



UNIVERSIDAD NACIONAL DEL SUR

TESIS DE DOCTOR EN CIENCIA Y TECNOLOGÍA DE LOS MATERIALES

“PREDICCIÓN DE PROPIEDADES DE SUSTANCIAS Y MATERIALES DE INTERÉS EN LA INDUSTRIA QUÍMICA A TRAVÉS DEL DESARROLLO DE MÉTODOS COMPUTACIONALES”

Damián Palomba

BAHIA BLANCA

ARGENTINA

2014

RESUMEN

El objetivo de esta Tesis es desarrollar métodos computacionales predictivos para propiedades específicas de compuestos de interés en la industria química, particularmente en la industria farmacéutica y de materiales poliméricos. Para desarrollar la metodología de trabajo se utilizó como herramienta la técnica Relación Cuantitativa Estructura/Propiedad (QSPR) (Quantitative Structure/Property Relationship), que consiste en relacionar cuantitativamente diferentes parámetros de una entidad química (por ejemplo una molécula pequeña o un polímero) con una propiedad bien definida de la misma. Este trabajo se plantea como un estudio interdisciplinario, de forma tal que la técnica QSPR sea enriquecida con el conocimiento del ensayo de medición de las propiedades que se buscan predecir y fundamentalmente con los aspectos físico-químicos involucrados.

La metodología de trabajo se aplicó en una primera instancia a la predicción de propiedades de drogas y compuestos orgánicos en general y, en una segunda, a propiedades de materiales poliméricos. Las propiedades que se exploraron vinculadas a las drogas y compuestos orgánicos fueron algunas de las físico-químicas relacionadas al comportamiento ADMET (absorción, distribución, metabolismo, excreción y toxicidad) de los mismos. Estas fueron la *absorción intestinal humana (AIH) (Human Intestinal Absorption)* y el *pasaje de la barrera hemato-encefálica (BHE) (Blood-Brain Barrier)*, ambas esenciales para el desarrollo de nuevos fármacos. Asimismo, se estudiaron los compuestos orgánicos volátiles (VOCs) (volatile organic compounds) que son gases emitidos de ciertos sólidos o líquidos. Se predijeron sus *coeficientes de distribución sangre-hígado ($\log P_{liver}$)*, que se pueden emplear en la evaluación de riesgos y toma de decisiones en políticas de salud pública. Por otro lado, con respecto al campo de los materiales poliméricos se exploraron diferentes propiedades. Una de ellas es una propiedad térmica, la *temperatura de transición vítrea (T_g)*, la cual se relaciona con el desempeño mecánico y la procesabilidad del material; las restantes son propiedades mecánicas derivadas del ensayo de tracción en una dimensión: *elongación a la rotura (Elongation at Break)*, *resistencia a la rotura (Strength at Break)* y *módulo elástico o de Young (Tensile Modulus)*. Estas propiedades mecánicas brindan información relacionada con la ductilidad, resistencia y rigidez de un material polimérico respectivamente, y junto con otras definen su perfil de aplicación estructural.

La Tesis se organiza, de modo general, en dos grandes bloques en relación con el material al cual se aplica la predicción: drogas y compuestos orgánicos volátiles (compuestos de interés farmacéutico y de salud pública) por un lado, y por el otro, materiales poliméricos (materiales de interés en la industria química). Esta estructura obedece a las significativas diferencias moleculares entre los compuestos de trabajo de los cuales se obtiene la propiedad a predecir, denominada propiedad objetivo o target, y por lo tanto de aquí surgen también los distintos enfoques con los que se plantearon cada una de las predicciones.

La contribución original en el área de las drogas y compuestos orgánicos volátiles fue el desarrollo de nuevos modelos de predicción para las propiedades previamente mencionadas, mediante un enfoque semi-automático (un método de selección automática de variables combinado con una selección manual guiada por el conocimiento experto) que se puede aplicar también para modelar otras propiedades y otros compuestos. También el aporte del conocimiento físico-químico durante la fase de modelado conduciendo a modelos más aceptables, ya que son más fáciles de interpretar y tienden a generalizar mejor a los compuestos de diseño (virtuales), es decir compuestos aún no sintetizados.

Con relación al campo de los materiales poliméricos, las contribuciones novedosas fueron generar diferentes modelos para predecir la propiedad térmica y las propiedades mecánicas nombradas. Se desarrolló un prototipo molecular sintético, consistente en una estructura trimérica, para representar a los polímeros. Se propusieron nuevos descriptores para materiales poliméricos mediante un enfoque original de las cadenas de los polímeros, distinguiendo los fragmentos que corresponden respectivamente a la cadena principal y a la cadena lateral. Se obtuvo un modelo de predicción para la T_g enriquecido con el conocimiento físico-químico subyacente del fenómeno estudiado y se presentó una explicación estructural detallada de los descriptores del modelo y su relación con la propiedad estudiada. Luego, se validó el prototipo molecular (trímico) en relación a estructuras más complejas (31 unidades repetitivas). Con respecto a las propiedades mecánicas, se presentó un set de datos de trabajo que se recopiló y depuró para polímeros sintéticos a partir de fuentes disponibles. Se propusieron descriptores: por un lado, nuevos de cadena de polímeros, y por el otro, parámetros experimentales. Finalmente, se demostró la utilidad de incorporar información experimental del ensayo de tensión junto con estrategias

Resumen

estructurales para abordar la predicción, generando así herramientas más inteligentes e interpretables para el diseño de nuevos materiales con un perfil de aplicación específico.

ABSTRACT

The goal of this Thesis is to develop predictive computational methods for specific properties of compounds of interest in the chemical industry, particularly in pharmaceutical and polymeric materials industry. In order to develop the working method, the Quantitative Structure/Property Relationship (QSPR) technique was utilized, which relates quantitatively different parameters of an entity (e.g. a molecule or polymer) with an own well-defined process, such as a property. This work is planned as an interdisciplinary study, with the aim of improving the QSPR technique by means of physicochemical comprehension and the knowledge of target property measurement test.

Firstly, the method was applied to predict properties of drugs and general organic compounds and, secondly, to predict polymeric materials properties. Physicochemical properties related to the ADMET (absorption, distribution, metabolism, excretion and toxicity) behavior of drugs and organic compounds were explored. These were the *Human Intestinal Absorption (HIA)* and the *Blood Brain Barrier (BBB) penetration*, both essential for drug development. Furthermore, the volatile organic compounds (VOCs) were studied, which are gases emitted from certain solids or liquids. Their *blood-to-liver partition coefficients ($\log P_{liver}$)* were predicted; it can be applied to risk assessment and decision making in public health policies. Regarding to the polymeric materials field, several properties were studied. One of them is a thermal property, the glass transition temperature (T_g), which is related to the processability and material mechanical performance; the remaining ones are tensile properties: *elongation at break*, *strength at break*, and *tensile modulus*. These mechanical properties provide information related to the ductility, strength, and stiffness of a polymeric material, respectively and, along with other ones, define its structural application profile.

This Thesis can be broadly divided into two main categories, according to the material that prediction is performed: drugs and volatile organic compounds (compounds of interest in pharmaceutical industry and public health) on the one hand, and polymeric materials (materials of interest in the chemical industry) on the other. This structure is due to significant molecular differences between the working compounds (organic and polymeric materials) from which the property to predict is

Abstract

obtained (target property), and therefore to the different approaches whereby each prediction was addressed.

The original contribution in the drugs and volatile organic compounds field was the development of new predictive models for the aforementioned properties, using a semi-automatic approach (an automatic-variable-selection method combined with a knowledge-aided-manual selection) that can also be applied so as to model another properties. Moreover, during the modeling phase, the contribution of the physical-chemical knowledge led to acceptable models since they are easier to interpret and tend to better generalize design compounds (virtual), *i.e.* not-yet-synthesized compounds.

Regarding the polymeric materials science, the generation of different models for predicting the already mentioned thermal property and the mechanical properties was a novel contribution. A molecular prototype, consisting of a trimeric structure, was used in order to represent the polymers. New descriptors were proposed for polymeric materials by means of a polymer chains approach, the main and side chain. A prediction model for T_g was obtained, enriched by the underlying physicochemical knowledge from the studied phenomenon, and a detailed structural explanation of the model descriptors and its relation to the studied property was presented. Afterwards, the molecular prototype (trimer) was validated against to more complex structures (31 repeating units). With respect to tensile properties, a tailor-made dataset was presented. Several descriptors were proposed: new ones of polymer chain, and alternatively, experimental parameters. Finally, we demonstrated the usefulness of considering experimental information from the tensile test along with structural strategies to tackle the prediction, thereby more intelligent tools for the design of new materials with a specific application profile are provided.

7.5 Referencias

- Adams N. & Schubert U. S., “From Data to Knowledge: Chemical Data Management, Data Mining, and Modeling in Polymer Science”, *Journal of Combinatorial Chemistry*, **6**(1), 12-23 (2004). **262**
- Brandrup J., Immergut E.H. & Grulke E.A. (eds.), In: *Polymer Handbook*. John Wiley & Sons, New York, United States (1999). **260**
- Callister Jr. W. D. (ed.), In: *Materials Science and Engineering: An Introduction*. John Wiley & Sons, New York, United States (2007). **261, 261, 266, 290**
- Chapter 6: guidance on the principle of mechanistic interpretation, guidance document on the validation of (quantitative) structure–activity relationships [(QSAR)] models, In: *Series on Testing and Assessment, No. 69, OECD Environment Health and Safety Publications*, (2007). **289**
- Dearden J.C., Cronin M.T.D. & Kaiser K.L.E., “How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR)”, *SAR and QSAR in Environmental Research*, **20**(3-4), 241-266 (2009). **289**
- *DRAGON for Windows (Software for Molecular Descriptor Calculations)*, Version 5.5, Talete srl: Milan, Italy (2007). <http://www.talete.mi.it/> **277, 278, 278, 278, 297, 301, 301**
- Eslick J. & Camarda K., “Polyurethane design using stochastic optimization”, *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*, 769-774 (2006). **263**
- Eslick J.C., Ye Q., Park J., Topp E.M., Spencer P. & Camarda K.V., “A computational molecular design framework for crosslinked polymer networks”, *Computers and Chemical Engineering*, **33**(5), 954–963 (2009). **263**
- Gramatica P., Chapter 12: Chemometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of organic pollutants (pp. 327–366). In: *Recent Advances in QSAR Studies: Methods and Applications*. Puzin T., Leszczynski J. & Cronin M.T.D. (eds.), Springer, Dordrecht, The Netherlands (2010). **289**

- Holder A. J. & Liu Y., “A quantum mechanical quantitative structure–activity relationship study of the flexural modulus of C, H, O, N-containing polymers”, *Dental Materials*, **26**(9), 840–847 (2010). **263**
- *HyperChemTM, Molecular Modeling System*, Release 8.0.7 for Windows, Hypercube, Inc.: Gainesville, Florida, USA (2009). <http://www.hyper.com/> **273**
- Katritzky A.R., Kuanar M., Slavov S., Hall C.D., Karelson M., Kahn I. & Dobchev D.A., “Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction”, *Chemical Reviews*, **110**(10), 5714–5789 (2010). **286**
- Kutz M. (ed.), In: *Handbook of Materials Selection*. John Wiley & Sons, New York, United States (2002). **260**
- Le T., Epa V. C., Burden F. R. & Winkler D. A., “Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties”, *Chemical Reviews*, **112**(5), 2889–2919 (2012). **261, 262, 262, 273, 282**
- Matlab, <http://www.mathworks.com/> **266**
- Palomba D., Vazquez G.E. & Díaz M.F., “Novel descriptors from main and side chains of high-molecular-weight polymers applied to prediction of glass transition temperatures”, *Journal of Molecular Graphics and Modelling*, **38**, 137–147 (2012). **274, 282**
- *PolyInfo*, http://polymer.nims.go.jp/index_en.html **258, 265**
- Seitz J. T., “The Estimation of Mechanical Properties of Polymers from Molecular Structure”, *Journal of Applied Polymer Science*, **49**(8), 1331-1351 (1993). **262**
- Seymour R.B & Carraher C. E. (eds.), In: *Introducción a la química de los polímeros*. Editorial Reverté, Barcelona, Spain (1997). **244, 293, 296**
- Smith D. A. & Ulmer C. W., Impact Resistant Polymers. PCT International Patent Application 98/37118 (1998). **263**
- Soto A.J., Cecchini R.L., Vazquez G.E. & Ponzoni I., “Multi-objective feature selection in QSAR using a machine learning approach”, *QSAR & Combinatorial Science*, **28**(11-12), 1509–1523 (2009). **279**
- *STATISTICA*, Version 8.0, StatSoft, Inc.: Tulsa, Oklahoma, USA (2007). <http://www.statsoft.com/> **280**

- Tetko I.V., Bruneau P., Mewes H.W., Rohrer D.C. & Poda G.I., “Can we estimate the accuracy of ADME-Tox predictions? ”, *Drug Discovery Today*, **11**(15-16), 700–707(2006). **278**
- Todeschini R. & Consonni V. (eds.), In: *Molecular Descriptors for Chemoinformatics*. Wiley-VCH, Weinheim, Germany (2009). **278, 278, 278**
- Ulmer C.W., Smith D.A., Sumpter B.G. & Noid D.I., “Computational neural networks and the rational design of polymeric materials: the next generation polycarbonates”, *Computational and Theoretical Polymer Science*, **8**(3-4), 311–321 (1998). **262**
- Utracki L.A. (ed.), In: *Polymer Blends Handbook*. Kluwer Academic Publishers, Dordrecht, The Netherlands (2002). **260, 260**
- Van Krevelen D.W. (ed.), In: *Properties of Polymers*. Elsevier, Amsterdam, The Netherlands (2009). **248, 260, 293**
- Ward I. M. & Sweeney J. (eds.), In: *Mechanical Properties of Solid Polymers*. John Wiley & Sons, Chichester, United Kingdom (2012). **261**
- Yap C.W., “PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints”, *Journal of Computational Chemistry*, **32**(7), 1466-1474 (2011). <http://padel.nus.edu.sg/software/padeldescriptor/> **277, 293**