

## RELACIONES CUANTITATIVAS ESTRUCTURA-DULZOR BASADAS EN LOS MÉTODOS DE CLASIFICACIÓN N3 Y PLSDA

Cristian Rojas<sup>1,2,\*</sup>, Roberto Todeschini<sup>3</sup>, Davide Ballabio<sup>3</sup>, Andrea Mauri<sup>4</sup>,  
Viviana Consonni<sup>3</sup>, Piercosimo Tripaldi<sup>5</sup> y Francesca Grisoni<sup>3</sup>

<sup>1</sup> Instituto de Investigaciones Físicoquímicas Teóricas y Aplicadas (INIFTA), CONICET, UNLP, Diag. 113 y 64, C.C. 16, Sucursal 4, 1900 La Plata, Argentina

<sup>2</sup> Decanato General de Investigaciones. Universidad del Azuay, Av. 24 de Mayo 7-77 y Hernán Malo. Apartado Postal 01.01.981. Cuenca, Ecuador

<sup>3</sup> Milano Chemometrics and QSAR Research Group, Dept. of Earth and Environmental Sciences, University of Milano-Bicocca, P.za della Scienza 1, 20126 Milano, Italy

<sup>4</sup> Kode s.r.l., Via Nino Pisano, 14, 56122 Pisa, Italy

<sup>5</sup> Laboratorio de Química-Física de Alimentos. Facultad de Ciencia y Tecnología, Universidad del Azuay, Av. 24 de Mayo 7-77 y Hernán Malo. Apartado Postal 01.01.981. Cuenca, Ecuador

\* Autor para correspondencia: [crojasvilla@gmail.com](mailto:crojasvilla@gmail.com)

### Resumen.

El objetivo de este trabajo fue el desarrollo de dos modelos preliminares basados en las relaciones cuantitativas estructura-actividad (QSARs) para clasificar moléculas dulces y no dulces. Se usó una base de datos de 649 compuestos (435 dulces y 214 no dulces). Para validación externa, la base de datos se dividió de forma casual y proporcional a la numerosidad de las clases en un grupo de calibración (70%) y predicción (30%). Para cada estructura se calcularon las improntas digitales de conectividad ampliada (ECFPs) [1] y los descriptores moleculares [2] independientes de la conformación mediante el programa Dragon [3]. Durante el análisis exploratorio de los datos, usando el escalado multidimensional (MDS) [4] y las ECFPs, se identificaron tres grupos consistentes de moléculas. Dos de estos grupos están bien definidos por moléculas dulces, mientras que el tercero agrupa las moléculas dulces y no dulces. Usando las 399 moléculas del tercer grupo y los 2130 descriptores se aplicó una reducción no supervisada de descriptores con el método V-WPS [5], seguido de una reducción supervisada basada en algoritmos genéticos [6] acoplado con los métodos de clasificación de los *N*-vecinos (N3) [7] y análisis discriminante de mínimos cuadrados parciales (PLSDA) [8]. Dos modelos con 6 descriptores cada uno fueron seleccionados por optimización de la tasa de aciertos (NER) en validación cruzada. Posteriormente, la capacidad predictiva de los modelos se midió usando el conjunto de moléculas de predicción. Los resultados para el modelo N3 ( $NER_{cal} = 0.75$ ,  $NER_{cv} = 0.74$ ,  $NER_{pred} = 0.72$ ) y el modelo PLSDA ( $NER_{cal} = 0.72$ ,  $NER_{cv} = 0.71$ ,  $NER_{pred} = 0.70$ ) indican buena capacidad de discriminación entre moléculas dulces y no dulces.

[1] D. Rogers, M. Hahn, *J. Chem. Inf. Model.*, 50 (2010) 742-754.

[2] R. Todeschini, Consonni, V., *Wiley-VCH, Weinheim*, 2009.

[3] Kode, srl, Dragon (version 7), Software for Molecular Descriptor Calculation, 2016.

[4] J.B. Kruskal, *Psychometrika*, 29 (1964) 1-27.

[5] D. Ballabio, V. Consonni, A. Mauri, M. Claeys-Bruno, M. Sergent, R. Todeschini, *Chemometr. Intell. Lab.*, 136 (2014) 147-154.

[6] R. Leardi, A.L. Gonzalez, *Chemometr. Intell. Lab.*, 41 (1998) 195-207.

[7] R. Todeschini, D. Ballabio, M. Cassotti, V. Consonni, *J. Chem. Inf. Model.*, 55 (2015) 2365-2374.

[8] S. Wold, M. Sjöström, L. Eriksson, *Chemometr. Intell. Lab.*, 58 (2001) 109-130.