

PREDICCIÓN DE CURVAS TEÓRICAS DE DISTRIBUCIÓN DE PESO MOLECULAR DE RESINAS POLIMERICAS

Fiorella Cravero^{1*}, María J. Martínez², Gustavo E. Vazquez³, Ignacio Ponzoni², Mónica F. Díaz¹

¹Planta Piloto de Ingeniería Química, Universidad Nacional del Sur – CONICET.
Camino La Carrindanga km. 7, Bahía Blanca, Argentina.

²Instituto de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur – CONICET. San Andrés 800, Campus de Palihue, Bahía Blanca, Argentina.

³Facultad de Ingeniería y Tecnologías, Universidad Católica del Uruguay. Av. 8 de Octubre 2738, Montevideo, Uruguay

* fcravero@plapiqui.edu.ar

Sección 14: Química Teórica y Computacional

Introducción

El término *quimioinformática* en su sentido más amplio hace referencia al uso de herramientas computacionales, así como métodos informáticos, para el estudio de diferentes problemas en el campo de la química. Las técnicas utilizadas son conocidas como *in silico* y han sido aplicadas ampliamente en el diseño racional de drogas. En la actualidad, un área en constante crecimiento es el uso y desarrollo de estas herramientas teóricas para el diseño de nuevos materiales poliméricos.

Los polímeros son macromoléculas de elevado peso molecular (PM) formadas por la repetición de unidades denominadas monómeros. Estos monómeros reaccionan por un proceso llamado polimerización que consta básicamente de tres etapas: iniciación, propagación y finalización. Entonces, los polímeros sintéticos son moléculas muy complejas, comparados con los fármacos: tienen pesos moleculares muy altos, comportamientos singulares y además la gran mayoría de las resinas comerciales son polidispersas. Esta distribución de PM define el perfil de comportamiento mecánico del material, entre otras propiedades fisicoquímicas, de aquí su importancia. Por lo tanto, al diseñar un nuevo material polimérico, no sólo es importante describir muy bien la estructura química, sino que es indispensable considerar la polidispersión deseada.

En el ámbito experimental existe, entre otras, una técnica para determinar la distribución de peso molecular de un polímero y es conocida como GPC (cromatografía por permeación de geles), un tipo especial de cromatografía por exclusión de tamaño (SEC) que separa los analitos según su volumen hidrodinámico. Esta técnica genera una curva de distribución de pesos a partir de la cual se pueden calcular los parámetros que describen a la población de tamaños: Mn (peso molecular promedio en número), Mw (peso molecular promedio en peso), etc. A partir de estos dos se obtiene el índice de polidispersión (PDI): M_w/M_n , el cual da una idea del ancho de la curva (Fig.1). De este modo, cuando se reportan resultados en la literatura acerca de materiales poliméricos sólo se informan los valores de Mn y Mw, es decir sólo se indica su polidispersión sin presentar la curva.

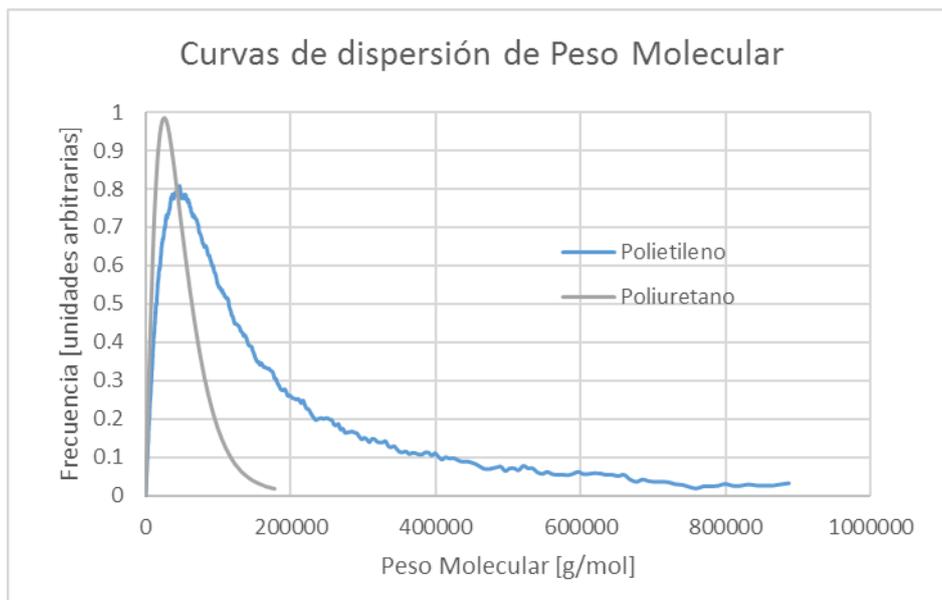


Figura 1. Distribución de PM de diferentes materiales poliméricos.

En los últimos años, nuestro grupo de investigación ha explorado el ámbito de la predicción de propiedades térmicas y mecánicas de materiales poliméricos, aplicado al campo del diseño de nuevos polímeros [1,2]. Se ha aplicado la técnica de modelado QSPR (Relación Cuantitativa Estructura-Propiedad), que relaciona una propiedad objetivo (ej. elongación a la rotura [2]) con descriptores moleculares calculados sobre la estructura molecular. Hasta el momento, se ha trabajado con representaciones moleculares sintéticas de los polímeros (monómeros, trímeros, etc.) como una primera aproximación de modelado. En la actualidad, nos encontramos con el desafío de describir al material real en términos de su polidispersión. Esto implica conocer sus curvas de distribución de peso pero, como se dijo antes, en la literatura sólo se reportan M_n y M_w . Por lo tanto, en el presente trabajo se discuten los primeros resultados de la reconstrucción de curvas teóricas de distribución de peso molecular a partir de los valores de M_n y M_w de los materiales de nuestra base de datos (Dataset 1 [2]).

Metodología y Resultados

Para atacar el problema de la reconstrucción de dispersión de pesos, se propone emplear una técnica de aprendizaje supervisado en 3 etapas (Fig.2). El objetivo del aprendizaje supervisado es crear una función capaz de predecir el valor de salida correspondiente (dato) a cualquier objeto de entrada válida (dato) después de haber aprendido con una serie de ejemplos, que son los datos de entrenamiento. Para entrenar estas herramientas se generó una nueva base de datos con curvas de distribución de PM reales (Dataset 2) para esta ocasión, cuyos materiales presentan una distribución de tipo *lognormal* (Ver apartado Dataset).

Etapa 1. A partir de cada curva del Dataset 2, se obtienen los 3 parámetros de la distribución *lognormal*: ubicación, media y desvío estándar, usando test de bondad de ajuste.

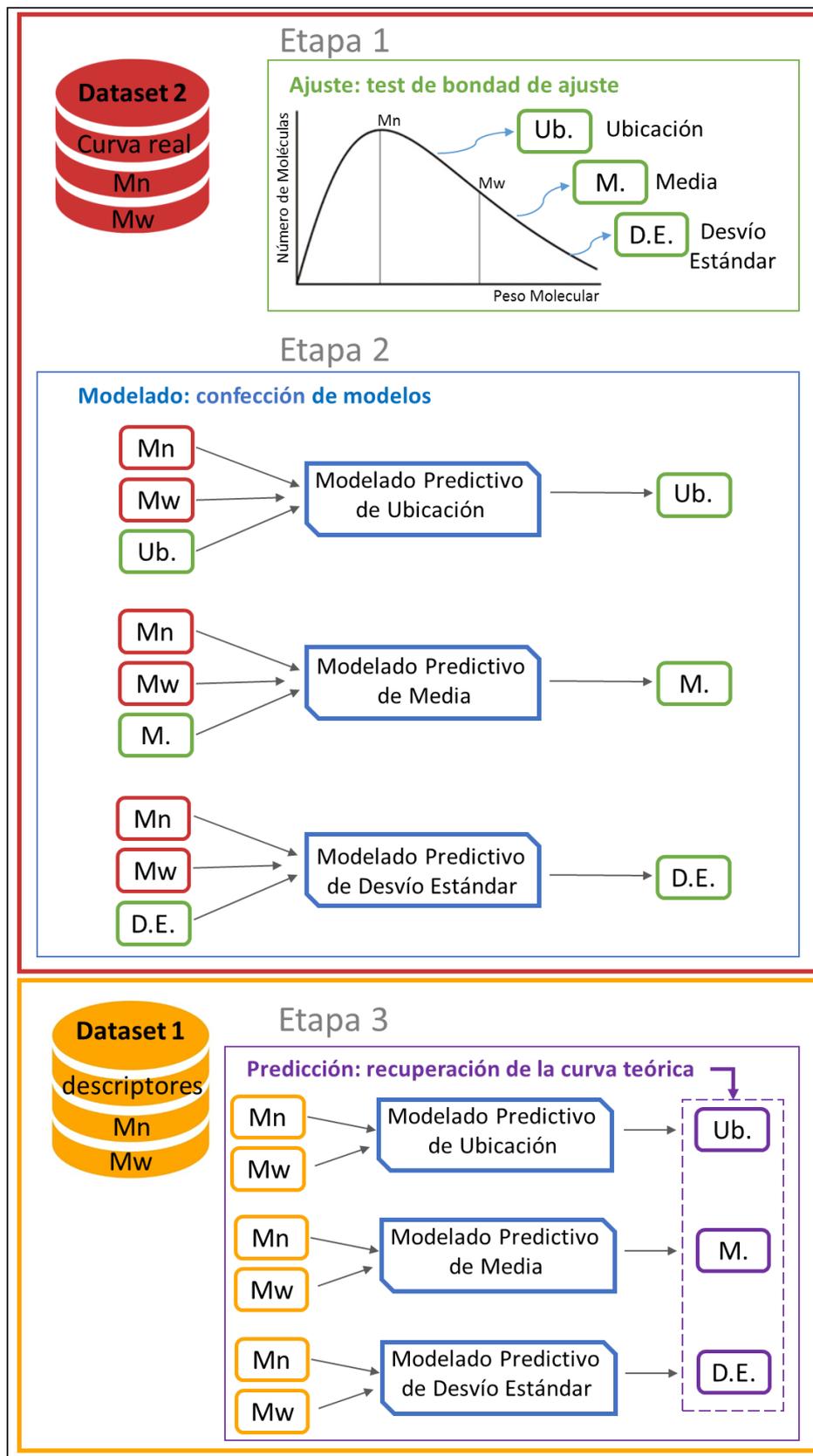


Figura 2. Metodología de predicción de curvas teóricas de distribución de pesos del Dataset 1.

Etapa 2. Se generan 3 modelos que relacionan Mn y Mw (Input) con cada parámetro *lognormal* usando la herramienta WEKA [3]. Los métodos utilizados para inferir los modelos de regresión fueron: *Linear Regression, Multilayer Perceptron, Simple Linear Perceptron, Random Committee, M5Rules, Decision Stump, Arbol M5P, Random Forest, Random Tree* (Tabla1). La calidad de predicción de cada uno de ellos fue evaluada con *leave one out* (14 *folds*) y validación cruzada de 7 *folds*. En la tabla 1 se reportan los mejores resultados. Sin embargo, no todos los valores son altos, razón por la cual se proyecta recabar mayor cantidad de datos para completar el Dataset 2 y así lograr obtener mejores desempeños.

Tabla 1. Desempeños estadísticos de los modelos de etapa 2.

		Linear Regression	M5Rules	Decision Stump	Arbol M5P	
14 folds	Ubicación	Coefficiente de Correlación	0,1322	0,1376	0,0491	0,1372
		Error Medio Absoluto	891,6529	805,9517	821,7997	796,5357
		Error Cuadrático Medio	1760,5743	1685,0439	1433,4194	1658,2004
		Error Relativo Absoluto	140,06%	126,60%	129,09%	125,12%
		Error Cuadrático Relativo	161,15%	154,24%	131,21%	151,78%
	Media	Coefficiente de Correlación	1	0,7608	0,137	0,7608
		Error Medio Absoluto	446125526,3	380689088,4	446861556	380689088,4
		Error Cuadrático Medio	866263363,4	851137639,7	963877931,9	851137639,7
		Error Relativo Absoluto	100%	85,33%	100,17%	85,33%
		Error Cuadrático Relativo	100%	98,25%	111,27%	98,25%
	Desv. Estándar	Coefficiente de Correlación	1	0,7686	0,1442	0,7686
		Error Medio Absoluto	1,19083E+16	1,01891E+16	1,19077E+16	1,01891E+16
		Error Cuadrático Medio	2,31185E+16	2,27292E+16	2,5726E+16	2,27292E+16
		Error Relativo Absoluto	100,00%	85,56%	99,99%	85,56%
		Error Cuadrático Relativo	99,99%	98,31%	111,27%	98,31%
7 folds	Ubicación	Coefficiente de Correlación	0,1653	0,132	0,6968	0,1235
		Error Medio Absoluto	9.895.146	7.793.261	5.345.359	778.734
		Error Cuadrático Medio	18.251.971	16.922.835	9.769.678	16.547.156
		Error Relativo Absoluto	1600182%	1260278%	864418%	1257929%
		Error Cuadrático Relativo	168527%	1562546%	902069%	1527858%
	Media	Coefficiente de Correlación	0,6798	0,591	0,1373	0,591
		Error Medio Absoluto	4.464.406.845.907	406.993.127.226	4.469.516.829.796	406.993.127.226
		Error Cuadrático Medio	8.689.249.965.386	8.589.250.428.841	9.640.059.068.359	8.589.250.428.841
		Error Relativo Absoluto	100%	91164%	1001145%	91164%
		Error Cuadrático Relativo	100%	988492%	1109424%	988492%
	Desv. Estándar	Coefficiente de Correlación	0,6787	0,5932	0,1442	0,5932
		Error Medio Absoluto	1,19081E+16	1,0872E+16	1,19077E+16	1,0872E+16
		Error Cuadrático Medio	2,31869E+16	2,29306E+16	2,5726E+16	2,29306E+16
		Error Relativo Absoluto	99,99%	91,29%	99,99%	91,29%
		Error Cuadrático Relativo	99,99%	98,89%	110,94%	98,89%

Etapa 3. Con los 3 modelos obtenidos en etapa 2, se cargan los datos de Mn y Mw del Dataset 1 y se predicen los 3 parámetros *lognormal* para recuperar la curva teórica de polidispersión.

Dataset.

Dataset 1: 77 polímeros de alto peso molecular, con sus respectivos Mn, Mw, descriptores calculados y datos sobre propiedades mecánicas medidas [2].

Dataset 2: 14 polímeros polidispersos (polietilenos, poliuretanos) con sus respectivos M_n , M_w y datos de la curva de distribución de pesos provenientes de un estudio de GPS.

Conclusiones

Los principales aportes de este trabajo fueron la construcción de un set de curvas de polidispersión de materiales reales (Dataset 2) y la generación de modelos predictivos para recuperar mediante aprendizaje supervisado las curvas teóricas de polidispersión de nuestro Dataset 1. En tal sentido, según nuestro conocimiento, esta es la primera metodología que intenta abordar esta problemática.

Referencias

1. Palomba D., Vazquez G., Díaz M.F. (2012). "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. Elsevier.
2. Palomba D., Vazquez G., Díaz M.F. (2014) "Prediction of Elongation at Break for Linear Polymers". *Chemometrics and Intelligent Laboratory Systems*, 139, 121–131. Elsevier.
3. Hall, M. et al. (2009). "The WEKA" Data Mining Software: An Update. *ACM SIGKDD Explorations Newsletter* 11(1), 10-18.