MOL2NET, International Conference Series on Multidisciplinary Sciences
http://sciforum.net/conference/mol2net-03

Mini-review on the applications and perspectives of a new simplex machine learning approach in chemistry and biology

Muhammad FARMAN[1], Asma HAMMAMI-SEMMAR[2], Abir SARRAJ-LAABIDI[3,4], Nabil SEMMAR[3,*]

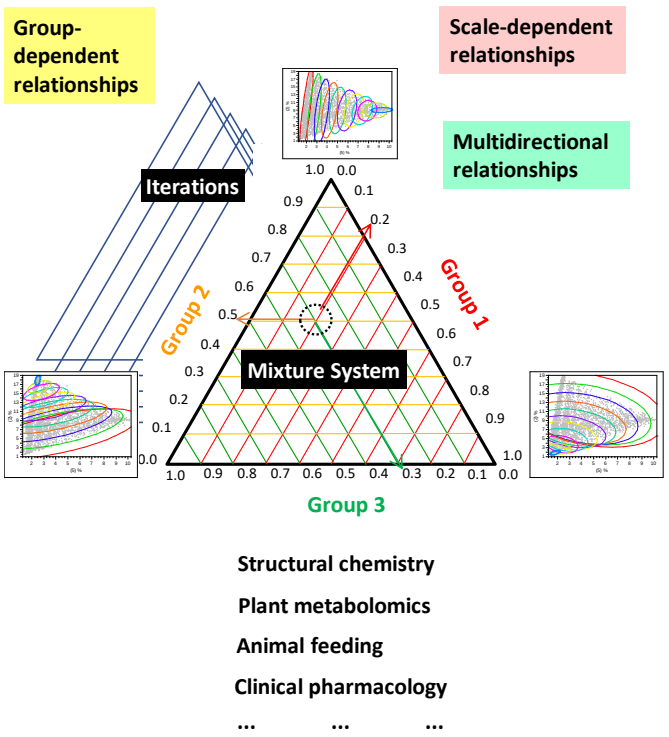
Affiliations:

1 Quaid-i-Azam University, Department of Chemistry, Islamabad, 45320, Pakistan
2 University of Carthage, National Institute of Applied Sciences and Technologies (INSAT), 1080, Tunis, Tunisia
3 Université de Tunis El Manar, Institut Pasteur de Tunis, Laboratory of Bioinformatics, Biomathematics and Biostatistics (BIMS), 1002, Tunis, Tunisia
4 Université de Tunis El Manar, Faculté des Sciences de Tunis, Campus Universitaire, 2092 Tunis, Tunisia

* **For correspondence**: nabilsemmar5@gmail.com

Keywords:
Chemical substitutions, chromatograms, feeding behaviors, hysteresis, metabolic pathways, polymorphism, stomach contents,

Graphical Abstract



Abstract.
This communication presents a two decades-updating on different applications of a new simulation method using simplex rule for smoothing regulation processes of several cooperative or competing components of complex mixture systems in biology and chemistry. Application fields covers different matter scales varying from atomes to ecosystems via molecular and biological organism levels. Simulation results issued from simplex approach helped for better understanding (i) inter-atomic regulation ways of molecular synthesis, (ii) organization metabolic pathways from chromatographic data, (iii) feeding behaviors of multi-prey foraging animal species, (iv) dynamic balances of drug derivatives in clinical populations. Application perspectives of the simplex machine-learning method are widely opened for highlighting multidirectional and multiscale regulation processes of polymorphism and diversification of multi-cluster complex systems from big data satisfying mass conservation principle.

Introduction

Simplex space is particularly conceived for analysis of mixture systems made by competing and complementary components linked by a unit sum of their relative levels [1]. Traditional applications of simplex method are well known in (i) analytical chemistry for optimization of multi-solvents-eluent compositions and (ii) pharmacology for preparation of potentially efficient multi-components drugs [2-4]. Under mathematical aspect, mixture designs and predictive equations were developed by Scheffé to control final mixture responses in relation to initial mixture compositions (factors) [5].

Beyond the search of optimal proportion of mixed components in a single complex system, a new simplex approach was developed to extract regulation law governing the differentiation of several exclusive complex systems characterized by distinct mixture patterns made by co-occurring components [6]. This computational approach provides a new dimension to mixture system analysis by considering inter-population (inter-group) aspect at higher compositional scale in addition to the classic intra-population analysis representing the basic lower scale. This approach is of machine-learning type. It was developed in 1999 by Semmar during his PhD on flavonoids of *Astragalus* genus [7, 8]. After this initiative work on plant metabolomics, application of the simulation approach for highlighting regulation processes of mass distribution variables favoring polymorphic development or system diversification was extended to different fields including structural chemistry, clinical pharmacology and animal feeding behavior [9-11].

Method

Simplex approach is fundamentally based on the mass conservation principle making a whole resource to be distributed according to several proportion or regulation ways under the conservative constraint of unit sum [**5, 6**]. Starting from a well classified population made by q groups characterized by additive variable profiles, the most hierarchical scale question of simplex approach consists in analyzing variability between and within groups to highlight regulation processes favoring the differentiation of each group. Between-group question refers to mixture system in which *q* groups compete for their developments the ones at the expense of the others under the governing constraint of unit sum linking the *q* weighted groups (**Fig. 1a**). At lower scale, within-group question refers to *p* additive variables characterizing groups' profiles with *p* relative levels representing *p* regulated parts of a whole unit sum representing the whole resource distributed within each individual profile (**Fig. 1b**).

The variation of the *q* weights of groups through a complete set of combinations between them results in a complete set of average profiles showing more or less high regulation levels of the *p* variables. Increase and decrease degrees and ways of average regulation levels of the *p* variables reflect variable processes governing the distribution of whole resources conditionally to the different groups [**2**]. Average patterns calculations from iterated combinations between weighted groups represent a machine-learning process helping to extract smoothed relationships between regulation levels of the *p* variables in each of the *q* groups. Smoothed relationships learned from iterated combinations in simplex approach are graphically highlighted at two scales: global and local scales (**Fig. 1c**). Global relationships concern general trajectories of variables leading to the development of different groups. However, local trends consist of systematic variations indicating transitory regulation processes within the governing global relationships.

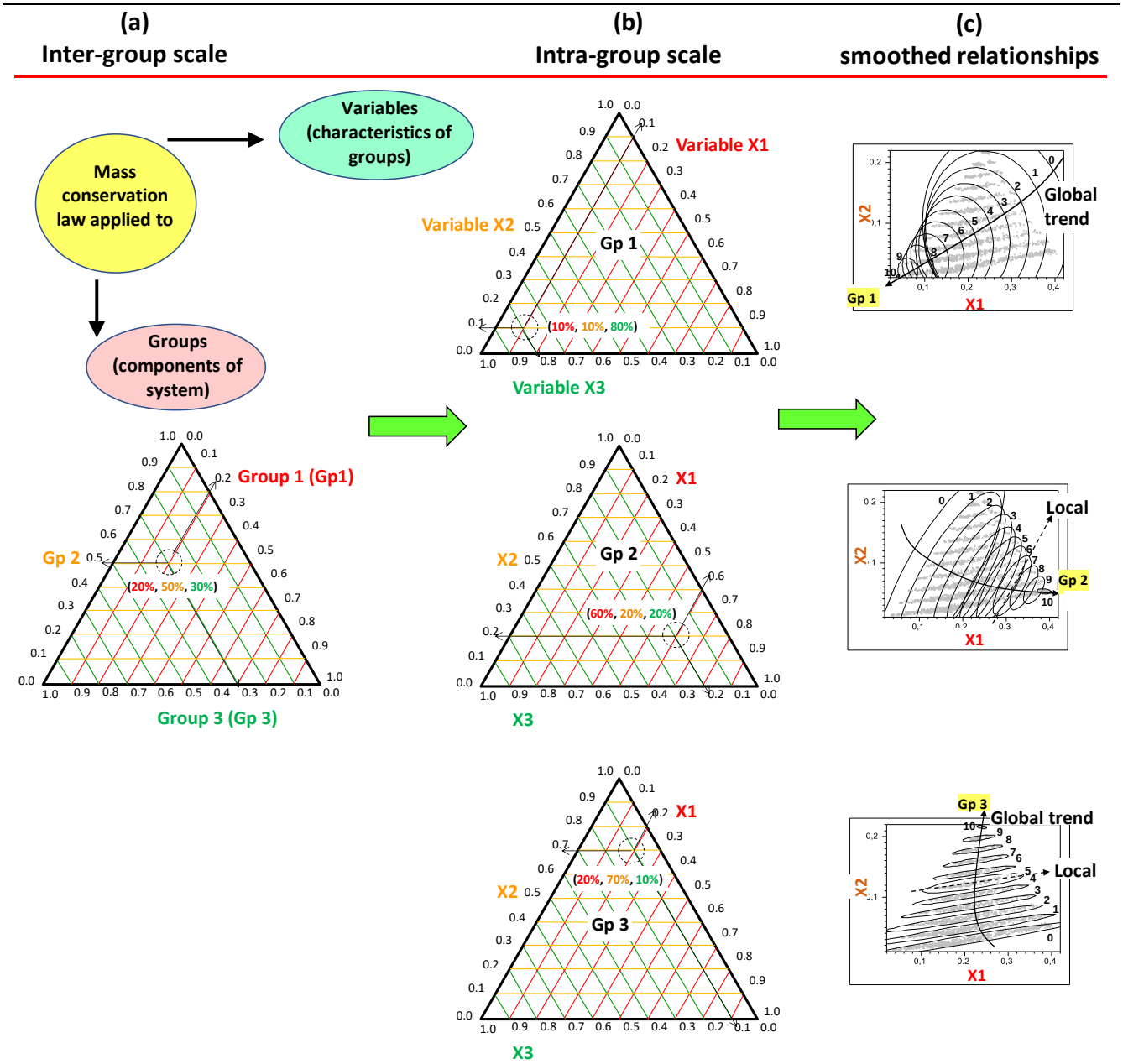


Figure 1. Scale-dependent basic methodological concepts of simplex-based simulation approach linked to mass conservation principle application at inter- and intra-group levels of studied system.

Results

Simplex approach was applied to different scale systems in biology and chemistry for regulation analysis of:

- Chemical substitutions of carbons between different metabolic classes (atomic and molecular scales) (**Fig. 2a**).
- Relative amounts of metabolites in different chromatograms representing different biological (plant) subpopulations (molecular and biological organism scales) (**Fig. 2b**).
- Feeding levels of different biological taxa consumed by a forger animal species showing different diet (biological population and ecosystem scales) (**Fig. 2c**).

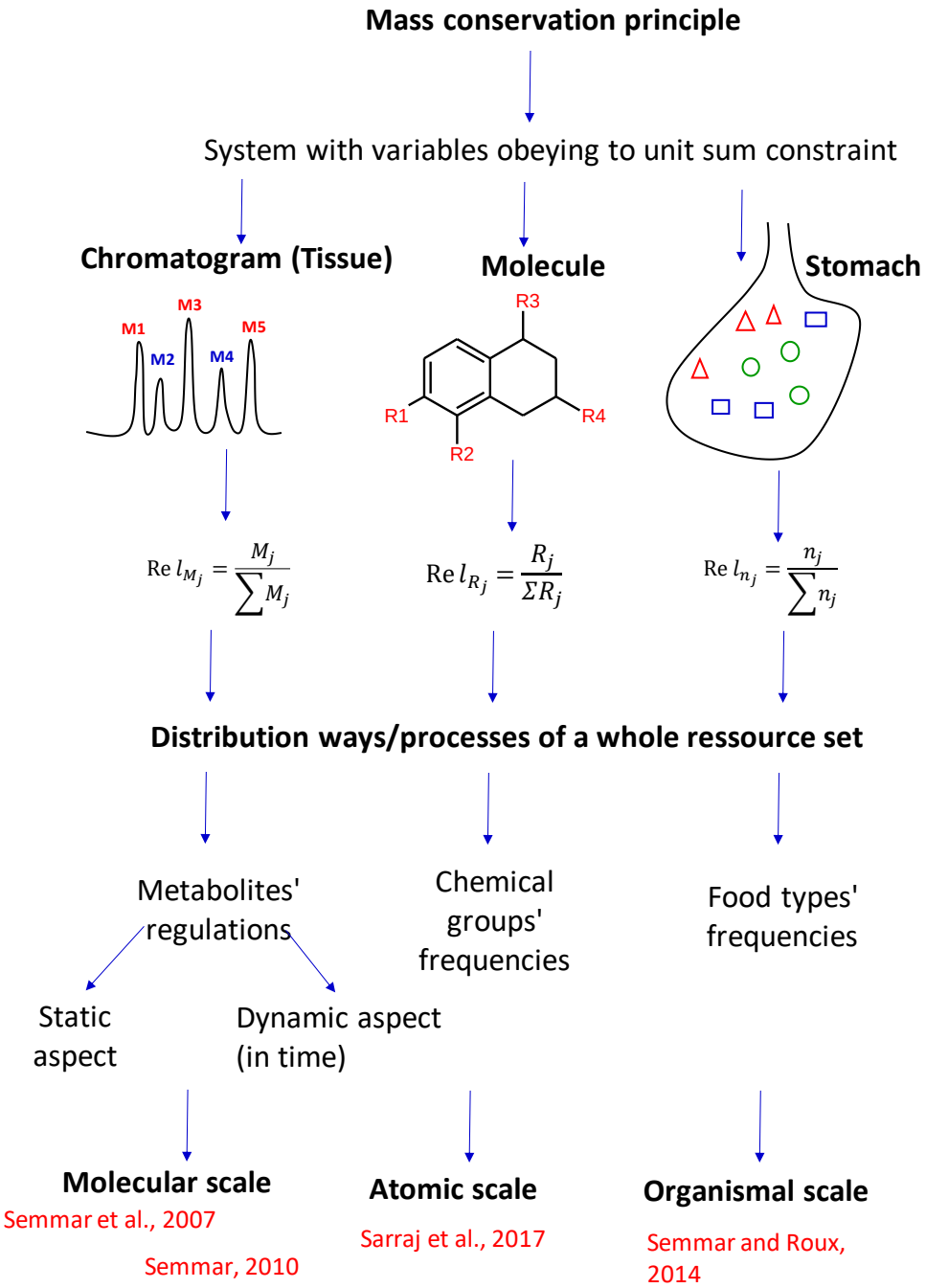


Figure 2: Different application fields of simplex approach

Simplex analysis of atomic-molecular scale systems

Simplex approach was applied to a big dataset of 156 cycloartane-based saponins synthesized by 49 *Astragalus* plant species [**11**]. Saponins were initially classified into three classes associated with increasing number of glycosylated molecular chains (desmosylation chains) Results highlighted multi-directional regulation trends between molecular carbons implying competing or sequential processes for glycosylation (*Gly*) leading to different desmosylation classes of saponins (**Fig. 3a**). Illustrative example shows conditional relationships between 6- and 3-*Gly*: for monodesmosides (one *Gly*-chain saponins), C3 tends to be initially glycosylated at the expense of C6 (C3 vs C6 competing process). Bidesmoside formation implied strong increase in 6-*Gly* regulation vs a decrease of relative levels of 3-*Gly*. Tridesmoside formation occurred by 25-*Gly* at the expense of relative level of 3-*Gly* (initially favored carbon in monodesmoside) under stable regulation of 6-*Gly*.

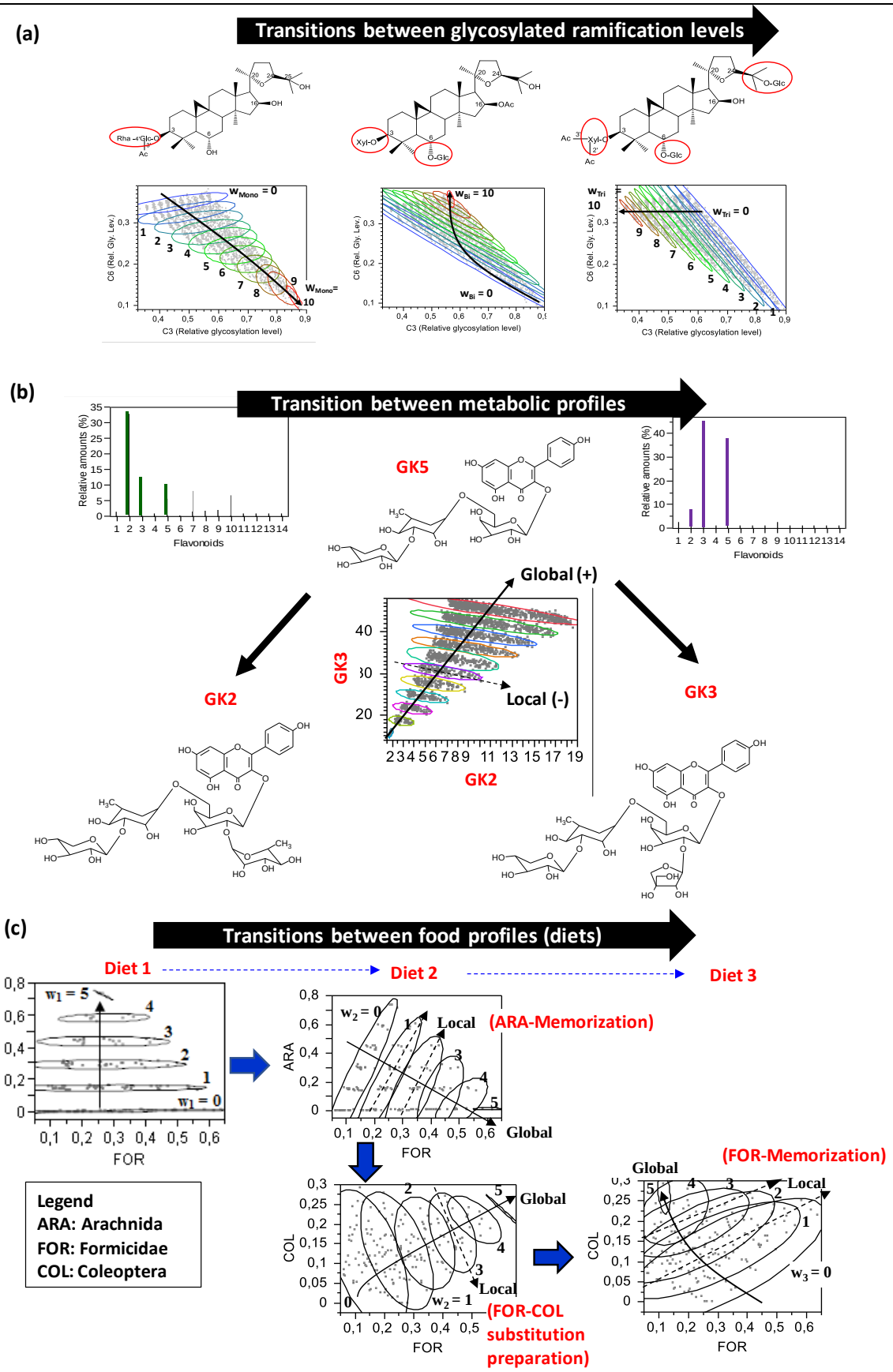


Figure 3. Simulation results given by simplex approach applied to different biological systems with different scales. (**a**) Atomic, (**b**) metabolic, (**c**) organismal scales.

Simplex analysis of metabolic-biological organism scales systems

Simplex approach was applied on a dataset of 400 chromatographic profiles of flavonoids issued from 400 individuals of the plant species *Astragalus caprinus* [**8**]. The 400 metabolic profiles were initially

classified into four chemotypes (or metabotypes) characterized by high regulation levels of some flavonoids. Simplex approach helped to highlight global and local relationships governing regulation levels of different flavonoids leading to the development of the four chemotypes. Positive and negative trends highlighted cooperative and competing processes between metabolites. Global and local scale trends (positive or negative) provided information on netwee- and within-metabolic pathway processes (**Fig. 3b**): illustrative example showed positive global trend and negative local variation between tetraglycosides of kaempferol *2* and *3* (*GK2*, *GK3*) [**8**]. Positive global trend was due to shared kaempferol glycosylation pathway of *GK 2* and *GK3* which compete together against other glycosylation pathways associated to other aglycones (e.g. quercetin pathway). Negative local variation highlighted two divergent downstream sub-pathways (*GK2* vs *GK3*) from the upstream common pathway of kaempferol. Divergence was due to the fact that tetraglycosides *GK2* and *GK3* are issued from the same precursor (triglycoside of kaempferol 5) to be glycosylated by a different sugar.

Simplex analysis of metabolic systems was also concerned with a study on drug derivatives in human Parkinson disease-suffering population [**10**]: L-Dopa and derivatives. In addition to global and local relationships between drug metabolites in different patient categories, dynamic application of simplex approach to several time-dependent data subsets helped to highlight anti-clockwise hysteresis between *DOPAC* (precursor) and *HVA* (derivative) (**Fig. 4**). Such a delayed regulation betseen two L-dopa derivatives was associated to sequential metabolic synthesis processes implying same enzymes (promiscuitary metabolic system).

Simplex analysis of population-ecosystem scales systems

Simplex approach was applied on a dataset of 45 stomach contents of lizard species (*Phrynosoma douglassi*) in order to analyze its foraging behavior toward several arthropod type preys (**Fig. 3c**) [**9**]. After classification of stomach content profiles into different diets, simplex analysis helped to extract in silico a serial of positive and negative global trends between several prey types. Such global trends provided highlighting on diet transition processes where the lizard species was revealed to be gradually motivated by the foraging of bigger and more mobile preys. Positive and negative local variation between consumed preys highlighted transition processes by which predator memorized then prepare to leave last potential preys, respectively, to learn feeding new (bigger, more mobile) preys in next diets.

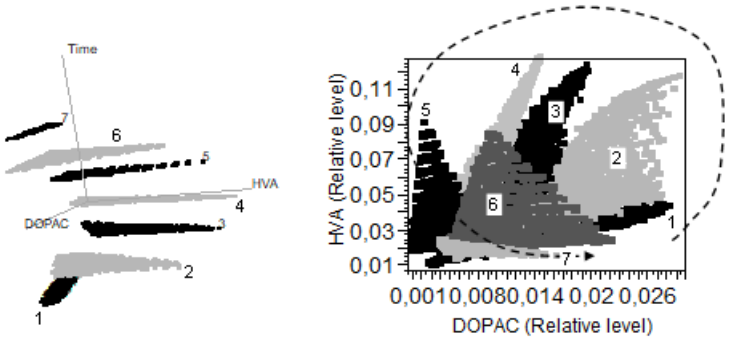


Figure 4. Hysteretic relationship highlighted by simplex approach between two metabolites (DOPAC, precursor and HVA, derivative) due to promiscuitary enzymatic system (same implied enzymes) [**10**]

Discussion

Simplex approach calls for wide promising perspective applications for studying of system control processes from big experimental datasets in biology and chemistry. Apart from the multiple fields, application breadth of simplex approach is increased by its flexibility of working a priori on different

stratification criteria in heterogeneous population making studies to benefit from inter-group aspect in addition to the intra-group one (traditionally considered in classic mixture system analysis).

References

1. Cornell, J, **2002**. *Experiments with Mixtures. Designs, Models, and the Analysis of Mixture Data*. 3rd Edition. Willey, New York, 649p.
2. Eide, I. Strategies for Toxicological Mixtures. *Food and Chemical Toxicology* **1996**, 34, 1147–1149.
3. Glajch, J. L., Kirkland, J. J., & Snyder, L. R. Practical optimization of solvent selectivity in liquid-solid chromatography using a mixture-design statistical technique. *Journal of Chromatography* **1982**, 238, 269–280.
4. Pattarino, F., Marengo, E., Gasco, M. R., & Carpignano, R. Experimental design and partial least squares in the study of complex mixtures: microemulsions as drug carriers *International Journal of Pharmaceutics* **1993**, 91, 157–165.
5. Scheffé H. Experiments with Mixtures. *J R Stat Soc Series B* **1958**, 20: 344–360.
6. Semmar N., **2013.** Two Computational Simplex Approaches to Graphical Highlighting Metabolic Phenotypes and their Functional Origins: Correspondence Analysis and Weighted Metabolic Profiles Analysis**.** In: *Metabolomics Coming of Age with its Technological Diversity. Advances in Botanical Research*, Vol. 67, 1st Edition, D. Rolin (Ed.). Academic Press, Burlington, pp. 441-492.
7. Semmar, N., **2000**. Analyse de la diversité chimique (Flavonoides et saponines) de l'espèce *Astragalus caprinus* (Fabaceae). PhD Thesis, University of Lyon 1, France.
8. Semmar N., Jay M., Nouira S., **2007**. A new approach to graphical and numerical analysis of links between plant chemotaxonomy and secondary metabolism from HPLC data smoothed by a simplex mixture design. *Chemoecology* 17, pp. 139-156.
9. Semmar N. and Roux M., **2014**. A New Simplex Approach to Highlight Multi-Scale Feeding Behaviors in Forager Species from Stomach Contents: Application to Insectivore Lizard Population. *BioSystems*, vol. 118, pp. 60-75.
10. Semmar N., **2010**. A New Mixture Design-Based Approach to Graphical Screening of Potential Interconnections and Variability Processes in Metabolic Systems. *Chemical Biology & Drug Design* 75(1), pp. 91-105.
11. Sarraj-Laabidi, A., Messai, H., Hammami-Semmar, A., & Semmar, N. **2017**. Chemometric Analysis of Inter-and Intra-Molecular Diversification Factors by a Machine Learning Simplex Approach. A Review and Research on *Astragalus* saponins. *Current topics in medicinal chemistry*, *17*(25), 2820-2848.