply of ω-6/ω-3 ratios. This work is aimed to look for a new classification model by means of Chemoinformatics, combined with an original experimental fatty acid distribution in ruminal microbial membrane.

On the other hand, Chemoinformatics is related to Machine Learning, Chemometrics and Bioinformatics[15], and it combines the scientific working fields of Chemistry, Information Science, and the areas of topology, chemical graph theory, and data mining in the chemical space. Corwin Hansch was one of the founders of modern Chemoinformatics, which is based on lipophilicity-activity relationship. A type of Hansch model is the following:[16]

$$f(\varepsilon_i) = a_0 + a_1 \cdot \log P_i + a_2 \cdot pK_a + a_3 \cdot MR - a_4 \cdot (\log P_i)^2 \quad (1)$$

It is well known that steric, electrostatic, and hydrophobicity factors may be biologically relevant[17, 18]. In this equation, the different parameters can be used as inputs to account for the factors: such as water/n-octanol partition coefficients (P_i), molecular refractivity (MR), acidity constants logarithmic (pKa), and other physicochemical parameters to quantify different global molecular properties.[19] The outputs of the model are the values of a molecular property (ε_i) or a function of this property $f(\varepsilon_i)$ for a given chemical compound or molecular entity (m_i). The innovations of these models are described as follows. **1**) The use of the linear regression to seek multivariate linear equations is able to predict the values of $f(\varepsilon_i)$, employing several input variables. **2**) Hansch also generalized the use of lipophilicity parameters by the formulation of parabolic models for non-linear relationships. **3**) The logarithmic terms ($\log P_i$) of P_i are commonly used as the measures of molecular lipophilicity and play an important role in the model. In turn, $\log P_i$ values can be predicted either by atomic methods (like XLogP or ALogP) or by chemical fragment methods (like $CLogP_i$ or similar methods).[20, 21] From a physical-chemistry point of view, Hansch's model is an extra-thermodynamic approach closely related to Linear Free Energy Relationships (LFER).[22, 23] The designation of Hansch's models as LFER equations come from the use of parameters depending on Gibbs free energy (G_i) of the i^{th} process.[24] The changes on the values of this potential during a process obey a logarithmic statistical thermodynamic relationship with equilibrium constant K_i [25].

$$\Delta G_i = -RT\log(K_i) \quad (2)$$

However, in these types of equations, other physicochemical parameters or molecular descriptors can also be used to quantify the effect of changes on the chemical structure over a characteristic of interest. It means that molecular descriptors for a given molecule can be used, which are not only thermodynamic constants, but also other theoretical measures of molecular lipophilicity, electronegativity, polarizability, or molecular topology properties[19], *etc*. The values of these input variables (^iV_k) may be calculated as physicochemical parameters or molecular descriptors of different types (k) for a given molecule (m_i). In fact, the notation can be extended including extra- thermodynamic functions or parameters as follows.

$$f(\varepsilon_i) = \sum_{j=1}^{jmax} a_k \cdot {}^iV_k + \sum_{j=1}^{jmax} b_j \cdot ({}^iV_k)^2 + e_0 \quad (3)$$

Overall, the basic assumption for Hansch's analysis is that similar molecules have similar activities.[26-28] This principle is also called Structure–Activity Relationship (SAR). The SAR paradox refers to the fact that not all cases that similar molecules along with similar activities. The underlying problem is therefore how to define a *small* difference on a molecular level. The problem is relevant since each kind of property, *e.g.*, solubility, reactivity, or metabolism, is expected to depend on another difference. It means that "*small*" variations or perturbations need to be quantified at the molecular structural level, which in turn imply a "*small*" linear change in the free energy of interaction of the drug with the receptor.

In our opinion, the ideas of the Perturbation Theory (PT)[29] can be used to account for this problem in the context of Chemoinformatics. That is why; in this work PT and LFER ideas were used to formulate a new PT-LFER approach. This PT-LFER approach is a generalization of the classic Hansch Extra-Thermodynamics method for Chemoinformatics. The proof-of-concept was also demonstrated with an experimental-theoretical study about complex networks of FA distribution in Lipidomics. To this end, first the experiments were carried out to determine LCFA composition in the rumen microbiome. Next, the Chemoinformatics study was included, starting with the definition, training, and validation of new PT-LFER classification models. Machine Learning methods such as Artificial Neural Networks (ANNs) were used to test PT-NLFER models (Non-Linear analogues of PT-LFER). Next, the best PT-LFER model found was used to predict the effect of perturbations on initial boundary conditions over a large complex network of FA distribution/uptake in the ruminal microbiome. The observed complex network for the data reported was constructed and compared for the first time with the predicted network and model random networks of similar large. Last, the theoretical section was completed with a comparative study of the PT-LFER classification model found with other non-linear models. This study was of major relevance due to previous results that point to a strong relationship between ω-6/ω-3 ratios of FA intake and human health.[30, 31] Accordingly, this work paves the way to evaluate the effect of perturbations on complex molecular systems involved in chemical structures and boundary experimental conditions.

2. Materials and methods

2.1. Experimental section

In the workflow of this experimental part (**Fig. 1**). The general details of the experimental procedures used in experiment 1 and experiment 2 were explained as follows.

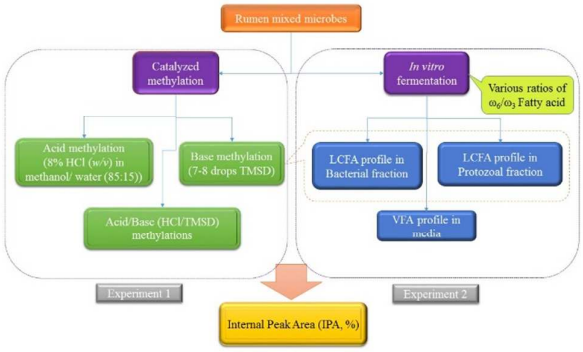


Fig. 1 Workflow of the experimental section (dataset): IPA values of each FA based on bacterial membrane FAs catalyzed with methylation methods (experiment 1) and IPA values obtained from bacteria, protozoan, and media fractions with the fermentation of various exogenous ω-6/ω-3 ratios supplementation with base methylation (experiment 2)

2.1.1. Animal welfare. Three adult male Pelibuey sheep with permanent rumen-fistula (body weight, 45.0 ± 5.0 kg) were used as inoculum donor according to the Mexican Official Standard (NOM-220-SSA1-2002). Nutritional composition of fodder for animal donors was according to National Research Council (NRC).[32] All the animal procedures and protocols were approved by the Animal Care Committee, National Center for Disciplinary Research in Animal Breeding and Physiology (CENID FyMA), National Institute of Forestry, Agriculture and Livestock (INIFAP), Queretaro, Mexico.

2.1.2. Details of *in vitro* fermentation. The *in vitro* details are according to the description of Tang, *el at.*,[33] with the particle-free rumen fluid mixed with the artificial saliva buffer solution[34] in a proportion of 1:2 (*v/v*) at 39 ℃ under continuous flushing of CO_2. Microbial fatty acids were prepared according to the method developed by Or-Rashid.[14] More specifically, the microbial and protozoan samples were separated by differential centrifugation according to the method described by Legay-Carmier and Bauchart.[35]

2.1.3. Specific procedures of the experiment 1. Ruminal mixed microbes without fermentation were catalyzed with acid methylation (8% HCl (w/v) dissolved in methanol/water (85/15)),[12] base methylation (Trimethylsilyldiazomethane, TMSD)[14], combined acid- and base-methylations (first catalyzed with 8% HCl, and subsequently catalyzed with TMSD), respectively. The values of Peak Area, $PA_{(i)}$, for each LCFA under different sets of experimental conditions c_j (different samples) were determined with GC [14, 36] (Model 6890N, Agilent Technologies Inc., USA) with HP-88 Column at laboratory of CENID FyMA, INIFAP, and VFAs were determined with DB-FFAP column. The values of peak area obtained were used to calculate the Internal Peak Area, IPA (%), as follows.

$$\mathrm{IPA}(\%)_{ij} = 100 \cdot \left(\frac{\mathrm{PA}_{(i)}}{\sum_{m_i \subset cj}^{nj} \mathrm{PA}_{(i)}} \right) \quad (4)$$

2.1.4. Specific procedures of the experiment 2. This study was conducted to evaluate the effect of various exogenous ω-6/ω-3 ratios on the biohydrogenation metabolism of the microbial microbiome. The ω-6 and ω-3 PUFAs, linoleic acid (LA, L1376-5g, Sigma-Aldrich) and α-linolenic acid (ALA, L2376-500mg, Sigma-Aldrich) with a total amount of 100 mg/g in substrates, were set as the ratios of 100:0, 90:10, 80:20, 66:33, 50:50 and 20:80, respectively.

Food components used to feed the animals were the same as those used in experiment 1. All fermentation lipid samples extracted with a chloroform–methanol mixture (2:1, *v/v*) [37] from bacterial and protozoan fractions were catalyzed with base methylation (TMSD, herein) [14]. LCFA profiles extracted from bacterial and protozoan fractions and the VFA profiles were determined to calculate IPA (%), and the concentration (mM) of VFA profiles were also calculated.

2.2. Theoretical section

2.2.1. Workflow used for PT-LFER Chemoinformatics study. In the second section, a Chemoinformatics study of the results obtained in the experimental section was carried out. **Fig. 2** shows the workflow diagram that states the integration of both (experimental and theoretical) sections. For the analysis, the chromatographic data about IPA (%) values of fatty acids were collected under different ω-6/ω-3 ratios and experimental conditions c_j. Next, we defined the PT-LFER model. After that, we calculated the values of input variables, including molecular descriptors (^iV_k) of class k^{th} for every i^{th} fatty acid molecule, and perturbation operators $\Delta\Delta V_k(c_j)$. After that we performed the statistical analysis and obtained the PT-LFER model. As follows, more details were explained in some steps.

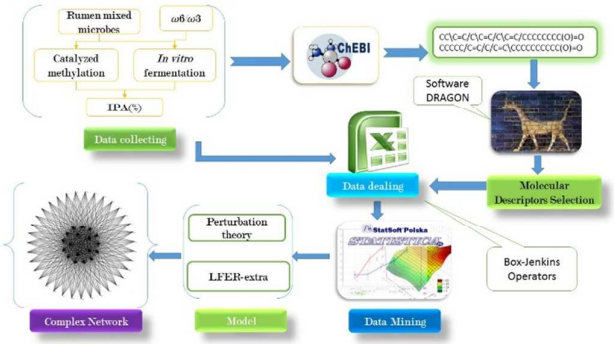


Fig. 2 Workflow used herein to seek PT-LFER models

2.2.2. Theoretical details of the PT-LFER models. In a recent study, Gonzalez-Díaz *et al.* [29] has formulated a general-purpose PT model for multiple-boundary Chemoinformatics problems. In this work, this theory is extended to the study of PT-LFER models of perturbations in complex networks. Let be a general function $f(L_{nr})$ useful to quantify the occurrence (L_{nr} = 1) or not (L_{nr} = 0) of a process involving a set of molecules (m_i) in a complex system. It is considered that all the possible states form a network of states. The network nodes are the initial or reference states (r) linked to their respective final or new states (n) reached by the system after a perturbation of the initial conditions. It separates into a set of multiple initial experimental boundary conditions $^{ref}c_j \equiv (c_0, c_1, c_2, c_3 \ldots c_n)$ (conditions of reference) and a different or new set of boundary conditions $^{new}c_j \equiv (c_0, c_1, c_2, c_3 \ldots c_n)$ (conditions of new) after one or multiple perturbations (changes in these conditions). The PT-LFER model proposed herein is a linear equation with the following form:

$$'f(L_{nr})_{new} = a_0 + a_1 \cdot f(\varepsilon_{ij})_{ref} + a_2 \cdot \langle f(\varepsilon_{ij}) \rangle_{ref} + \sum_{j=1,k=1}^{j=jmax,\ k=kmax} a_{kj} \cdot \Delta\Delta V_k(c_j) \quad (5)$$

The output function '$f(L_{nr})_{new}$ is a score used in Linear Discriminant Analysis (LDA) to calculate the outputs or *posteriori* probability of binary classification of inputs $L_{nr} = 1$ or $L_{nr} = 0$ [38]. The vectors v_i = [$f(\varepsilon_{ij})_{ref}$, $<f(\varepsilon_{ij})>_{ref}$, ω-3, ω-6, $^iV_1$, …$^iV_{kmax}$, $\Delta V_1(c_1)$, …$\Delta V_k(c_j)$,… $\Delta V_{kmax}(c_{jmax})$, $\Delta\Delta V_1(c_1)$, …$\Delta\Delta V_k(c_j)$,… $\Delta\Delta V_{kmax}(c_{jmax})$] are the inputs of this model. Each vector v_i, represents a statistical case (i^{th} case) out of a total of n = 407,655 cases (perturbations). These statistical cases encoded by v_i vectors are perturbations of one entry or state of reference (changes in input parameters) that yield an output or new state. The input vectors v_i include the value of $f(\varepsilon_{ij})_{ref}$ for the state of reference (known value). The vectors v_i also take into account the amounts of ω-6 and ω-3 for the new state (after perturbation). The values of molecular descriptors (^iV_k) used in a classic Hansch's analysis were also included. Last, the inputs also consider the values of the PT-LFER operators $\Delta\Delta V_k(c_j)$.

2.2.3. Calculation of molecular descriptors. In the first work of this series, we used the mean values of atomic electronegativity of the chemical structure descriptors (^iV_k) of a drug [39]. In another recent work, the method for the prediction of peptide epitopes was adapted using the perturbation theory [40]. In the present work, the previous PT models are extended to other direction. Herein, PT models and Hansch's LFER equations are combined to carry out a PT-LFER analysis for the first time. To this end, the following steps were taken. First, structural variables (^iV_k) were used as a new set of molecular descriptors. The values of these variables were calculated with the DRAGON software[41-43]. The first molecular descriptor calculated was V_1 = Mw (Molecular weight). The molecular descriptors V_2 = AEigv, V_3 = AEige, and V_4 = AEigp were included, which are the average eigenvalues of the topological distance matrices weighted with atomic van der Waals volumes (v), Polarizabilities (p), or Electronegativities (e). Last, V_5 = MR (Molecular Refractivity), V_6 = LogP (logarithm of the n-Octanol/Water partition coefficient) were also proposed. The structures of fatty acids were uploaded to DRAGON in a form of Simplified Molecular-Input Line Entry System (SMILES) codes. SMILES codes are very useful to manage molecular structures[44-46] and for further calculation of molecular descriptors[21, 47] (**Table 1**). In our work, the SMILES codes of corresponding fatty acids were downloaded from the website data of Chemical Entities of Biological Interest (ChEBI: http://www.ebi.ac.uk/chebi/).

Table 1. Molecular descriptors (^iV_k) of fatty acids obtained from the ChEBI database

Name of fatty acids in ChEBI [a]	*cis/trans* pattern [b]	Molecular descriptors of FA- Inputs [c]					
		V_1	V_2	V_3	V_4	V_5	V_6
Lauric Acid	*l*	200.4	137.4	127.6	139.5	58.7	4.5
Myristic Acid	*l*	228.4	179.1	169.0	181.2	67.9	5.5
Myristoleic Acid	*c*	226.4	172.1	162.0	174.2	69.0	5.0
Pentadecanoic Acid	*l*	242.5	202.0	191.8	204.1	72.5	5.9
cis-10-Pentadecenoic acid	*c*	240.4	194.8	184.5	196.9	73.6	5.5
Palmitic Acid	*l*	256.5	226.3	216.0	228.4	77.1	6.4
Palmitoleic Acid	*c*	254.5	217.9	207.6	220.1	78.2	5.9
Heptadecanoic acid	*l*	270.5	252.0	241.6	254.1	81.7	6.8
Stearic acid *cis*-10-Heptadecenoic acid	*c*	268.5	243.2	232.8	245.4	82.8	6.4
Stearic acid	*l*	284.5	279.1	268.6	281.2	86.3	7.3
Elaidic acid	*t*	282.5	269.5	259.0	271.7	87.4	6.8
Oleic acid	*c*	282.5	269.5	259.0	271.7	87.4	6.8
Linolelaidic acid	*tt*	280.5	260.9	250.4	263.1	88.5	6.4
Linoleic acid	*cc*	280.5	260.9	250.4	263.1	88.5	6.4
Arachidic acid	*l*	312.6	337.4	326.8	339.6	95.5	8.2
γ-Linolenic acid	*ccc*	278.5	251.5	240.9	253.7	89.6	5.9
Linolenic acid	*ccc*	278.5	255.1	244.5	257.3	89.6	5.9
cis-11.14-Eicosadienoic acid	*ct*	308.6	318.0	307.3	320.2	97.7	7.3
Behenic acid	*l*	340.7	401.3	390.6	403.5	104.7	9.1
cis-8.11.14-Eicosatrienoic acid	*ctt*	306.5	307.4	296.7	309.7	98.8	6.9
Erucic acid	*t*	338.6	390.2	379.4	392.4	105.8	8.7
Acetic acid	*l*	60.1	13.1	6.4	14.4	12.6	-0.2
Propionic acid	*l*	74.1	19.2	11.8	20.7	17.3	0.4
Isobutyric acid	*l*	88.1	24.6	16.7	26.2	21.8	0.9
Butyric acid	*l*	88.1	26.7	18.8	28.3	21.9	0.9
Isovaleric acid	*l*	102.2	33.2	24.9	34.9	26.4	1.1
Valeric acid	*l*	102.2	35.6	27.4	37.4	26.5	1.3

[a] Fatty acids measured for our linear discriminant analysis PT-LFER model.

[b] *cis/trans* pattern, *l* represents *linear*, *c* as *cis-*, *t* as *trans-* PUFAs. The order of *c* or *t* represents the order of initial isomerization characteristics with the tails of PUFAs.

[c] Molecular descriptors (^iV_k) calculated with DRAGON software: V_1 = Mw, V_2 = Aeigv, V_3 = Aeige, V_4 = Aeigp, V_5 = AMR, and V_6 = LogP.

Physical chemistry chemical physics

Journal Name

ARTICLE

2.2.4. Calculation of PT operators. When the previous equation of the PT-LFER model was expanded, two types of input terms can be observed. The first type of term is the function $f(\varepsilon_{ij})_{ref}$. This function takes the values, $f(\varepsilon_{ij})_{ref} = \langle\varepsilon_{ij}\rangle_{ref}$ = IPA(%)$_i$ for each sample. IPA(%)$_i$= 100·(PA_i/PA_{jmax}) is the Internal Peak Area proportion, used to quantify the experimental proportion of a fatty acid determined by GC. It means that $f(\varepsilon_{ij})_{ref}$ is the measured value of the proportion of a fatty acid under the same conditions c_j. The second class refers to the perturbation terms $\Delta\Delta V_k(c_j)$. The parameters $\Delta\Delta V_k(c_j)$ are useful to quantify the effect of perturbations of different boundary conditions (c_j) over the output $'f(L_{nr})_{new}$, which was defined herein as a discrete value function (occurrence or not of links in the network) for the classification purposes. The difference $\Delta\Delta V_k(c_j)$ between the final or new state ($\Delta V_k(c_j)_{new}$) and the initial or reference state ($\Delta V_k(c_j)_{ref}$) is the additive perturbation for a component in the $\Delta V_k(c_j)$. When the output of this equation is $'f(L_{nr})_{new} > 'f(L_{nr})_{ref}$ => L_{nr} = 1 => IPA(%)$_{new}$ > IPA(%)$_{ref}$; and consequently (=>) the distribution or proportion of the FA in the new state is higher than in the reference state, otherwise L_{nr} = 0.

$$'f\left(L_{nr}\right)_{new} = a_0 + a_1 \cdot 'f\left(L_{nr}\right)_{ref} + a_1 \cdot \left\langle \varepsilon_{ij} \right\rangle_{ref} + \sum_{j=1,\,k=1}^{j=jmax,\;k=kmax} a_{ijk} \cdot \Delta\left({}^{i}V_k - \left\langle V_k\left(c_j\right)\right\rangle\right) \quad (6)$$

$$\Delta\Delta V_k\left(c_j\right) = p\left(c_j\right)_{new} \cdot \left({}^{i}V_k - \left\langle V_k\left(c_j\right)\right\rangle\right)_{new} - p\left(c_j\right)_{ref} \cdot \left({}^{i}V_k - \left\langle V_k\left(c_j\right)\right\rangle\right)_{ref} \quad (7)$$

2.2.5. Calculation of Box-Jenkins Operators. A close inspection of the perturbation terms shows that they are probability-weighted differences (Δ) of Box-Jenkins Operators $\Delta V_k(c_j)$. The values of Box-Jenkins Operators $\Delta V_k(c_j)$ of the molecular descriptors ($^{i}V_k$) were calculated to quantify the effect of deviations of a molecule (m_i) from the average behavior of all molecules measured under the same set of conditions (c_j) of the complex system. Deviations due to the changes on different boundary conditions (c_j) were taken into account. The boundary conditions refer to preliminary operational conditions, c_1 = refers to the use of different experimental treatments, c_2 = with/without fermentation, c_4 = is the gas chromatographic protocol used, and c_5 = refers to the use of replicate experiments. Others are more directly related to the posterior distribution and nature of the FA, c_3 = the biological phase of distribution of LCFAs, and c_6 = quantify information about *cis/trans* geometric pattern present in the LCFAs. All data was processed in an Excel file. In Excel, the values of $\Delta V_k(c_j)$ were calculated considering various experimental boundary conditions (c_j). The probabilities are $p(c_j) = n_j/n_{total}$; n_j number of experimental entries for condition c_j and n_{total} = 744 total number of experimental entries (the total number of IPA(%) measured in this work). The average value $\langle\Delta V_k(c_j)\rangle$ (**Table 2**) is the difference of the function value $^{i}V_k$ with the average molecular descriptors $\langle V_k\rangle$ for a specific boundary condition c_j, see the equations:

$$\Delta V_k\left(c_j\right) = \left({}^{i}V_k - \left\langle V_k\left(c_j\right)\right\rangle\right) \quad (8)$$

$$\left\langle V_k\left(c_j\right)\right\rangle = \frac{1}{n_j}\left(\sum_{m_i \subset c_j}^{n_j} {}^{i}V_k\right) \quad (9)$$

Table 2. Average values of input variables ($\langle V_k\rangle$) for experimental boundary conditions (c_j)

Experimental boundary condition		Average eigenvalues of input variables $\langle V_k(c_j)\rangle$ [aa]						$p(c_j)$ [ab]
Conditions (c_j)	level	V_1	V_2	V_3	V_4	V_5	V_6	
c_1 => Treatments [a]	BM	274.4	258.3	247.9	260.5	84.8	6.6	0.113
	AM	274.4	258.3	247.9	260.5	84.8	6.6	0.113
	CM	274.4	258.3	247.9	260.5	84.8	6.6	0.113
	BA	261.9	234.6	224.3	236.8	80.3	6.2	0.258
	PA	261.9	234.6	224.3	236.8	80.3	6.2	0.258
	MA	85.8	25.4	17.7	27.0	21.1	0.7	0.145
c_2 => fermentation [b]	***0***	274.4	258.3	247.9	260.5	84.8	6.6	0.339
	1	223.2	188.7	178.9	190.7	67.3	5.0	0.661
c_3 => Phase [c]	Bacteria fraction	269.0	248.1	237.7	250.3	82.9	6.4	0.597
	Protozoa fraction	261.9	234.6	224.3	236.8	80.3	6.2	0.258
	Media fraction	85.8	25.4	17.7	27.0	21.1	0.7	0.145
c_4 => Column of GC [d]	HP-88 (112-88A7)	266.8	244.0	233.6	246.2	82.1	6.4	0.855
	DB-FFAP	85.8	25.4	17.7	27.0	21.1	0.7	0.145

c_5 => Replicate (*r*-error) [e]	***0***	242.8	213.9	203.8	216.0	73.9	5.6	0.391
	0.1	242.8	213.9	203.8	216.0	73.9	5.6	0.391
	0.2	205.8	173.6	164.2	175.6	61.6	4.4	0.133
	0.3	274.4	258.3	247.9	260.5	84.8	6.6	0.085
c_6 => *cis*/*trans* pattern [f]	*linear*	203.2	180.8	171.5	182.8	59.6	4.6	0.238
	cis	278.5	269.1	258.6	271.3	86.1	6.7	0.111
	trans	310.6	329.8	319.2	332.1	96.6	7.7	0.032
	trans, trans	280.5	260.9	250.4	263.1	88.5	6.4	0.016
	cis, cis	308.6	320.8	310.1	323.0	97.7	7.3	0.032
	cis, trans	289.9	279.6	269.0	281.8	91.6	6.7	0.048
	trans, cis	280.5	261.1	250.5	263.3	88.5	6.4	0.016
	cis, cis, cis	286.5	269.9	259.3	272.1	92.3	6.2	0.111
	cis, trans, trans	292.5	280.0	269.3	282.2	94.2	6.4	0.032
	trans, cis, trans	278.5	254.6	244.0	256.8	89.6	5.9	0.016
	trans, trans, trans	278.5	252.0	241.4	254.2	89.6	5.9	0.032
	cis, cis, trans	278.5	252.5	241.9	254.8	89.6	5.9	0.016
	cis, trans, cis	278.5	252.0	241.4	254.2	89.6	5.9	0.032

[a] "BM" means base methylation without fermentation; "AM" means acid methylation without fermentation; "CM" means acid- and base-combined methylation; "BA" means fatty acids from bacterial fraction after 48h fermentation; "PA" means fatty acids from protozoan fraction after 48h fermentation; "MA" means volatile fatty acids from media fraction after 48h fermentation.

[b] "***0***" means the dataset from experiment 1 without fermentation; "***1***" means the dataset from experiment 2 with fermentation of omega 6 and omega 3.

[c] "phase", means the dataset: long chain fatty acids including from the bacterial membrane (bacterial fraction), protozoan membrane (protozoan fraction), volatile fatty acids from fermentation media (media fraction).

[d] Column of GC, "HP-88 (112-88A7)" means the column of GC for determining long chain fatty acids; "DB-FFAP" means the column for determining volatile fatty acids.

[e] "***0***" means the original data, "***0.1***, ***0.2***, or ***0.3***" means the 1, 2, or 3 replicates, respectively.

[f] *cis*/*trans* pattern: "*linear*" means LCFA without double bonds; "*cis*" means LCFA with *cis* isomerization; "*trans*" means LCFA with *trans* isomerization; and the number of cis or trans means LCFA with the same number of *cis* or *trans* double bonds.

[aa] $\langle V_k(c_j) \rangle$ means the average of Molecular descriptors (V_k) for different conditions (c_j); the descriptors are V_1 = Mw, V_2 = Aeigv, V_3 = Aeige, V_4 = Aeigp, V_5 = AMR, and V_6 = LogP.

[ab] $p(c_j) = n_j / n_{total}$; n_j number of experimental entries for condition c_j and n_{total} = 744 total number of experimental entries.

2.2.6. Dataset. Predicting the effect of perturbations in input conditions over the output property is the aim of this model. For it, we need to infer the value of the property in a new set of conditions using a known experimental value as reference. It means that we need to predict the variation of the experimental properties for pairs of data cases (reference and new). Consequently, if we have an original dataset with *n* cases we need to explore a total of n^2 cases for an exhaustive investigation of the data space (all pairs of data). If *n* is big, the number of pairs increases notably. Consequently, we carry out a random sampling procedure. We generated as many as possible pairs of data that we can process with one MS Excel sheet selecting at random both the reference and the new state. The MS Excel function random, if has been used to generate pairs of random numbers between one and *n*. The very high number of 407,655 perturbations was the higher number of pairs of cases we was able to handle in Excel with our processing power.

2.2.7. Classification models. The Linear Discriminant Analysis (LDA)[38] algorithm implemented in the STATISTICA software was used to find the best PT-LFER model. Sometimes the relationship between the input variables and the output is more complex and the linearity cannot solve the problem. Therefore, the non-linear models could provide a better solution, but with the drawback of not being able to interpret the model and the relations between the variables. Thus, the Artificial Neural Networks (ANNs) [48] were tested: Linear Neural Networks (LNNs), which are similar to the LDA models, and non-linear Multi-Layer Perceptron (MLPs) [38]. The full datasets were randomly split into training series ("t", 75%) used for model construction, and validation series ("v", 25%) used for model validation. In addition, a cross-validation variable was added to the dataset with the test values of "t" and "v". All independent variables were unified and standardized with the STATISTICA software, prior to model construction.

2.2.8. Complex Networks study. Both the observed and predicted networks were constructed in Excel and saved in the **.net** (lists of pairs of nodes) file format. The links of the observed network coincide with the classes to be predicted by the previous LDA model. The existence of a link corresponding to the condition L_{nr} = 1, if $IPA(\%)_{obs} > IPA(\%)_{ref}$ for each fatty acid at both the initial and final states, L_{nr} = 0 otherwise. A number of pairs of states as high as possible was generated, calculating the existence of observed links with the previous rule, and they were also predicted with the model. These files were processed with the CentiBiN software described by Junker *et al*.[49], to calculate the average indices of the topology of the network. The indices calculated were the average values of the vertex-vertex topological distance [50], node degree, and closeness of the giant component of the observed, predicted and the two similar random ones. Two models of random networks (random network 1 and 2) were also built. The model of the

random network selected was the Erdös–Rényi graph (ER), which often used as a random network model.

3. Results and discussion

3.1. Catalyzed methylation in ruminal microbes

Saponification followed by methylation is a classic method for the preparation of FA methyl esters. **Table 3** shows the results obtained in the experimental determination of the values of IPA (%) after different acid-/base- methylations. This table reports the Average and Standard Deviation (SD) values of IPA (%) for those fatty acids for the first time. In general, base-catalyzed methylation proceeds more rapidly under mild temperature conditions than acid-catalyzed reactions.[51]

Table 3. Internal peak area values, IPA (%) [a], of LCFA profiles of ruminal mixed microbes by acid- and/or base-catalyzed methylation

Name of fatty acids		Average [b]			SD		
		B	A	A&B	B	A	A&B
Lauric Acid	C12:0	0.55	0.52	1.87	0.07	0.16	0.76
Myristic Acid	C14:0	1.06	1.24	2.12	0.05	0.26	0.69
Myritoleic Acid	C14:1	1.95	2.31	2.38	0.11	0.49	0.69
Pentadecanoic Acid	C15:0	1.20	1.47	1.60	0.02	0.35	0.43
cis-10-Pentadecenoic acid	C15:1	0.40	0.53	0.68	0.08	0.10	0.25
Palmitic Acid	C16:0	17.91	18.86	19.70	0.54	1.73	0.98
Palmitoleic Acid	C16:1	1.85	2.30	2.14	0.07	0.32	0.18
Heptadecanoic acid	C17:0	0.99	1.39	1.49	0.25	0.20	0.28
cis-10-heptadecanoic acid	C17:1	0.00	0.17	0.53	0.00	0.20	0.29
Stearic acid	C18:0	49.76	49.17	40.21	1.02	1.62	0.80
Elaidic acid	C18:1 n9t	8.68	8.23	8.76	0.65	0.69	1.30
Oleic acid	C18:1 n9c	9.00	8.69	8.17	0.09	0.88	2.12
Linolelaidic acid	C18:2 n6t	0.65	0.74	0.85	0.04	0.05	0.64
Linoleic acid	C18:2 n6*c*	2.37	2.63	2.81	0.06	0.11	0.36
Arachidic acid	C20:0	0.84	0.68	1.23	0.17	0.07	1.50
γ-Linolenic acid	C18:3 n6	0.45	0.58	0.80	0.06	0.14	0.34
Linolenic acid	C18:3 n3	0.00	0.00	1.52	0.00	0.00	1.25
cis-11,14-Eicosadienoic acid	C20:2	0.34	0.14	0.75	0.67	0.16	0.63
Behenic acid	C22:0	0.59	0.13	1.33	0.15	0.15	0.78
cis-8,11,14-Eicosatrienoic acid	C20:3 n6	0.58	0.05	0.82	0.22	0.10	1.11
Erucic acid	C22:1 n9	0.82	0.17	0.24	0.42	0.19	0.47
Unsaturated fatty acids		27.10	26.55	30.44	1.46	1.08	3.29
Long chain fatty acids >= 18 carbons		74.07	71.20	67.50	0.95	3.51	3.76
18 carbons unsaturated fatty acids		21.15	20.88	22.91	0.68	1.46	4.03
cis-fatty acids		11.36	11.32	10.98	0.06	0.79	1.88
trans-fatty acids		9.34	8.97	9.61	0.61	0.64	1.89
Ratios (*cis*/*trans*)		1.22	1.26	1.14	0.07	0.05	0.11
Ratios (Stearic acid: Palmitic acid)		2.78	2.61	2.04	0.03	0.32	0.12
Odd-carbon fatty acids		2.60	3.56	4.30	0.31	0.59	0.85
Even-carbon saturated fatty acids		70.71	70.59	66.47	1.25	0.69	2.96
Even-carbon unsaturated fatty acids		26.70	25.85	29.23	1.52	1.13	3.76
Saturated/unsaturated fatty acids		2.69	2.77	2.28	0.19	0.15	0.40

[a] Internal peak area values, IPA(%), means the relative proportion of different fatty acids in the corresponding individual sample. [b] "A" = acid methylation, "B" = base methylation, or "A&B" = acid methylation with subsequent base methylation. Average and Standard Deviation (SD) of IPA (%) values for long chain fatty acids.

3.2. LCFA profiles in bacteria and protozoan

This work is focused on the lipid metabolism of exogenous FAs by direct determination of the IPA (%) values of LCFA from ruminal microbe/protozoa biological membrane, including FAs from bacteria (**Table 4**) and protozoan (**Table 5**) biological membrane under different experimental conditions (c_j).

It is well known that the imbalance of ω-6/ω-3 ratios in the diet has the potential to induce inflammation, asthma, arthritis, vascular disease [52], but high level of ω-3 exert a suppressive effect [53-56]. As expected, the FA composition of ruminal bacteria and protozoa biological membrane, VFAs in media were indeed changeable with the exogenous ω-6/ω-3 PUFAs ratios. This study had no significance statistical difference in the main IPA (%) of fatty acids (e.g., C16:0, C18:0), but some valuable information has still been extracted from

the results. First of all, *cis*-FAs content increased with exogenous ω-3 PUFA, *trans*-FAs decreased in bacteria biological phase. For example, the *cis*-FA profiles in ω-6/ω-3 = 20:80 was 1.76 times (bacterial phase) and 1.60 times (protozoan phase) than that in ω-6/ω-3 = 100:0, and *trans*-FA profiles in ω-6/ω-3 = 100:0 was 1.24 times (bacterial phase) and 0.98 time (protozoan phase) than that in ω-6/ω-3 = 20:80. This directly results in the increasing ratio of *cis*/*trans*- fatty acid compositions with the increasing of the exogenous ω-3 PUFA, such as 2.18 times in bacterial phase and 1.58 times in protozoan phase, when ω-6/ω-3 = 20:80 compared with ω-6/ω-3 = 100:0, respectively. It means exogenous PUFAs are degraded by rumen microorganisms, or have more complex metabolism processes to leading to intermediary metabolism with both of *cis*- and *trans*- unsaturated FAs formulation. The biohydrogenation of linoleic acid (LA, *cis* 9, *cis* 12- C18:2) in rumen is isomerized to *cis* 9, *trans* 11- C18:2 isomer (conjugated linoleic acid, CLA), conversion of this isomer to *trans* 11- C18:1 (vaccenic acid), and reduction to stearic acid (C18:0) [57]. Whereas the bio-hydrogenation of α-linolenic acid (ALA) is characterized by isomerization to 9, 11, 15- *cis, trans, cis*- C18:3 isomer and subsequent reduction via *cis*, *trans* isomers C18:2, C18:1 and then to stearic acid [58]. This research showed that ω-3 PUFA (α-linolenic acid) could increase the *cis*-FAs content compared to ω-6 PUFA (linoleic acid) on both of bacteria and protozoa phases.

Secondly, IPA values of C16:0 and C18:0 in bacterial phase were 18.7% and 57.8%, whereas those in protozoan phase were 13.3% and 67.5%, respectively. The exogenous ω-6/ω-3 ratios have no significant effect on these two major fatty acids in both bacterial and protozoan phases. However, the minor difference of lipid composition like ratios of palmitic/stearic acid, or unsaturated/saturated fatty acids on bacterial and protozoan biological membrane, may trigger a great difference in the function of membranes of bacteria and protozoan (*e.g.*, membrane fluidity, permeability, hydrophobicity and stability)[59, 60], or further in the functional groups [61, 62] such as specific peptide, enzymes, or channels, *etc*. Stearic acid in protozoa membrane is higher (about 14.0%) than that in bacteria membrane, unlike other fatty acids.

Table 4. Internal peak area values, IPA (%), of LCFA profiles in bacterial fraction [a]

Name of fatty acids	Various exogenous ω-6/ω-3 PUFAs Ratios (x: [100-x]; in total of 100 mg/g substrate)						Average
	100-0	*90-10*	*80-20*	*66-33*	*50-50*	*20-80*	
Lauric Acid	0.62	0.67	0.56	0.55	1.89	0.52	0.80
Myristic Acid	1.53	1.45	1.31	1.27	1.71	1.23	1.42
Myritoleic Acid	2.66	2.61	2.32	2.20	1.90	2.06	2.29
Pentadecanoic Acid	2.13	1.86	1.82	1.68	1.53	1.61	1.77
cis-10-Pentadecenoic acid	0.83	0.73	0.64	0.58	0.45	0.44	0.61
Palmitic Acid	18.38	18.73	18.82	18.82	18.79	18.39	18.65
Palmitoleic Acid	1.45	1.91	1.97	1.29	1.72	1.94	1.71
Heptadecanoic acid	1.77	2.16	2.01	1.73	1.65	1.92	1.87
cis-10-Heptadecenoic acid	0.62	0.41	0.16	0.00	0.25	0.25	0.28
Stearic acid	56.93	56.17	58.64	58.97	58.22	57.58	57.75
Elaidic acid	5.95	6.61	5.71	5.59	4.89	5.01	5.63
Oleic acid	2.93	3.86	3.07	3.59	3.90	5.33	3.78
Linolelaidic acid	2.09	0.97	1.18	1.65	1.05	1.47	1.40
Linoleic acid	1.03	1.07	1.04	1.02	1.24	1.62	1.17
Arachidic acid	0.56	0.59	0.46	0.68	0.55	0.33	0.53
γ-Linolenic acid	0.53	0.22	0.28	0.37	0.25	0.31	0.33
Unsaturated fatty acids, %	18.07	18.38	16.38	16.29	15.66	18.43	17.20
Long chain fatty acids ≥ 18 carbons, %	70.02	69.49	70.39	71.87	70.09	71.65	70.58
18 carbons unsaturated fatty acids, %	12.52	12.73	11.28	12.22	11.33	13.74	12.30
cis- fatty acids, %	3.95	4.93	4.12	4.61	5.13	6.95	4.95
trans- fatty acids, %	8.04	7.58	6.89	7.24	5.94	6.48	7.03
Ratios (*cis*-/*trans*-)	0.492	0.650	0.598	0.637	0.864	1.072	0.72
Ratios (Stearic acid: Palmitic acid)	3.098	2.999	3.116	3.134	3.098	3.131	3.10
Odd-carbon fatty acids, %	5.34	5.15	4.63	4.00	3.89	4.22	4.54
Even-carbon saturated fatty acids, %	78.03	77.60	79.80	80.29	81.16	78.04	79.15
Even-carbon unsaturated fatty acids, %	16.63	17.24	15.58	15.71	14.95	17.74	16.31
Saturated/unsaturated fatty acids	4.53	4.44	5.11	5.14	5.39	4.43	4.81

[a] Internal peak area values, IPA (%), means the relative proportion (%) of different fatty acids in the corresponding individual sample.

Journal Name

ARTICLE

Thirdly, even-carbon saturated FA in the treatment of ω-6/ω-3 = 100:0 and ω-6/ω-3 = 20:80 were 85.1% and 81.7% in the protozoan phase, whereas even-carbon unsaturated FA in those treatments were 13.0% and 15.8%, respectively (**Table 5**). Herein, the highest value of even-carbon unsaturated FA appeared in the treatment of ω-6/ω-3 = 20:80, this phenomenon also happened in the bacterial phase, which means that, according to this study, a high amount of ω-3 PUFAs has the tendency to increase the even-carbon unsaturated FAs compared to ω-6 PUFAs.

An interesting fact is that linolelaidic acid (C18:2, 6*t*) proportion in ω-6/ω-3 = 100:0 was a little higher than other treatments in bacterial phase, whereas a higher proportion appeared in protozoan phase when ω-6/ω-3 = 20:80. This might be due to the biohydrogenation of PUFAs (such as linoleic and linolenic acid) in rumen resulting in the production of primarily *trans*- fatty acids and stearic acid.[63] All the difference in fatty acid distributions between bacterial and protozoan phases can reflect the protozoan and bacteria having a different and complex metabolism in the processes of assimilation, absorption, degradation or *de novo* synthesis. Thanks to these differences, that makes the different phases (bacterial and protozoan phase), as an important input variable, more reasonable in our new Hansch Perturbation Theory – LFER model. The LCFAs and VFAs distribution was stated in the entire fermentation system of ruminal micro-niche environment, shown in **Fig. 3**.

Table 5. Internal peak area values, IPA (%), of LCFAs in protozoan fraction [a]

Name of fatty acids	Various exogenous ω-6/ω-3 PUFAs Ratios (x: [100-x]; in total of 100 mg/g substrate)						Average
	100-0	*90-10*	*80-20*	*66-33*	*50-50*	*20-80*	
Lauric Acid	0.61	0.50	1.09	0.25	0.28	0.21	0.49
Myristic Acid	0.76	1.07	1.16	0.73	0.93	0.68	0.89
Myritoleic Acid	0.63	0.96	0.88	0.75	0.87	0.59	0.78
Pentadecanoic Acid	0.79	1.24	1.04	0.93	1.00	0.73	0.96
cis-10-Pentadecenoic acid	0.48	0.67	0.66	0.32	0.59	0.48	0.53
Palmitic Acid	11.39	13.46	15.78	13.16	14.07	11.93	13.30
Palmitoleic Acid	0.39	0.89	0.92	0.99	1.10	0.89	0.86
Heptadecanoic acid	0.44	1.72	1.70	1.26	1.59	1.19	1.31
cis-10-Heptadecenoic acid	0.20	0.15	0.17	0.17	0.19	0.10	0.16
Stearic acid	71.29	66.40	63.53	68.44	66.89	68.55	67.52
Elaidic acid	7.29	6.67	6.63	6.50	6.20	5.83	6.52
Oleic acid	3.04	3.50	3.90	4.04	3.98	5.08	3.93
Linolelaidic acid	0.69	0.59	0.60	0.88	0.95	2.25	0.99
Linoleic acid	0.78	1.44	1.05	0.83	0.94	1.05	1.02
Arachidic acid	1.01	0.42	0.53	0.47	0.24	0.31	0.50
γ-Linolenic acid	0.21	0.33	0.35	0.28	0.16	0.14	0.24
Unsaturated fatty acids, %	13.71	15.20	15.17	14.77	14.99	16.41	15.04
Long chain fatty acids ≥ 18 carbons, %	84.31	79.34	76.60	81.45	79.37	83.21	80.71
18 carbons unsaturated fatty acids, %	12.01	12.52	12.54	12.54	12.24	14.36	12.70
cis- fatty acids, %	3.82	4.94	4.96	4.87	4.92	6.13	4.94
trans- fatty acids, %	7.98	7.25	7.23	7.38	7.16	8.09	7.51
Ratios (*cis*-/*trans*-)	0.480	0.681	0.685	0.661	0.688	0.758	0.66
Ratios (Stearic acid: Palmitic acid)	6.258	4.933	4.027	5.202	4.753	5.746	5.15
Odd-carbon fatty acids, %	1.90	3.78	3.56	2.68	3.37	2.50	2.96
Even-carbon saturated fatty acids, %	85.07	81.85	82.10	83.05	82.42	81.67	82.69
Even-carbon unsaturated fatty acids, %	13.04	14.38	14.34	14.28	14.21	15.83	14.34

Saturated/unsaturated fatty acids	6.29	5.58	5.60	5.78	5.67	5.09	5.65

[a] Internal peak area values, IPA (%), means the relative proportion (%) of different fatty acids in the corresponding individual sample.

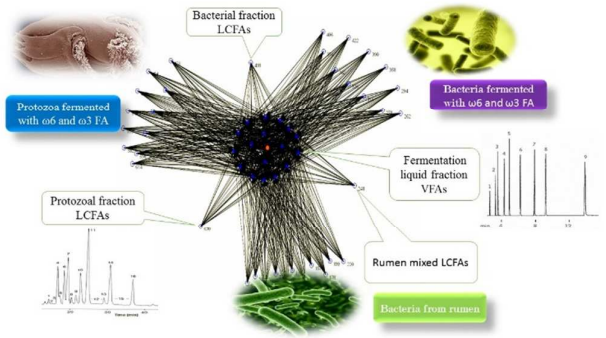


Fig. 3 Illustration of the Sub-network of LCFAs and VFAs in rumen micro-niches

3.3. VFA profiles in media provided with exogenous ω-6/ω-3 PUFAs

In this study, the peak area (PA) of VFA was determined in each sample to calculate the internal peak area, IPA (%), at 48 h fermentation. On the other hand, the absolute concentration of VFA was also calculated, using the PA combined with the corresponding standard curve in the same situation (**Table 6**). Volatile fatty acids (VFAs) are widely regarded as secondary metabolites to reflect the hydrogenation metabolism of lipids, microbial degradation enzyme activity, and the lifecycle of microbial organism in the rumen micro-ecological niche.[64, 65] The acetic, propionic and butyric acids are the major VFAs, with proportions of 43.6%, 23.1% and 19.2%, respectively. Meanwhile, the residue VFAs, including isobutyric, isovaleric and valeric acid had a total proportion of 14.1%. It is noteworthy that there was no significant difference in VFAs when supply of different exogenous ω-6/ω-3 PUFAs ratios. The values of all VFAs in the treatment of ω-6/ω-3 = 80:20 were lower than others, which might be the result of the bottle cap of storage containers which was broken or was not sealed properly. However, this study is focused on the proportion peak areas of VFA on the same sample or treatment. It can be concluded that acetic acid had a minor decrease, but propionic acid had a slight increase, with an increasing proportion of ω-3 PUFAs in the total supplementation exogenous PUFAs. Even if both acetic acid and propionic acid changed a little in terms of different proportion of ω-6 and ω-3 PUFAs in a total of 100 mg/g substrate, the ratio of acetic and propionic acid was regularly decreased, as increasing proportion of ω-3 PUFAs with an average of 1.89.

Table 6. Internal peak area values, IPA (%), and absolute Concentration (mM) of volatile fatty acids (VFAs) in media fraction [a]

VFA name	IPA (%) values of VFAs supplemented with various exogenous ω-6/ ω-3 PUFAs Ratios [b]						Average
	100-0	90-10	80-20	66-33	50-50	20-80	
Acetic acid	44.72	43.86	43.61	43.62	42.83	42.74	43.56
Propionic acid	22.80	22.62	22.90	23.15	23.36	23.66	23.08
isobutyric acid	2.48	2.57	2.78	2.37	2.39	2.65	2.54
Butyric acid	18.31	19.06	19.55	19.21	19.66	19.62	19.24
isovaleric acid	5.93	5.99	5.84	5.89	5.77	5.81	5.87
Valeric acid	5.77	5.89	5.32	5.77	5.99	5.51	5.71
Ac/Pro	1.96	1.94	1.90	1.88	1.83	1.81	1.89
VFA name	Absolute concentration (mM) of VFAs supplemented with various exogenous ω-6/ ω-3 PUFAs Ratios [c]						(r^2) [d]
	100-0	90-10	80-20	66-33	50-50	20-80	
Acetic acid	124.4 ±1.66	104.2 ± 2.89	75.6 ± 2.26	114.1 ± 1.85	109.4 ± 1.01	110.5 ±0.71	0.9964
Propionic acid	31.6 ± 0.54	26.8 ± 0.71	19.8 ± 0.56	30.2 ± 0.59	29.7 ± 0.37	30.5 ± 0.22	0.9974
isobutyric acid	2.3 ± 0.04	2.1 ± 0.02	1.6 ± 0.07	2.1 ± 0.02	2.1 ± 0.02	2.3 ± 0.27	0.9985
Butyric acid	19.1 ± 0.41	16.9 ± 0.87	12.6 ± 0.22	18.8 ± 1.15	18.8 ± 0.35	19.0 ± 0.10	0.9983
isovaleric acid	5.0 ± 0.11	4.3 ± 0.13	3.1 ± 0.10	4.6 ± 0.21	4.4 ± 0.04	4.5 ± 0.04	0.9988
Valeric acid	4.8 ± 0.10	4.2 ± 0.10	2.8 ± 0.22	4.5 ± 0.03	4.6 ± 0.04	4.3 ± 0.25	0.9986
Ac/Pro [e]	3.93 ± 0.02	3.89 ± 0.01	3.83 ± 0.01	3.78 ± 0.02	3.68 ± 0.04	3.63 ± 0.01	-

[a] Internal peak area values, IPA(%), means the relative proportion (%) of different fatty acids in the same individual sample.
[b] The entire supplementation amount of ω-6/ ω-3 PUFAs was standard: 100 mg/g alfalfa substrate, means ± standard errors.
[c] Standard curve was used to calculate the values of each VFA; the standard curve equation is $f(x) = a \cdot x + b$, with $f(x)$ = concentration of the VFAs, x = Peak Area (PA), a = coefficient of peak area, b = intercept. The values of (a, b) found for different VFAs were the following: for Acetic acid (3,000,000; 3,000,000), Propionic acid (6,000,000; 1,000,000), isobutyric acid (9,000,000; 465,874), Butyric acid (8,000,000; 767,690), isovaleric acid (10,000,000; 501,483), and Valeric acid (10,000,000; 644,543).
[d] "r^2" the correlation coefficient square of standard curve for calculating each corresponding volatile fatty acid.
[e] Ac/Pro = the ratio of acetic acid with propionic acid.

Physicochemical Interactions

Journal Name

ARTICLE

3.4. PT-LFER model for FA distribution network

A new model was developed useful to predict the proportion of FAs (LCFAs and VFAs) in different phases of ruminal microbiome within/without exogenous PUFAs after perturbations in chemical molecular descriptors (V_k) and under initial experimental boundary conditions (c_j). Each value represents a corresponding coefficient in the new model for predicting the IPA(%)$_{new}$ of each FA (**Table 7**). As explained, this model can classify as high (L_{nr} = 1)/ low (L_{nr}= 0) the expected proportion of FAs (LCFA/VFA) between the new and reference states after changing the boundary conditions c_j. The parameter n(L_{nr} = 1) represents the number of cases in the sub-set with L_{nr} = 1 (links in the network), or the same with IPA(%)$_{new}$ of new sub-set is higher than that of reference IPA(%)$_{ref}$. On the other hand, n(L_{nr} = 0) represents the number of cases observed and predicted in the sub-set with L_{nr} = 0 (not connected nodes) or explained as that the IPA(%)$_{new}$ value is lower than IPA(%)$_{ref}$ value. The best PT-LFER model found using the LDA algorithm has only 12 variables and it is described by the following algorithm.

$$
\begin{aligned}
{}'f(L_{nr})_{new} = {} & -0.021 \cdot f(\varepsilon_{ij})_{ref} + 0.0026 \cdot \langle f(\varepsilon_{ij}) \rangle_{ref} + 0.3713 \cdot {}^{new}V_6 + 1.0709 \cdot {}^{new}\omega 6 - 1.1264 \cdot {}^{new}\omega 3 \\
& + 0.0237 \cdot \Delta\Delta V_5(c_1) - 0.0063 \cdot \Delta\Delta V_6(c_2) + 0.0044 \cdot \Delta\Delta V_1(c_4) - 0.0037 \cdot \Delta\Delta V_1(c_5) \\
& - 0.0036 \cdot \Delta\Delta V_4(c_6) - 0.1682 \cdot ({}^{new}V_6)^2 + 0.0182 \cdot (\Delta\Delta V_6(c_3))^2 - 13.7236 \\
& N = 407,655; \quad \chi^2 = 244,532.9; \qquad p < 0.005
\end{aligned}
\tag{10}
$$

Table 7. Details of the PT-LFER model for the distribution network of fatty acids

Coeff.	Variable [a]	Value	Classic Symbols [b]	PT operators ($\Delta\Delta V_k(c_j)$) [c]
a_0	Intercept	-13.7236	-	-
a_1	$f(\varepsilon_{ij})_{ref}$	-0.0210	-	-
a_2	$\langle f(\varepsilon_{ij}) \rangle_{ref}$	0.0026	-	-
a_3	${}^{new}V_6$	0.3713	V_6 = LogP	-
a_4	${}^{new}V_7$	1.0709	$V_7 = (LogP)^2$	-
b_5	${}^{new}\omega$-6	-1.1264	-	-
b_6	${}^{new}\omega$-3	0.0237	-	-
b_7	$\Delta\Delta V_5(c_1)$	-0.0063	V_5 = MR	$= \Delta V_5(c_1)_{new} - \Delta V_5(c_1)_{ref} = p(c_1)_{new} \cdot ({}^{new}V_5 - \langle V_5(c_1) \rangle_{new}) - p(c_1)_{ref} \cdot ({}^{ref}V_5 - \langle V_5(c_1) \rangle_{ref})$; $\langle V_5(c_1) \rangle$ = Average of MR for c_1
b_8	$\Delta\Delta V_6(c_2)$	0.0044	V_6 = LogP	$= \Delta V_6(c_2)_{new} - \Delta V_6(c_2)_{ref} = p(c_2)_{new} \cdot ({}^{new}V_6 - \langle V_6(c_2) \rangle_{new}) - p(c_2)_{ref} \cdot ({}^{ref}V_6 - \langle V_6(c_2) \rangle_{ref})$; $\langle V_6(c_2) \rangle$ = Average of LogP for c_2
b_9	$\Delta\Delta V_1(c_4)$	-0.0037	V_1 = Mw	$= \Delta V_1(c_4)_{new} - \Delta V_1(c_4)_{ref} = p(c_2)_{new} \cdot ({}^{new}V_1 - \langle V_1(c_4) \rangle_{new}) - p(c_2)_{ref} \cdot ({}^{ref}V_1 - \langle V_1(c_4) \rangle_{ref})$; $\langle V_1(c_4) \rangle$ = Average of Mw for c_2
b_{10}	$\Delta\Delta V_1(c_5)$	-0.0036	V_1 = Mw	$= \Delta V_1(c_5)_{new} - \Delta V_1(c_5)_{ref} = p(c_5)_{new} \cdot ({}^{new}V_1 - \langle V_1(c_5) \rangle_{new}) - p(c_5)_{ref} \cdot ({}^{ref}V_1 - \langle V_1(c_5) \rangle_{ref})$; $\langle V_1(c_5) \rangle$ = Average of Mw for c_5
b_{11}	$\Delta\Delta V_4(c_6)$	-0.1682	V_4 = Aeigp	$= \Delta V_4(c_6)_{new} - \Delta V_4(c_6)_{ref} = p(c_6)_{new} \cdot ({}^{new}V_4 - \langle V_4(c_6) \rangle_{new}) - p(c_6)_{ref} \cdot ({}^{ref}V_4 - \langle V_4(c_6) \rangle_{ref})$; $\langle V_4(c_6) \rangle$ = Average of Aeigp for c_6
b_{12}	$(\Delta\Delta V_6(c_3))^2$	0.0182	V_6 = LogP	$= \Delta V_6(c_3)_{new} - \Delta V_6(c_3)_{ref} = p(c_3)_{new} \cdot [({}^{new}V_6 - \langle V_6(c_3) \rangle_{new}) - p(c_3)_{ref} \cdot ({}^{ref}V_6 - \langle V_6(c_3) \rangle_{ref})]^2$; $\langle V_6(c_3) \rangle$ = Average of LogP for c_3

[a] $f(\varepsilon ij)_{ref} = \langle \varepsilon ij \rangle_{ref} = \langle IPA(\%) \rangle_{ref}$ average of reference entries for conditions of c_1 = treatments, c_2 = with/without fermentation, c_3 = phase, c_4 = gas chromatography protocol, c_5 = replicates, and c_6 = cis/trans pattern;

[b] Symbols of molecular descriptors calculated with the DRAGON software: V_1 = Mw, V_2 = Aeigv, V_3 = Aeige, V_4 = Aeigp, V_5 = AMR, V_6 = LogP, and $V_7 = (LogP)^2$. The parameters $\Delta V_k(c_j)$ are Moving Averages, and $\Delta\Delta V_k(c_j) = p(c_j)^{new} \cdot ({}^{new}V_k - \langle V_k(c_j) \rangle^{new}) - p(c_j)ref \cdot ({}^{ref}V_k - \langle V_k(c_j) \rangle^{ref})$ are PT operators.

Where, the output function $'f(L_{nr})_{new}$ is a function of connectivity pattern (L_{nr}) in the complex network for the co-distribution of FAs in reference and new state (predicted values). The output function $'f(L_{nr})_{new}$ is useful to classify the pairs of states (pairs of nodes).The statistical parameters used were specificity (Sp), sensitivity (Sn), and accuracy (Ac). Consequently, the other input terms were expanded as follows. For instance, $\Delta\Delta V_k(c_j) = p(c_j)_{new} \cdot \Delta Vk(c_j)_{new} - p(c_j)_{ref} \cdot \Delta V_k(c_j)_{ref}$. This can be further expanded in turn as $\Delta\Delta V_k(c_j) = p(c_j)_{new}$ $({}^{new}V_k - \langle V_k(c_j) \rangle_{new})$ - $p(c_j)_{ref}$ $({}^{ref}V_k - \langle V_k(c_j) \rangle_{ref})$, among $\langle V_k(c_j) \rangle$ = Average of V_k for c_j. This new model

found predicted the effects of perturbations under the initial conditions (c_j) over FA distribution with Sensitivity, Specificity, and Accuracy > 80% for a total of = 303,712 cases in training and = 103,943 cases in external validation series (**Table 8**). These results are considered good for any LDA model.

Table 8. Results of LDA PT-LFER model for the perturbation network of fatty acid distribution in ruminal microbiome

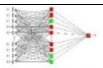
Data sub-set [a]	Statistical parameter	Prediction rates (%)	Prediction cases	
			No.(L_{nr} = 0)	***No.***(L_{nr} = 1)
Training dataset				
No.(L_{nr}= 0)	Specificity	*82.9*	127,292	26,275
No.(L_{nr} = 1)	Sensitivity	*91.1*	13,403	136,742
Train total	Accuracy	*86.9*	90.5% [b]	83.9% [c]
Validation dataset				
No.(L_{nr} = 0)	Specificity	*81.7*	40,118	8,986
No.(L_{nr} = 1)	Sensitivity	*84.9*	8,282	46,557
Validation total	Accuracy	*83.4*	82.9% [b]	83.8% [c]

[a] Number in total = 407,655; ***No.***(L_{nr}= 0) represents the number of cases in the sub-set with L_{nr} = 0 (not connected nodes) or the same with new and reference states when $IPA(\%)_{new} \leq IPA(\%)_{ref}$. ***No.***($L_{nr}$= 1) represents the number of cases in the sub-set with L_{nr} = 1 (links in the network), or the same with new and reference states when $IPA(\%)_{new} > IPA(\%)_{ref}$; [b] NPV: negative predictive value; [c] PPV: positive predictive value.

3.5. PT-NLFER model for a FA distribution network

Additional tests have been conducted with STATISTICA, using linear and non-linear ANNs (LNNs and MLPs) methods in order to compare them with the above LDA model. LNNs have one input layer and one output layer, but no hidden layer. Therefore, the predicted output is a linear combination of the input neuron values, similar with the LDA model. MLPs have at least one hidden layer of neurons. The ANNs have been used in the literature to find diverse classification models[66-68]. Accordingly, and strictly speaking, the LNN models are also PT-LFER models because they are linear relationships. However, they are included in this section because they are a particular case of ANNs. In contrast, the MLP models can be classified as PT-NLFER (PT-Non Linear Free Energy Relationships), because they consider non-linear relationships between the input PT operators and the output.

Table 9. Comparative study of PT-LFER *vs.* PT-NLFER models [a]

Model ***No.***-description	Typical topology	Statistical parameters [b]			
		TP (%)	*VP* (%)	*TE* (%)	*VE* (%)
PT-NLFER models					
1- MLP 12:12-13-1:1		87.61	87.91	59.97	85.07
2- MLP 5:5-10-1:1		82.25	88.31	56.22	47.65
3- MLP 6:6-8-1:1		87.60	90.07	49.93	51.87
4- MLP 7:7-10-1:1		88.22	90.40	46.13	40.90
5- MLP 12:12-10-1:1		92.54	92.10	43.61	53.88
6- MLP 12:12-11-1:1		92.54	93.73	41.81	41.18
PT-LFER models					
7- LNN 8:8-1:1		85.68	81.73	33.94	35.85
8- LNN 9:9-1:1		86.60	82.12	33.50	36.09
9- LNN 10:10-1:1		86.73	82.14	33.46	36.11
10- LNN 11:11-1:1		86.63	82.00	33.45	36.16
11- LNN 12:12-1:1		86.62	81.98	33.43	36.23
12- LDA 12:12-1:1		86.94	83.39	15.03	19.92

[a] PT-LFER: Perturbation Theory- Linear Free Energy Relationships; PT-NLFER: Perturbation Theory- Non Linear Free Energy Relationships.
[b] *TP* (%) = Training Performance, *VP* (%) = Validation Performance, *TE* (%) = Training Error, *VE* (%) = Validation Error.

Table 9 presents the best 11 ANN models with the corresponding statistics for the best LDA classification. The MLP models have different input variables, from 5 to 12 and the LNN models are based on 8 to 12 variables. The results demonstrate the prediction power of the non-linear ANNs (MLPs) against the linear models (LDA and LNNs). The best MLP model (no. 6: MLP 12:12-11-1:1) has 12 input variables and only one hidden layer with 11 neurons. It can predict 93.73% of the test cases and it classifies 92.54% of the training cases. This model has around 10% more prediction power compared to the LDA PT-LFER model but only 5.6% more classification power in training. The PT-NLFER model obtained with MLP number 6 classified our dataset better than the LDA PT-LFER model. However, PT-LFER is notably simpler and shows a direct relationship between the input variables and the output. If the results are sorted by the test classification (validation preference), the order of the models is the following: MLPs - 6, 5, 4, 3, 2, 1; 12 (LDA model); LNNs – 9, 8, 10, 11, 7. Thus, the LDA model has a better prediction capacity than all LNNs but less than MLPs. Between MPLs, the models with only 5 input variables can be observed (compared with the 12 ones for LDA), but the LDA model classifies 4.70% more of the training set, even if the MLP one can predict 4.92% more of the test set. Another advantage of the LDA model are the low training and validation errors compared to all ANNs (around 25% of the ANN errors). In conclusion, the MLP models were better problem solvers, but notably more complicated.

The number of nodes = fatty acids (sum of input results i^{th}) of the complex networks was 744 (the full details are presented in http://dx.doi.org/10.6084/m9.figshare.1408852). It can be concluded that the classification results, obtained with this new PT-LFER equation, are promising and confirm the potential of the present methodology. The present model is the result of combining Hansch analysis with LDA models, Box-Jenkins Operators, and Perturbation Theory ideas. Our group and other authors[69-74] have used LDA models alone or combined with Box-Jenkins Operators to predict properties of complex systems[75-78], these models may include or not perturbation theory considerations[79]. However, in this paper these ideas are extended to the Hansch analysis for the first time.

3.6. Construction of FA distribution network using the PT-LFER model

Network biology [80] is accepted as a very useful approach to shed light on the functional organization of the cell. With this idea in mind, the observed complex networks were built for perturbations in FA metabolism/distribution between ruminal media and bacterial

or protozoan individuals. In so doing it was considered that two states are connected (L_{nr} = 1) if both $f(\varepsilon_{ij})_{new}$ = IPA(%)$_{obs}$ and $f(\varepsilon_{ij})_{ref}$ = IPA(%)$_{obs}$ - IPA(%)$_{ref}$ > 0, and L_{nr} = 0 otherwise. This condition indicates that the level of both fatty acids in the new state is higher than that of fatty acids in the state of reference (initial state). Consequently, our network is a network of co-distribution of fatty acids. In addition, the model was used to predict the same complex network. To this end, it was considered that L_{nr} = 1 (nodes linked) when both values of '$f(\varepsilon_{ij})_{new}$ and '$f(\varepsilon_{ij})_{ref}$ predicted by the model have the probability $p(c_{ij})$ > 0.5 of having $f(\varepsilon_{ij})_{ref}$ = IPA(%)$_{obs}$ - IPA(%)$_{ref}$ > 0.

Last, two models of random networks (random network **1** and **2**) were also built (**Table 10**). Each model was defined with a number of nodes and links as similar as possible to the observed and predicted networks respectively. The objective was to understand the overall nature of the FA metabolism/distribution data (similarity to a random process or not). The average values of some topological indices were calculated to compare quantitatively the structure of these networks. The indices calculated were the average values of the vertex-vertex topological distance [50], node degree, and closeness of the giant component of the observed, predicted network models and the two similar random ones. Erdős–Rényi (ER) random networks where, apparently, similar to the observed and predicted networks. In fact, the average values of the topological distance, node degree, and closeness are similar, halfway between the observed and predicted network (1.83 *vs*. 1.77, 72.75 *vs*. 80.29, and 0.000755 *vs*. 0.000836, respectively).

Table 10. Giant components of the observed, predicted, and random networks

Observed networks	Value	Average indexes [a]	Value	Predicted networks
	1.8 72.7 0.0008	Distance Degree Closeness	1.8 80.3 0.0008	
ER random network [b] **1**	Value	Average indexes	Value	ER random network **2**
	1.9 68.6 0.0007	Distance Degree Closeness	1.9 80.0 0.0008	

[a] Distance means the average values of the vertex-vertex topological distance, Degree and Closeness means the node degree, and closeness of the giant component, respectively.

[b] ER random network means Erdős–Rényi random network.

4. Conclusions

Mixed experimental-theoretical methodology can be used to study the effect of multiple factors over fatty acids distribution networks on ruminal microbiome. PT and LFER ideas can be combined to develop a PT-LFER model of fatty acid distribution network. Box-Jenkins and PT Operators of physicochemical parameters are useful inputs in this sense. ANN algorithms are also useful to test the performance of alternative PT-NLFER; Non-Linear models. Last, ER random network models can be employed to carry out comparative studies with the observed and predicted networks in order to study the overall effect of perturbations on the fatty acid distribution processes.

Acknowledgements

Yong Liu acknowledges the Mexican Council for Science and Technology (CONACyT), scholarship (2014-MZO2015) and student interchange (Ref. 290842), for financial support for doctoral researches. The authors also thank CENID FyMA, INIFAP for providing experimental animals and laboratory for the analysis of the fatty acid compositions. This work was supported by the "Galician Network for Colorectal Cancer Research (REGICC)" (Ref. R2014/039), funded by the Xunta de Galicia, by "Collaborative Project on Medical Informatics (CIMED)" PI13/00280, funded by the Carlos III Health Institute from the Spanish National plan for Scientific and Technical Research and Innovation 2013−2016, and the European Regional Development Funds (FEDER). The authors acknowledge the support by the Galician Network of Drugs R+D REGID (Xunta de Galicia R2014/025).

References

1. A. P. Simopoulos, *Biomedicine & Pharmacotherapy*, 2002, **56**, 365-379.
2. A. P. Simopoulos, *Experimental Biology and Medicine*, 2008, **233**, 674-688.
3. A. P. Simopoulos, *Biomedicine & Pharmacotherapy*, 2006, **60**, 502-507.
4. A. P. Simopoulos, *The American Journal of Clinical Nutrition*, 1999, **70**, 560s-569s.
5. S. Kwak, S. Myung, Y. Lee, H. Seo and f. Korean Meta-analysis Study Group, *Archives of Internal Medicine*, 2012, **172**, 686-694.
6. G. C. Burdge and P. C. Calder, *Nutrition Research Reviews*, 2006, **19**, 26-52.
7. H. C. Bucher, P. Hengstler, C. Schindler and G. Meier, *American Journal of Medicine*, 2002, **112**, 298-304.
8. E. C. Rizos, E. E. Ntzani, E. Bika, M. S. Kostapanos and M. S. Elisaf, *The Journal of the American Medical Association*, 2012, **308**, 1024-1033.
9. S. L. Kronberg, E. J. Scholljegerdes, G. Barcelo-Coblijn and E. J. Murphy, *Lipids*, 2007, **42**, 1105-1111.
10. H. V. Petit, *Journal of dairy science*, 2002, **85**, 1482-1490.
11. T. M. Hess, J. K. Rexford, D. K. Hansen, M. Harris, N. Schauermann, T. Ross, T. E. Engle, K. G. Allen and C. M. Mulligan, *Journal of animal science*, 2012, **90**, 3023-3031.
12. K. i. Ichihara and Y. Fukubayashi, *Journal of Lipid Research*, 2010, **51**, 635-640.
13. J. G. Kramer, V. Fellner, M. R. Dugan, F. Sauer, M. Mossoba and M. Yurawecz, *Lipids*, 1997, **32**, 1219-1228.
14. M. M. Or-Rashid, N. E. Odongo and B. W. McBride, *Journal of animal science*, 2007, **85**, 1228-1234.
15. A. Varnek and I. I. Baskin, *Molecular Informatics*, 2011, **30**, 20-32.
16. C. Hansch, A. R. Steward and J. Iwasa, *Molecular Pharmacology*, 1965, **1**, 87-92.

17. K. C. Chou and Y. D. Cai, *J Chem Inf Model*, 2005, **45**, 407-413.
18. E. Estrada, E. Uriarte, E. Molina, Y. Simon-Manso and G. W. Milne, *J Chem Inf Model*, 2006, **46**, 2709-2724.
19. B. Ehresmann, M. J. de Groot and T. Clark, *J Chem Inf Model*, 2005, **45**, 1053-1060.
20. I. V. Tetko and G. I. Poda, *J Med Chem*, 2004, **47**, 5601-5604.
21. I. V. Tetko, V. Y. Tanchuk, T. N. Kasheva and A. E. Villa, *J Chem Inf Comput Sci*, 2001, **41**, 246-252.
22. K. Roy and J. T. Leonard, *J Chem Inf Model*, 2005, **45**, 1352-1368.
23. E. E. Anslyn and D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, 2006.
24. J. W. Gibbs, *A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces*, The Academy, 1871.
25. W. Greiner, D. Rischke, L. Neise and H. Stöcker, *Thermodynamics and Statistical Mechanics*, Springer New York, 2000.
26. C. Hansch, D. Hoekman, A. Leo, D. Weininger and C. D. Selassie, *Chem Rev*, 2002, **102**, 783-812.
27. C. Hansch, A. Kurup, R. Garg and H. Gao, *Chem Rev*, 2001, **101**, 619-672.
28. C. Hansch, W. E. Steinmetz, A. J. Leo, S. B. Mekapati, A. Kurup and D. Hoekman, *J Chem Inf Comput Sci*, 2003, **43**, 120-125.
29. H. Gonzalez-Diaz, S. Arrasate, A. Gomez-SanJuan, N. Sotomayor, E. Lete, L. Besada-Porto and J. M. Ruso, *Current topics in medicinal chemistry*, 2013, **13**, 1713-1741.
30. G. Calviello, H. M. Su, K. H. Weylandt, E. Fasano, S. Serini and A. Cittadini, *BioMed Research International*, 2013, **2013**.
31. P. C. Calder, C. Glaser, M. Klingler, B. Koletzko, N. D. Riediger, S. Serini, I. Innocenti, E. Piccioni and G. Calviello, in *Vitamins in the Prevention of Human Diseases*, 2011, pp. 629-696.
32. *Nutrient Requirements of Small Ruminants: Sheep, Goats, Cervids, and New World Camelids*, The National Academies Press, Washington, DC, 2007.
33. S. X. Tang, G. O. Tayo, Z. L. Tan, Z. H. Sun, L. X. Shen, C. S. Zhou, W. J. Xiao, G. P. Ren, X. F. Han and S. B. Shen, *Journal of animal science*, 2008, **86**, 1164-1172.
34. K. H. Menke and H. Steingass, *Animal Research and Development*, 1988, **28**, 7-55.
35. F. Legay-Carmier and D. Bauchart, *British Journal of Nutrition*, 1989, **61**, 725-740.
36. S. Na, *Analysis of Fatty Acids in Infant Formulas Using an Agilent J&W HP-88 Capillary GC Column*, Agilent Technologies (China) Co. Ltd., 2011.
37. J. Folch, M. Lees and G. H. Sloane Stanley, *The Journal of biological chemistry*, 1957, **226**, 497-509.
38. T. Hill and P. Lewicki, *STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*, StatSoft, Tulsa, 2006
39. H. Hong, Q. Xie, W. Ge, F. Qian, H. Fang, L. Shi, Z. Su, R. Perkins and W. Tong, *Journal of Chemical Information and Modeling*, 2008, **48**, 1337-1344.
40. H. Gonzalez-Diaz, L. G. Perez-Montoto and F. M. Ubeira, *Journal of immunology research*, 2014, **2014**, 768515.
41. A. M. Helguera, R. D. Combes, M. Pérez González and M. N. D. S. Cordeiro, *Current topics in medicinal chemistry*, 2008, **8**, 1628-1655.
42. E. Papa, F. Villa and P. Gramatica, *J Chem Inf Model*, 2005, **45**, 1256-1266.
43. A. Mauri, V. Consonni, M. Pavan and R. Todeschini, *MATCH, communications in mathematical and in computer chemistry*, 2006, **56**, 237-248.
44. A. A. Toropov and E. Benfenati, *Current drug discovery technologies*, 2007, **4**, 77-116.
45. A. Karwath and L. De Raedt, *J Chem Inf Model*, 2006, **46**, 2432-2444.
46. M. A. Siani, D. Weininger and J. M. Blaney, *J Chem Inf Comput Sci*, 1994, **34**, 588-593.
47. D. Vidal, M. Thormann and M. Pons, *J Chem Inf Model*, 2005, **45**, 386-393.
48. S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, 1998.
49. B. Junker, D. Koschutzki and F. Schreiber, *BMC Bioinformatics*, 2006, **7**, 219.
50. M. E. Newman, S. H. Strogatz and D. J. Watts, *Physical review. E, Statistical, nonlinear, and soft matter physics*, 2001, **64**, 026118.
51. J. Salimon, T. A. Omar and N. Salih, *The Scientific World Journal*, 2014, **2014**, 10.
52. P. C. Calder, *The American Journal of Clinical Nutrition*, 2006, **83**, 1505S-1519S.
53. P. C. Calder, *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 2006, **75**, 197-202.
54. P. C. Calder, *Molecular Nutrition & Food Research*, 2008, **52**, 885-897.
55. P. C. Calder, *British Journal of Clinical Pharmacology*, 2013, **75**, 645-662.
56. A. Cittadini, G. Calviello, H. M. Su and K. Weylandt, *BioMed Research International*, 2013, **2013**.
57. I. S. Nam and P. C. Garnsworthy, *Journal of Applied Microbiology*, 2007, **103**, 551-556.
58. P. F. Wilde and R. M. Dawson, *Biochemical Journal*, 1966, **98**, 469-475.
59. P. Kankaanpää, B. Yang, H. Kallio, E. Isolauri and S. Salminen, *Applied and Environmental Microbiology*, 2004, **70**, 129-136.
60. H. Goldfine, *The Journal of Lipid Research*, 1984, **25**, 1501-1507.
61. D. C. Bay and R. J. Turner, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2013, **1828**, 260-270.
62. P. A. Janmey and P. K. Kinnunen, *Trends in Cell Biology*, 2006, **16**, 538-546.
63. A. Lock and D. Bauman, *Lipids*, 2004, **39**, 1197-1206.
64. M. J. Allison, M. P. Bryant, I. Katz and M. Keeney, *Journal of Bacteriology*, 1962, **83**, 1084-1093.
65. P. Fontanille, V. Kumar, G. Christophe, R. Nouaille and C. Larroche, *Bioresource Technology*, 2012, **114**, 443-449.
66. H. González-Díaz, D. M. Herrera-Ibatá, A. Duardo-Sánchez, C. R. Munteanu, R. A. Orbegozo-Medina and A. Pazos, *Journal of Chemical Information and Modeling*, 2014, **54**, 744-755.
67. A. Duardo-Sánchez, C. R. Munteanu, P. Riera-Fernández, A. López-Díaz, A. Pazos and H. González-Díaz, *Journal of Chemical Information and Modeling*, 2013, **54**, 16-29.
68. H. González-Díaz, F. Prado-Prado, E. Sobarzo-Sánchez, M. Haddad, S. Maurel Chevalley, A. Valentin, J. Quetin-Leclercq, M. A. Dea-Ayuela, M. Teresa Gomez-Muños, C. R. Munteanu, J. José Torres-Labandeira, X. García-Mera, R. A. Tapia and F. M. Ubeira, *Journal of Theoretical Biology*, 2011, **276**, 229-249.
69. A. Duardo-Sanchez, C. R. Munteanu, P. Riera-Fernandez, A. Lopez-Diaz, A. Pazos and H. Gonzalez-Diaz, *J Chem Inf Model*, 2014, **54**, 16-29.

70. H. Gonzalez-Diaz, D. M. Herrera-Ibata, A. Duardo-Sanchez, C. R. Munteanu, R. A. Orbegozo-Medina and A. Pazos, *J Chem Inf Model*, 2014, **54**, 744-755.
71. N. Alonso, O. Caamano, F. J. Romero-Duran, F. Luan, D. S. C. MN, M. Yanez, H. Gonzalez-Diaz and X. Garcia-Mera, *ACS chemical neuroscience*, 2013, **4**, 1393-1403.
72. H. Gonzalez-Diaz and P. Riera-Fernandez, *J Chem Inf Model*, 2012, **52**, 3331-3340.
73. G. Aguero-Chapin, A. Antunes, F. M. Ubeira, K. C. Chou and H. Gonzalez-Diaz, *J Chem Inf Model*, 2008, **48**, 2265-2277.
74. G. Aguero-Chapin, H. Gonzalez-Diaz, G. de la Riva, E. Rodriguez, A. Sanchez-Rodriguez, G. Podda and R. I. Vazquez-Padron, *J Chem Inf Model*, 2008, **48**, 434-448.
75. A. Speck-Planche, V. V. Kleandrova, F. Luan and M. N. Cordeiro, *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences*, 2012, **47**, 273-279.
76. A. Speck-Planche, V. V. Kleandrova, F. Luan and M. N. Cordeiro, *Anti-cancer agents in medicinal chemistry*, 2013, **13**, 791-800.
77. A. Speck-Planche, V. V. Kleandrova and M. N. Cordeiro, *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences*, 2013, **48**, 812-818.
78. A. Speck-Planche, V. V. Kleandrova, F. Luan and M. N. Cordeiro, *Current Alzheimer research*, 2013, **10**, 117-124.
79. H. Gonzalez-Diaz, S. Arrasate, A. Gomez-San Juan, N. Sotomayor, E. Lete, L. Besada-Porto and J. M. Ruso, *Current topics in medicinal chemistry*, 2013.
80. A. L. Barabasi and Z. N. Oltvai, *Nature reviews*, 2004, **5**, 101-113.