



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA QUÍMICA Y BIOTECNOLOGÍA

DISEÑO Y ANÁLISIS DE SIGNIFICANCIA ESTADÍSTICA DE DESCRIPTORES EN  
LA ELECCIÓN DE MOLÉCULAS A REFORMAR UTILIZANDO CAMPD

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL QUÍMICO

GUILLERMO ANDRÉS VALENZUELA VENEGAS

PROFESOR GUÍA:  
FELIPE DÍAZ ALVARADO

MIEMBROS DE LA COMISIÓN:  
FRANCISCO GRACIA CAROCA  
J. CRISTIAN SALGADO HERRERA

SANTIAGO DE CHILE  
MAYO 2014

RESUMEN DE LA MEMORIA  
PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL QUÍMICO  
POR: GUILLERMO ANDRÉS VALENZUELA VENEGAS  
FECHA: MAYO 2014  
PROF. GUÍA: FELIPE DÍAZ ALVARADO

## DISEÑO Y ANÁLISIS DE SIGNIFICANCIA ESTADÍSTICA DE DESCRIPTORES EN LA ELECCIÓN DE MOLÉCULAS A REFORMAR UTILIZANDO CAMPD

Hoy en día, para cubrir las necesidades energéticas a nivel industrial, de transporte y social, se utilizan combustibles fósiles, los que dañan el medio ambiente. Una alternativa para reemplazar esta fuente es el hidrógeno, el que puede ser producido a través del reformado de compuestos orgánicos como el metanol, metano, etanol, glicerol, pentano, hexadecano, entre otros. Pero, ya que se tiene un conjunto de compuestos reformables, cabe pensar lo siguiente ¿Son éstas las mejores moléculas para reformar? ¿Cuál es la mejor? ¿Cómo se podría determinar? Estas preguntas resultan difíciles de responder, ya que corresponden a un problema inverso y de optimización, cuya resolución es lenta y costosa. La solución que se propone, es utilizar CAMPD (*Computer Aided Molecular and Process Design*), que considera la definición de descriptores y QSPRs, para estimar el comportamiento de las propiedades macroscópicas (variables de salida que se desean optimizar) a partir de propiedades microscópicas (variables de entrada), para así facilitar el ciclo de optimización. CAMPD por su parte, está vinculado con el concepto de *piezas*, que se utilizan para generar los compuestos candidatos que cumplen las propiedades macroscópicas objetivo.

Es por eso, que el presente trabajo, se enmarca en la primera etapa de CAMPD, lo que corresponde a definir descriptores estadísticamente significativos para el desempeño de moléculas en el proceso de reformado y proponer un tamaño de *piezas* que pueda representar al conjunto de moléculas a analizar y que genere la menor cantidad estructuras infactibles fisicoquímicamente.

En cuanto al tamaño de piezas, se selecciona al *grupos de átomos*, ya que logran representar la totalidad del conjunto de compuestos y generar alrededor de 140 estructuras infactibles.

Por otro lado, se determina que los descriptores que se obtienen para las propiedades macroscópicas: rendimiento de hidrógeno, emisiones y costo de producción; son el número de carbonos ( $Z_1$ ), hidrógenos ( $Z_2$ ) y grupos  $OH$  ( $Z_3$ ). Por su parte, a través del PCA, se obtienen dos variables,  $v_1 = Z_1$  y  $v_2 = 0.42Z_2 + 0.903Z_3$  que resumen y explican el comportamiento de las tres propiedades macroscópicas. Cabe señalar, que podrían existir más descriptores (como el ángulo de enlace o la accesibilidad al carbono), pero no se opta por ellos, ya que su cálculo es muy complejo.

Por último, se debe señalar que muchos de los datos que se utilizan en el análisis cuantitativo y cualitativo, son estimados, por lo que las relaciones obtenidas, podrían no ser tan acertadas. Además, el análisis cuantitativo, solo determina si existe alguna relación lineal entre las variables, por lo que puede generar sesgo en los resultados. Sin embargo, se debe destacar el desarrollo de una estrategia para abordar este problema (o alguno similar), una vez que se posean todos los datos necesarios y, así, obtener las relaciones correctas.

*A todos los que buscan una  
respuesta en vez de una solución.*

# Agradecimientos

Nunca he sabido cómo empezar un párrafo, y mucho menos cómo empezar uno relacionado con agradecimientos, pero llegó la hora de tener que hacerlo. No es una obligación, pero siento que si es de agradecer, tendría que hacerlo a todas las personas que conocí y también a las que no conocí, ya que influyeron en lo estoy completando. Pero no puedo mencionar a todos ellos, por lo que lo haré en particular a algunos, los que siempre han estado conmigo, me han apoyado e incluso desalentado en alguna oportunidad.

Quiero partir dándoles las gracias a mis papás, ya que sin su cariño, apoyo incondicional, consejos y todo, no habría llegado acá donde estoy. Además, darle las gracias por el ejemplo que me han dado, ya que son los mejores padres que uno pueda tener (y son míos!), y ojalá cuando me llegue la oportunidad, poder hacerlo tan bien como ustedes, dedicados totalmente. Quiero agradecer, también, a mi hermano, ya que sin sus experiencias y sin los caminos que recorrió antes que yo, no habría sabido qué dirección tomar. Muchas gracias!

Quiero darle las gracias a mi profesor guía, Felipe, quien además de soportar mis horrores en redacción y ayudarme en la memoria, es una de las personas que me sirve de ejemplo, día a día, para terminar esta nueva etapa que estoy entrando y llegar a ser tan gran profesor como lo es él. Gracias.

También agradezco a los profes, Cristian Salgado y Francisco Gracia, que fueron parte de los primeros profes que inspiraron el camino que estoy siguiendo, que es el de ser investigador y académico. Gracias por sus aportes y total disposición.

La estadía en la U, no me significó solo estudio, sino que también conocer gente (mucho). Por eso, quiero agradecer a mis amigos que conocí a lo largo de este período, con los que compartí todos esos momentos de sufrimiento y diversión (?): a Wax (~~Mauricio~~), a Camilo, a Clon (~~Mauricio~~), a Juan Lusho, cuándo sus tacas?. También, no puedo olvidar agradecerles a la Alicia (~~cabra-chica-gritona~~), a Juan Pablo, a Sergio, a la Made y a Marie, uxa que se pasaba bien cuando estaba con ustedes! Les agradezco también, Mentor, Fernando y Con3jo, por todas esas salidas “intelectuales” (~~y terminar en un medonalds~~) que disfruté (espero disfrutar más). También, quiero agradecer a mi buena amiga y compañera de grupo Vale! somos chantas y convencemos a la gente! A mi amigo Bayroncillo, sé que serás un gran ingeniero, tienes todo para serlo. A todos los niños de la rama de Tenis de Mesa, compartí gran parte de mi estadía en la U con ustedes. También agradecer a Seba, gracias por darme los datos rosa y por ir a comer a las fuentes (sé que me dirás cuando lo leas). A la Clau por escucharme, y por confiar en mí. Y por último, y no menos importante, a mi mejor amiga, la Sole, aunque no nos hemos visto tanto como antes, siempre sé que contaré con tu ayuda y

apoyo incondicional!

Debo dejar este párrafo para agradecer a mis ángeles (~~si no se pueden enojar~~ xd), por transformar esos momentos que uno necesita olvidar todo, en risas y más risas, y hacerme recordar que en verdad depende de uno el pasarlo bien en cada momento. No cambien chiquillas, Pilly, Javi y Caro! y... no las conozco! (*pobre globo u.u*).

Obviamente no puedo olvidarme de mi *seudo* generación, si bien, no entré con ustedes a la u, fue lo mejor entrar a la especialidad con todos, tenerlos de compañeros y amigos, ir en una gira de estudios (parecíamos colegio) y ahora titularnos casi la mayoría juntos. Los quiero! y obviamente nos veremos más adelante.

Casi creo que estoy llegando al fin, y bueno, aún no te he mencionado, pero no sé cómo darte las gracias por aguantar mis gruñidos, mis tincadas, mis caprichos, mis colapsos, mis locuras, por aceptarme como soy y como no soy. Te amo mucho! Y sé que no puedo encontrar la forma de agradecerte todo eso, pero también sé que sabes que cada vez que te miro a tus ojitos, es un gracias por ser tú. Gracias Macarena del Pilar Tapia Paredes, gracias por aparecer en el tiempo y lugar preciso. Te adoro.

Finalmente, quiero agradecerme a mí (creo que nadie se agradece a uno mismo), por permitirme poder hacer todo lo que he hecho, y todo lo que haré, por no dejar que abandone mi camino, aunque a veces sepa que voy contra un muro, por levantarme y seguir, por apuntar más allá de donde puedo mirar. Gracias!

Y así, creo que terminé mis agradecimientos, posiblemente me faltaron personas a quienes incluir, pero solo es porque mi memoria siempre me falla.

Ah, y se me olvidaba, a las personas que no conocí, Gracias!.

# Tabla de contenido

Índice de tablas	viii
Índice de ilustraciones	x
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes generales	1
1.1.1. Contexto	1
1.1.2. Reformado	3
1.1.3. Problema inverso	5
1.1.4. Descripción general de CAMPD	6
1.2. Objetivos y alcances	8
1.2.1. Objetivos generales	8
1.2.2. Objetivos específicos	8
1.2.3. Alcances	9
<b>2. Marco Teórico</b>	<b>10</b>
2.1. CAMD	10
2.2. CAMPD	13
2.3. Metodología CAMPD	14
2.4. Estrategia para el problema general	15
<b>3. Metodología</b>	<b>16</b>
3.1. Selección de propiedades macroscópicas	16
3.2. Búsqueda de distintas moléculas a reformar	17
3.2.1. Restricción por composición	18
3.2.2. Restricción por estructura molecular	18
3.2.3. Otras restricciones e hipótesis	19
3.2.4. Propiedades	19
3.3. Diseño de descriptores	20
3.4. Base de datos	20
3.5. Elección del tamaño de <i>piezas</i>	21
3.6. Análisis de significancia estadístico	22
3.6.1. Análisis cuantitativo y cualitativo	24
3.6.2. Análisis de componentes principales (PCA)	27
<b>4. Resultados y discusiones</b>	<b>29</b>
4.1. Propiedades macroscópicas	29

4.1.1.	Cálculo de emisiones . . . . .	30
4.1.2.	Cálculo del Costo y el rendimiento de hidrógeno . . . . .	31
4.1.3.	Almacenamiento de los datos . . . . .	31
4.2.	Conjunto de moléculas reformables . . . . .	32
4.2.1.	Propiedades y datos almacenados . . . . .	34
4.3.	Candidatos a descriptores . . . . .	40
4.4.	Piezas . . . . .	43
4.4.1.	Cardinalidad del universo de moléculas representables cuando el tamaño de piezas es atómico . . . . .	44
4.4.2.	Cardinalidad del universo de moléculas representables cuando el tamaño de piezas corresponde a grupo de átomos . . . . .	46
4.4.3.	Selección de piezas . . . . .	48
4.4.4.	Datos y supuestos . . . . .	49
4.5.	Significancia estadística . . . . .	51
4.5.1.	Análisis cuantitativo y cualitativo . . . . .	51
4.5.2.	PCA . . . . .	56
<b>5.</b>	<b>Conclusiones</b>	<b>59</b>
5.1.	Recomendaciones y trabajos futuros . . . . .	61
	<b>Bibliografía</b>	<b>62</b>
	<b>Anexos</b>	<b>68</b>
<b>A.</b>	<b>Propiedades Macroscópicas</b>	<b>69</b>
A.1.	Matriz de propiedades macroscópicas . . . . .	70
<b>B.</b>	<b>Base de Datos</b>	<b>71</b>
B.1.	Base de datos . . . . .	71
B.2.	Cálculo de la reacción general y sus coeficientes estequiométricos . . . . .	77
<b>C.</b>	<b>Candidatos a Descriptores</b>	<b>79</b>
C.1.	Matriz de Descriptores . . . . .	80
<b>D.</b>	<b>Gráficos y tablas: significancia estadística</b>	<b>81</b>
D.1.	Gráficos: Análisis cualitativo . . . . .	81
D.2.	Resultados del PCA . . . . .	86
<b>E.</b>	<b>Funciones Matlab</b>	<b>87</b>
E.1.	Búsqueda en base de datos . . . . .	87
E.2.	Conteo de átomos de una molécula . . . . .	88
E.3.	Cálculo coeficientes estequiométricos . . . . .	89
E.4.	Estimación de datos . . . . .	89
E.4.1.	Estimación de entalpías de fusión y vaporización . . . . .	90
E.4.2.	Estimación de calor específico (sólido-gas) . . . . .	91
E.4.3.	Estimación de calor específico (líquido-gas) . . . . .	91
E.5.	Cálculo de la entalpía de reacción . . . . .	92

E.5.1.	Cálculo de entalpía de formación del compuesto a la temperatura de reformado . . . . .	93
E.5.2.	Cálculo de entalpía de formación del agua . . . . .	94
E.5.3.	Cálculo de entalpía de formación del hidrógeno . . . . .	94
E.5.4.	Cálculo de entalpía de formación del $CO_2$ . . . . .	95
E.6.	Cálculo de la matriz de propiedades macroscópicas . . . . .	95
E.6.1.	Normalización del costo del compuesto . . . . .	96
E.6.2.	Función auxiliar de entalpía de reacción . . . . .	96
E.6.3.	Función auxiliar de entalpía de reacción . . . . .	97
E.7.	Cálculo coeficiente de correlación de Spearman y <i>p-value</i> . . . . .	98
E.8.	Rutina para análisis cualitativo . . . . .	99
E.9.	Rutina para agrupar descriptores . . . . .	100



# Índice de tablas

1.1. Compuestos orgánicos que han sido reformados [1–6]. . . . .	4
3.1. Características de los hidrocarburos presentes en un sustituto para combustible de avión, obtenida de <i>Shi et al.</i> [7]. . . . .	18
4.1. Coeficientes estequiométricos de la ecuación 4.4 . . . . .	37
4.2. Ventajas y desventajas para los diferentes casos si se cambia supuesto . . . . .	39
4.3. Factores promedios para la estimación de datos . . . . .	40
4.4. Resumen de parámetros cuantificados para la selección de piezas. . . . .	48
4.5. Valores obtenidos para el coeficiente de correlación en el estudio de relación entre propiedades macroscópicas y candidatos a descriptores. Un valor cercano a 1 en módulo significará que existe relación lineal marcada entre las dos variables. . . . .	52
4.6. Valores obtenidos para cada test de significancia entre propiedades macroscópicas y cada uno de los candidatos a descriptores ( <i>p-value</i> ). Un valor inferior a $5 \times 10^{-2}$ indica que se puede confiar en que el valor para el coeficiente de correlación asociado es correcto. . . . .	53
4.7. Varianza acumulada para cada nueva variable del PCA realizado con respecto al 4.7a rendimiento de hidrógeno, 4.7b las emisiones y 4.7c costo de producción. . . . .	58
A.1. Matriz que contiene en sus entradas las propiedades macroscópicas para cada uno de los compuesto a analizar. En la primera columna se encuentra el número CAS, en la segunda Eficiencia de hidrógeno ( $[\frac{kmolH_2}{kmolOrg}]$ ), en la tercera emisiones ( $[\frac{kmolCO_2}{kmolOrg}]$ ) y en la cuarta costos ( $[\frac{CLP}{ml}]$ ). En las filas se representan cada uno de las moléculas del conjunto. . . . .	70
B.1. Compuestos y datos almacenados en la base de datos. . . . .	71
B.2. Compuestos y datos almacenados en la base de datos. . . . .	72
B.3. Compuestos y datos almacenados en la base de datos. . . . .	73
B.4. Compuestos y datos almacenados en la base de datos. . . . .	74
B.5. Compuestos y datos almacenados en la base de datos. . . . .	75
B.6. Compuestos y datos almacenados en la base de datos. . . . .	76
C.1. Matriz que contiene en sus entradas la información de los descriptores definidos para cada compuesto. Las filas representan las moléculas y las columnas los candidatos a descriptores. El orden de estas últimas es el siguiente: CAS, Número de carbonos, Número de hidrógenos, Número de grupos <i>OH</i> . . . . .	80

- D.1. Matriz que contiene en sus entradas los valores de los coeficientes de las componentes principales (ver Sección 3.6.2) para el grupo de descriptores relacionados con la eficiencia de hidrógeno. Las columnas representes a las nuevas variables, ordenadas de mayor a menor varianza, y las filas los candidatos a descriptores. . . . . 86
- D.2. Matriz que contiene en sus entradas los valores de los coeficientes de las componentes principales (ver Sección 3.6.2) el grupo de descriptores relacionados con las emisiones. Las columnas representes a las nuevas variables, ordenadas de mayor a menor varianza, y las filas los candidatos a descriptores. 86
- D.3. Matriz que contiene en sus entradas los valores de los coeficientes de las componentes principales (ver Sección 3.6.2) para el grupo de descriptores relacionados con el costo. Las columnas representes a las nuevas variables, ordenadas de mayor a menor varianza, y las filas los candidatos a descriptores. 86

# Índice de ilustraciones

1.1.	Estudio de la producción mundial de energía [8]. GTPE = Gigatoneladas de petróleo equivalente . . . . .	1
1.2.	Costos de producción de Hidrógeno para el año 2009 y estimado para el 2030 [9]	2
1.3.	(1.3a) esquema de resolución para problema directo; (1.3b) esquema de resolución para problema inverso, basados en <i>Eslick et al.</i> [10]. . . . .	5
1.4.	Esquema de metodología CAMPD, basado en <i>Eslick et al.</i> [10]. . . . .	7
1.5.	Diagrama del ciclo de optimización, basado en <i>Roughton et al.</i> [11]. . . . .	7
2.1.	Pasos Básicos de CAMD [12]. . . . .	12
4.1.	Forma de almacenamiento para los valores de las propiedades macroscópicas medidas. Las filas representan los compuestos y las columnas las propiedades almacenadas. . . . .	32
4.2.	Estructura de datos que almacenó la información acerca de los compuestos reformables y algunos datos de ellos. . . . .	33
4.3.	Escenarios posibles para las moléculas en la base de datos. Las flechas indican el punto de partida ( $25[^\circ C]$ ) y el punto final ( $600[^\circ C]$ ). En rojo, se representa el estado gaseoso del compuesto; en verde, el estado líquido; y en azul, el estado sólido. . . . .	38
4.4.	Tamaños de piezas propuestos. . . . .	43
4.5.	Cardinalidad de moléculas para caso de 1 carbono y 2 carbono. . . . .	44
4.6.	Cadena lineal de n carbonos, y los posibles candidatos para completar uno de sus enlaces. . . . .	45
4.7.	Piezas y cardinalidad para piezas de <i>grupo atómico</i> . . . . .	46
4.8.	Posibles moléculas que se pueden formar con 3 carbonos. . . . .	47
4.9.	Posibles moléculas que se pueden formar con n carbonos. . . . .	47
4.10.	Esquema de la matriz que almacena los valores del coeficiente de correlación y <i>p-value</i> entre descriptores y propiedades macroscópicas. El valor de la celda ( $i, j$ ), que tiene valor 0,8 (para el ejemplo), indica que el coeficiente de correlación o el <i>p-value</i> (dependiendo del caso), asociado al candidato a descriptor $i$ y propiedad macroscópica $j$ , obtiene dicho valor. . . . .	52
4.11.	Gráfico de relación entre Número de hidrógenos y Rendimiento de hidrógeno.	53
4.12.	Dependencia entre rendimiento de hidrógeno y emisiones. . . . .	54
D.1.	Gráfico de relación entre Número de carbonos y Rendimiento de hidrógeno. .	81
D.2.	Gráfico de relación entre Número de hidrógenos y Rendimiento de hidrógeno.	82
D.3.	Gráfico de relación entre Número de grupos <i>OH</i> y Rendimiento de hidrógeno.	82

D.4. Gráfico de relación entre Número de carbonos y Emisiones. . . . .	83
D.5. Gráfico de relación entre Número de hidrógenos y Emisiones. . . . .	83
D.6. Gráfico de relación entre Número de grupos <i>OH</i> y Emisiones. . . . .	84
D.7. Gráfico de relación entre Número de carbonos y Costos. . . . .	84
D.8. Gráfico de relación entre Número de hidrógenos y Costos. . . . .	85
D.9. Gráfico de relación entre Número de grupos <i>OH</i> y Costos. . . . .	85

# Capítulo 1

## Introducción

### 1.1. Antecedentes generales

#### 1.1.1. Contexto

Hoy en día, los combustibles fósiles abastecen las necesidades energéticas sociales, industriales y de transporte. De acuerdo a proyecciones que se han realizado (ver Figura 1.1), la producción de estos combustibles se mantendrá en los próximos años e incluso irá en aumento, transformándolo, así, en la mayor fuente de energía en el futuro. Sin embargo, a través de la quema de este tipo de combustibles se obtienen significativos flujos de contaminantes, razón que suma atractivo a otros combustible no fósiles. Una de las alternativas disponibles es la utilización de hidrógeno, la que, mediante el uso de la tecnología de *Fuel Cell*, provee una mayor eficiencia energética que los motores de combustión interna [13,14], y además, no genera flujos de contaminantes significativos.

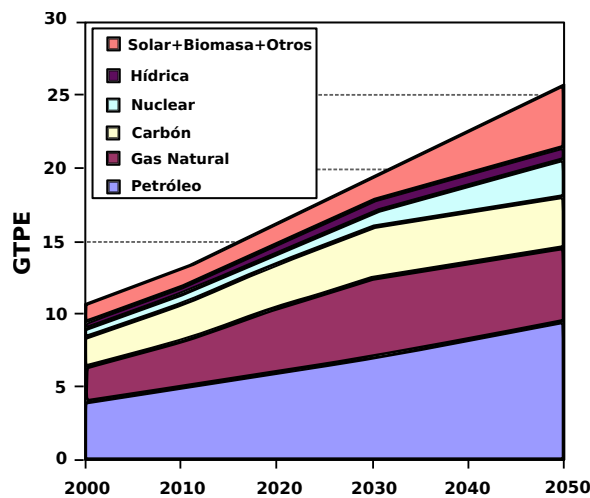


Figura 1.1: Estudio de la producción mundial de energía [8]. GTPE = Gigatoneladas de petróleo equivalente

Por otro lado, la obtención de hidrógeno no es una tarea simple a nivel industrial, ya que no se puede separar directamente de la naturaleza en su estado molecular ( $H_2$ ), sino que es necesario obtenerlo a partir de otras moléculas mediante reacciones químicas y separaciones, como por ejemplo: agua, hidratos de carbono o hidrocarburos [15].

En la Figura 1.2 se puede observar las estimaciones del costo de producción del hidrógeno a partir de distintas materias primas [9], comparando el valor correspondiente al año 2009 (situación que se considerará como actual) y la situación para el año 2030 (situación futura). Se deduce que, en general, hay más opciones atractivas en cuanto a producción centralizada que descentralizada (solo es rentable el uso de gas natural para la producción descentralizada, frente al uso de gas natural, carbón y biomasa de la centralizada), lo que se debe a la existencia de economías de escala [15]. Dentro del primer tipo (producción centralizada), una alternativa de materia prima, que compite con la utilización de gas natural (con y sin secuestro de  $CO_2$ ) y carbón para producir hidrógeno, es la utilización de biomasa y compuestos orgánicos, ya que el costo de ésta disminuirá a futuro (ver Figura 1.2). Cabe señalar que existen más alternativas para la obtención de hidrógeno, como nuclear y electrólisis, pero no son competitivas económicamente en la situación futura.

De lo anterior se desprende que una de las alternativas rentables para obtener  $H_2$  es la utilización de biomasa y compuestos orgánicos renovables. El último tipo será la materia prima seleccionada para este trabajo.

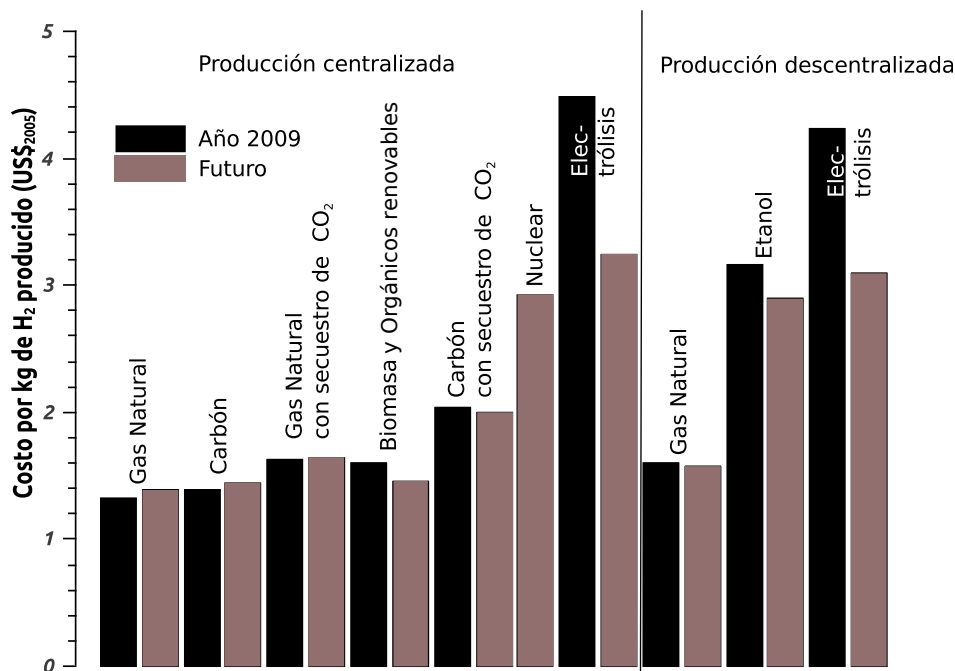


Figura 1.2: Costos de producción de Hidrógeno para el año 2009 y estimado para el 2030 [9]

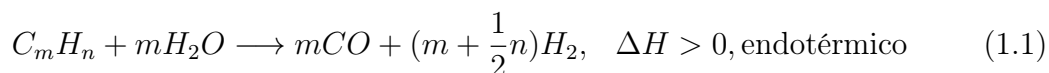
Una manera de procesar estos compuestos orgánicos renovables para su transformación a  $H_2$ , es la técnica de *reformado*, que consiste en la obtención de hidrógeno molecular a partir de moléculas como metano, gas natural, alcoholes, entre otras. Esta técnica se detallará a continuación.

### 1.1.2. Reformado

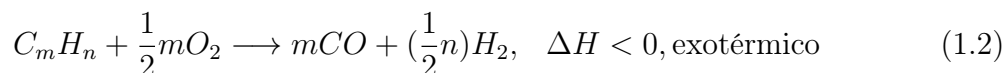
El reformado es la alternativa más utilizada para la producción industrial de hidrógeno, donde compuestos orgánicos, como metano, gas natural, alcoholes, cadenas largas de carbono (hasta  $C_{16}H_{34}$  [16]), reaccionan con agua (o aire y eventualmente, oxígeno) para así obtener el producto deseado: hidrógeno.

Existen tres alternativas para reformar compuestos orgánicos, *Steam Reforming* (SR), *Partial Oxidation* (POX), y *Autothermal Reforming* (ATR); éstas difieren principalmente en la utilización de vapor, oxígeno, o una mezcla de ambas como elemento reformante. A continuación se explicarán brevemente cada uno de estos métodos [16]<sup>1</sup>.

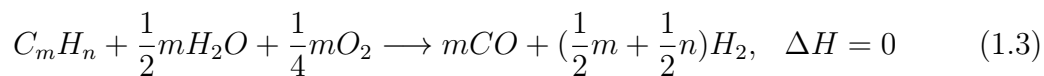
- **Steam Reforming (SR):** Tipo de reformado en que la reacción es endotérmica, ya que requiere de calor (fuente externa) y su eficiencia de hidrógeno es alta en comparación con las otras alternativas. La reacción de esta tecnología es la siguiente:



- **Partial Oxidation (POX):** En este tipo de tecnología, los hidrocarburos son transformados en gas de síntesis a través de una oxidación parcial. Este tipo de reacción es un proceso exotérmico y se alcanzan temperaturas de operación entre 1100°C a 1200°C para prevenir la formación de coque. La ecuación para esta reacción es:



- **Autothermal Reforming (ATR):** Este tipo de tecnología de reformado combina la característica exotérmica de una reacción POX y lo endotérmico del SR, es decir, utiliza el calor liberado por la reacción de oxidación parcial en el reformado por vapor. La temperatura de operación que alcanza esta tecnología es entre 900°C y 1150°C. La ecuación se muestra a continuación:



A través de las tres alternativas explicadas anteriormente, se ha encontrado una gran gama de compuestos orgánicos reformables, que van desde moléculas pequeñas como metano y metanol, hasta moléculas más grandes como hexadecano, o incluso cadenas cíclicas, como ciclooctano. En la Tabla 1.1, se pueden ver algunos compuestos que han sido reformados y existe registro científico de esto<sup>2</sup>.

---

<sup>1</sup>Las reacciones que se mencionan son solo ejemplos, pero servirán para desarrollar una reacción general (dado un compuesto) en las secciones próximas.

<sup>2</sup>Es importante destacar que las moléculas mencionadas en la Tabla 1.1, son solo una fracción de todos los compuestos.

Tabla 1.1: Compuestos orgánicos que han sido reformados [1–6].

Metanol [1]	$CH_4O$
Etanol [1]	$C_2H_6O$
n-Propanol [2]	$C_3H_8O$
Glicerol [3]	$C_3H_8O_3$
Metano [5]	$CH_4$
Etano [4]	$C_2H_6$
Iso-octano [6]	$C_8H_{18}$
n-Decano [6]	$C_{10}H_{22}$
n-Dodecano [6]	$C_{12}H_{26}$
n-Tetradecano [6]	$C_{14}H_{30}$
n-Hexadecano [6]	$C_{16}H_{34}$
Ciclooctano [6]	$C_8H_{16}$
1,2,4,5-tetrametilbenceno [6]	$C_{10}H_{14}$
m-Xileno [6]	$C_8H_{10}$
1-Metilnaftaleno [6]	$C_{11}H_{10}$

Dado que se tiene un conjunto de moléculas que se pueden reformar, entonces cabe pensar en lo siguiente: **¿Son éstas las mejores moléculas para reformar? ¿Cuál es la mejor? ¿Cómo se podría discriminar?**

Para responder las preguntas planteadas, primero es necesario definir qué se entenderá por *mejores moléculas para reformar*. Desde el punto de vista de las características de un proceso general, se busca obtener un alto rendimiento del producto deseado a un bajo costo y, a la vez, generar un bajo flujo de contaminantes. Es por eso que para el caso presente, se considerará como *mejor*, a la molécula que tenga menor costo de producción al reformar, que genere un alto rendimiento de  $H_2$ , y cuyas emisiones sean bajas, dada una cierta estructura y composición. Naturalmente, será necesario decidir qué se entenderá por emisiones, costos y rendimiento.

Por lo tanto, para encontrar solución a las interrogantes planteadas, se deberá resolver un problema de optimización, donde los criterios serán los anteriormente definidos (costo, eficiencia de hidrógeno y emisiones). Se presume que la función objetivo de la optimización será *no lineal*, y además las restricciones que se deben imponer serán *mixtas*, ya que mezclará variables discretas y continuas. Por otro lado, el problema planteado en sí entra en la categoría de *problema inverso*, ya que se estaría buscando las propiedades moleculares (microscópicas) como composición y estructura, que harán que ciertas propiedades macroscópicas tales como costo de producción, eficiencia de hidrógeno y emisiones, adquieran un valor específico. Cabe destacar que las propiedades de composición y estructura individualizan a una molécula, por lo que al fijarlas se estará determinando una molécula específica.

A continuación, se explicará en mayor detalle en qué consiste un problema inverso de predicción de propiedades macroscópicas.



### 1.1.3. Problema inverso

Primero, es necesario entender cuál es el camino a seguir para resolver problemas de ingeniería (problema directo), y posteriormente, cuál sería el camino para resolver el problema planteado (problema inverso). Específicamente, se explicará por qué ha sido categorizado como problema inverso.

Comúnmente, el camino directo se utiliza para predecir alguna propiedad macroscópica, es decir, desde propiedades microscópicas (como la composición y estructura) se estima el valor de algunas propiedades mesoscópicas o propiedades intermedias (como potenciales, energías de enlace) y se concluye con la asignación de valor para la propiedad macroscópica deseada a través de las anteriores. En cambio, el camino inverso para este mismo tipo de problema, parte definiendo el valor para la propiedad macroscópica y luego se determina la propiedad microscópica responsable de este comportamiento. Dicho en otras palabras, analizándolo como un proceso, en donde las variables de entrada son las propiedades microscópicas y las de salida son las macroscópicas, el camino directo sería determinar las variables de salida conociendo las entrada, y el inverso sería determinar las variables de entrada que provocan un conjunto de variables de salida.

En la Figura 1.3a, se observa que el camino para el problema directo de predicción de propiedades macroscópicas, es unidireccional (parte con las propiedades moleculares y termina con propiedades macroscópicas), mientras que de la Figura 1.3b, muestra que el camino inverso realiza un retroceso desde las propiedades macroscópicas, hasta las moleculares (o microscópicas). Cabe notar que la realización de este viaje tiene una dificultad explícita, ya que, a priori, no se sabe qué propiedad microscópica es la responsable de las propiedades macroscópicas deseadas, por lo que se debería realizar el viaje directo, además del inverso, para comprobar si la propiedad determinada cumple con esto.

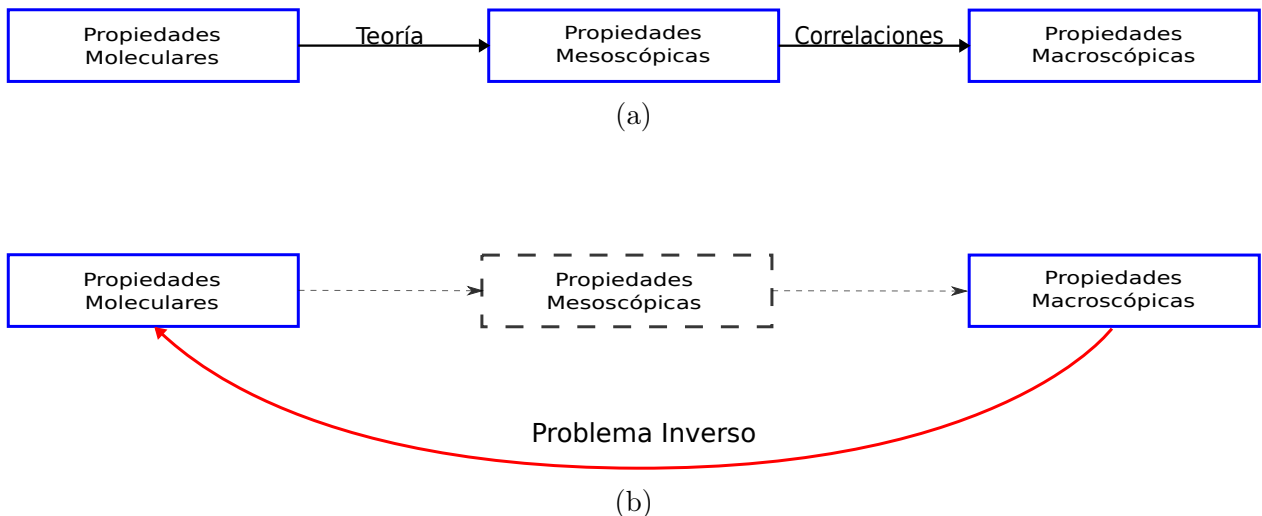


Figura 1.3: (1.3a) esquema de resolución para problema directo; (1.3b) esquema de resolución para problema inverso, basados en *Eslick et al.* [10].

Si se considera un ejemplo de estimación de costos para bombear un fluido a través de una tubería, el camino directo sería tomar como propiedad microscópica la geometría y distribución de cargas de la molécula, estimar el campo eléctrico (por ejemplo), posteriormente, obtener la viscosidad y, finalmente, obtener el valor para dicho costo relacionándolo con la viscosidad. Para el camino inverso, la ruta sería fijar el valor del costo de bombear un fluido que se desee y determinar que la propiedad microscópica responsable de este valor podría ser la geometría de la molécula y su distribución de carga. Como se mencionó antes, para el último camino no se sabe en una primera instancia que el responsable del valor del costo de separación es el tamaño de la molécula, por lo que se debería comprobar realizando el camino directo.

Volviendo al escenario planteado, el problema actual es la resolución de un ciclo de optimización, donde en cada etapa se resolvería un problema inverso de estimación de propiedades macroscópicas. Esto se debe a que en cada ciclo de optimización, se busca encontrar los valores para la composición y estructura de una molécula (propiedades que la definen), que provocan que el costo de producción, eficiencia de hidrógeno y emisiones, vayan minimizando o maximizando su valor, según el caso.

En resumen, el problema planteado es complicado de resolver, tanto por ser un problema de optimización no lineal como también por ser un problema inverso, donde encontrar en cada ciclo los valores de las propiedades microscópicas, es muy costoso y lento. Entonces, la preguntas que resultan de este análisis son: **¿existirá alguna forma de facilitar la resolución del problema de optimización?** O más específicamente, **¿se podrán encontrar las propiedades microscópicas de una forma sencilla?** **¿Existe alguna propiedad del nivel microscópica que se relacione con el rendimiento de hidrógeno?** Una solución a estas interrogantes es una metodología llamada *Computer Aided Molecular and Process Design* (CAMPD). Ésta considera la creación de descriptores que ayudarán a estimar el comportamiento de las propiedades macroscópicas a partir de propiedades microscópicas, facilitando la determinación de estas últimas en el problema inverso. Con esto, entonces, se permitiría bajar costos (recursos de cálculo o métodos de determinación de propiedades), acortar tiempo de cálculo y, finalmente, facilitar la implementación del ciclo de optimización. A continuación se realizará una explicación general de la metodología utilizada, y en el Capítulo 2 se abordará con mayor detalle.

#### 1.1.4. Descripción general de CAMPD

El principal objetivo de esta metodología, es la generación de estructuras moleculares que luego de aplicarles restricciones específicas, se obtendrían las moléculas que se ajusten a las propiedades macroscópicas que se requieren (propiedades macroscópicas objetivo). Para ello, un paso importante es la estimación de estas propiedades macroscópicas, ya que ésta debe ser rápida y precisa para facilitar los cálculos posteriores. En la Figura 1.4, se puede observar el esquema de resolución con la metodología CAMPD. En éste se ve la existencia de un elemento llamado *Descriptor*, el que ayudará a estimar comportamientos macro a través de alguna propiedad micro; y otro llamado *QSPR*, que cumple el rol de entregar el valor para las propiedades estimadas a partir de propiedades microscópicas utilizando una

función matemática que entreguen esto (finalmente éstas serán las que se optimizarán). Así, la metodología general de CAMPD es la siguiente:

- Fijar propiedades macroscópicas objetivo (criterios de optimización).
- Determinar propiedades microscópicas que se variarán para la creación de estructuras moleculares (para el problema serían largo de cadena y composición).
- Crear descriptores.
- Definir QSPRs.
- Resolver el ciclo de optimización, repitiendo hasta que se cumplan los criterios. El esquema de este ciclo se puede observar en la Figura 1.5, donde los pasos involucrados son:
  - Definir los valores de las propiedades macroscópicas deseadas.
  - Generar estructuras moleculares (utilizando *piezas*).
  - Estimar propiedades macroscópicas a través de descriptores y QSPRs.
  - Filtrar estructuras según criterios de optimización. Detener el ciclo si se cumplen con propiedades objetivo.

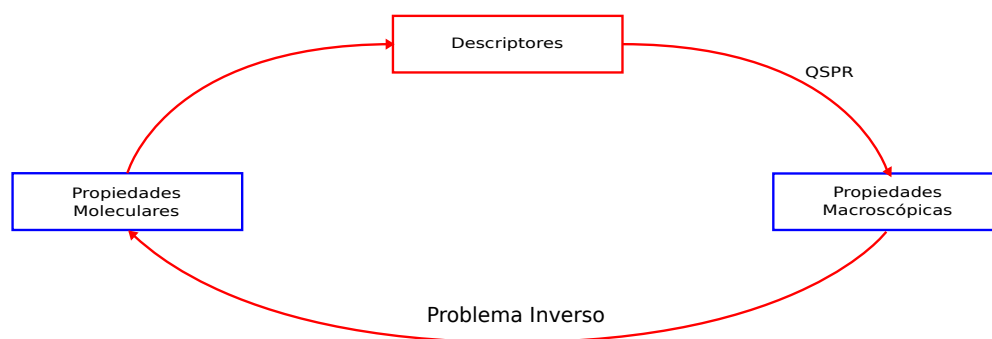


Figura 1.4: Esquema de metodología CAMPD, basado en *Eslick et al.* [10].

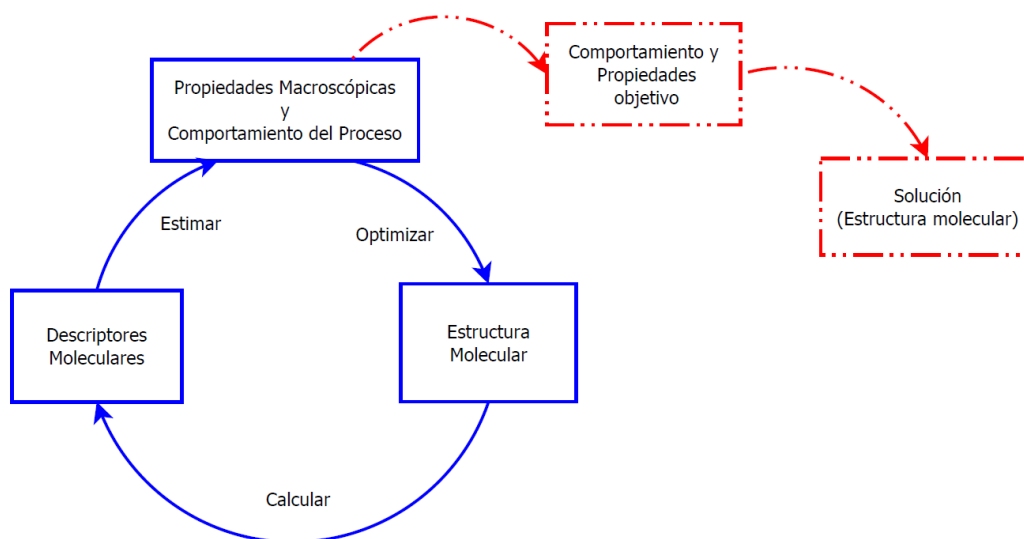


Figura 1.5: Diagrama del ciclo de optimización, basado en *Roughton et al.* [11].

Dicho en términos matemáticos, para utilizar la metodología, se debe definir:

$$\vec{y} = f(\vec{x}) \quad (1.4)$$

En la ecuación anterior,  $\vec{x}$  es el vector de propiedades micro influyentes en las propiedades macro;  $\vec{y}$  es el vector de propiedades macro a estimar (costos, impactos, disponibilidad, entre otras);  $f(\cdot)$  es una función que entrega las estimaciones de propiedades macro. Con lo cual se podría finalmente plantear el problema, en que la función objetivo a optimizar es  $f(\vec{x})$ .

## 1.2. Objetivos y alcances

A continuación se muestran los objetivos generales, específicos y limitaciones que se alcanzan en el presente trabajo.

### 1.2.1. Objetivos generales

- Definir descriptores estadísticamente significativos para el desempeño de moléculas en el proceso de reformado.
- Proponer ciertas estructuras (tamaño de piezas) que permitan representar el conjunto de moléculas a analizar de manera óptima.

### 1.2.2. Objetivos específicos

- Definir propiedades macroscópicas como rendimiento de hidrógeno, costo de producción, emisiones, entre otras, que se desean optimizar (maximizar o minimizar).
- Seleccionar un conjunto de moléculas para analizar.
- Determinar propiedades microscópicas cuantificables cuya relación con las propiedades macroscópicas sea justificable.
- Construir una base de datos con el conjunto de moléculas y las propiedades mencionadas anteriormente (propiedades microscópicas y macroscópicas).
- Definir convenientemente posibles tamaños de piezas.
- Calcular cardinalidad del universo de moléculas representables de acuerdo a los tamaños definidos.
- Elegir el mejor tamaño de piezas.
- Analizar la significancia estadística de las propiedades microscópicas con respecto a las propiedades macroscópicas definidas.

### 1.2.3. Alcances

El presente trabajo solo abarcará hasta la definición de descriptores con el análisis de significancia estadístico correspondiente, dejando la definición de QSPRs y función de optimización, y posterior resolución de éste, como trabajo futuro.

# Capítulo 2

## Marco Teórico

Como se mencionó anteriormente, la metodología general propuesta para resolver el problema se llama CAMPD, siendo ésta una modificación del método CAMD que será explicado a continuación.

### 2.1. CAMD

El diseño molecular asistido por computadora, en inglés *Computer aided molecular design*, o CAMD, busca confeccionar moléculas que cumplan con características previamente definidas. Este tipo de diseño está asociado con el concepto de *piezas*, el que se refiere a un conjunto de uno o varios elementos unitarios (por ejemplo: átomos), que al unirse entre sí en una determinada configuración, formarán una unidad macro específica (por ejemplo: Moléculas). En base a esto, CAMD se define como [12]:

*Dado un conjunto de piezas y un conjunto específico de propiedades macroscópicas, se determinará la molécula o estructuras moleculares que se ajusten a estas propiedades.*

La definición anterior hace referencia a un problema inverso de predicción de propiedades, en el que, como se muestra en la Figura 1.3b, dadas las propiedades macroscópicas objetivo y un conjunto dado de piezas, se deben determinar las moléculas (creadas a través del conjunto de *piezas*) que exhiben el comportamiento macroscópico deseado, modificando las propiedades moleculares de cada una de ellas, como por ejemplo su composición y estructura. Para ello, a partir de la metodología CAMD, se construirán diferentes posibles moléculas sobre las que se aplicarán restricciones de existencia, de diseño, u otras relacionadas con el valor de las estimaciones realizadas [12]; para finalmente obtener las estructuras moleculares que se requieren para el problema (que exhiban las propiedades macroscópicas deseadas).

La definición de las *piezas* es una tarea importante porque a partir de esto se determinarán tanto las estructuras moleculares que se puedan crear, como también el costo computacional

asociado a las restricciones de factibilidad fisicoquímicas que se impondrán sobre estas<sup>1</sup>. El grado de complejidad y el tamaño de las mismas son dos aspectos que se deben considerar, ya que, en una etapa posterior, la estimación de las propiedades macroscópicas dependerá de la contribución de las *piezas* que conforman la estructura molecular (cada pieza aportará a la estimación). Esta estimación puede ser de acuerdo a la unión por grupos moleculares (como presencia de grupos funcionales) [17], o también por la posición donde éstas se enlacen (indexación) [18].

Las etapas que se pueden identificar en el método CAMD, definidas por *Harper et al.* [17, 19], ejemplificadas a través de la elección de un solvente con ciertas características (ver Figura 2.1), son las siguientes<sup>2</sup>:

- **Pre-diseño:** El objetivo de esta etapa es identificar el problema y, a la vez, diseñar *metas*, es decir, propiedades objetivo, para así poder entregar esta información al método de solución. Es importante destacar que al definir las propiedades objetivo debe existir alguna forma de estimar estas propiedades. También en esta etapa se definen los bloques (*piezas*) que se utilizarán en la solución del problema y que darán origen a estructuras moleculares nuevas o ya existentes. En el ejemplo (ver Figura 2.1), esta etapa va desde precisar que se desea un conjunto de compuestos (alcoholes, cetonas, aldehídos y éteres) con propiedades similares al benceno, hasta la definición de las *piezas* y restricciones que se utilizarán para formar las moléculas que componen este conjunto.
- **Diseño:** En esta etapa se crean todos los posibles candidatos (compuestos) que se pueden formar a partir de las *piezas* definidas en la etapa anterior. Para evitar el crecimiento exponencial de candidatos, se utiliza una aproximación de múltiple nivel, la que consiste en que los posibles candidatos se comparan según criterios de diseño<sup>3</sup> y además se imponen filtros de existencia en una base de datos. Siguiendo con el ejemplo, aquí se forman estructuras moleculares a través de subconjuntos de las *piezas* definidas en la etapa anterior (por ejemplo el isobutil etil éter: 3 CH<sub>3</sub>, 1 CH<sub>2</sub>, 1 CH, 1 CH<sub>2</sub>O), eliminando las que no cumplan las restricciones impuestas.
- **Post-diseño:** El objetivo es analizar las soluciones obtenidas en la etapa anterior (moléculas), comparando propiedades y comportamientos que podrían no ser parte de las consideraciones de diseño. En esta etapa es necesario utilizar otras bases de datos, otros métodos de predicción de propiedades, etc., para así comprobar la validez del compuesto diseñado. Para el ejemplo, se estiman propiedades adicionales de isobutil etil éter para validar la respuesta obtenida a través del diseño.

---

<sup>1</sup>Cuando se habla de *estructura molecular*, se refiere a estructuras que no necesariamente existen en la realidad o son factibles fisicoquímicamente (por ejemplo, CH<sub>3</sub>, C, CH, entre otras).

<sup>2</sup>Estas etapas fueron definidas por los autores, pero para el presente trabajo se considerarán como generales para CAMD.

<sup>3</sup>Criterios de diseño, se refiere a ciertas propiedades que provocan que el compuesto exhiba el comportamiento deseado

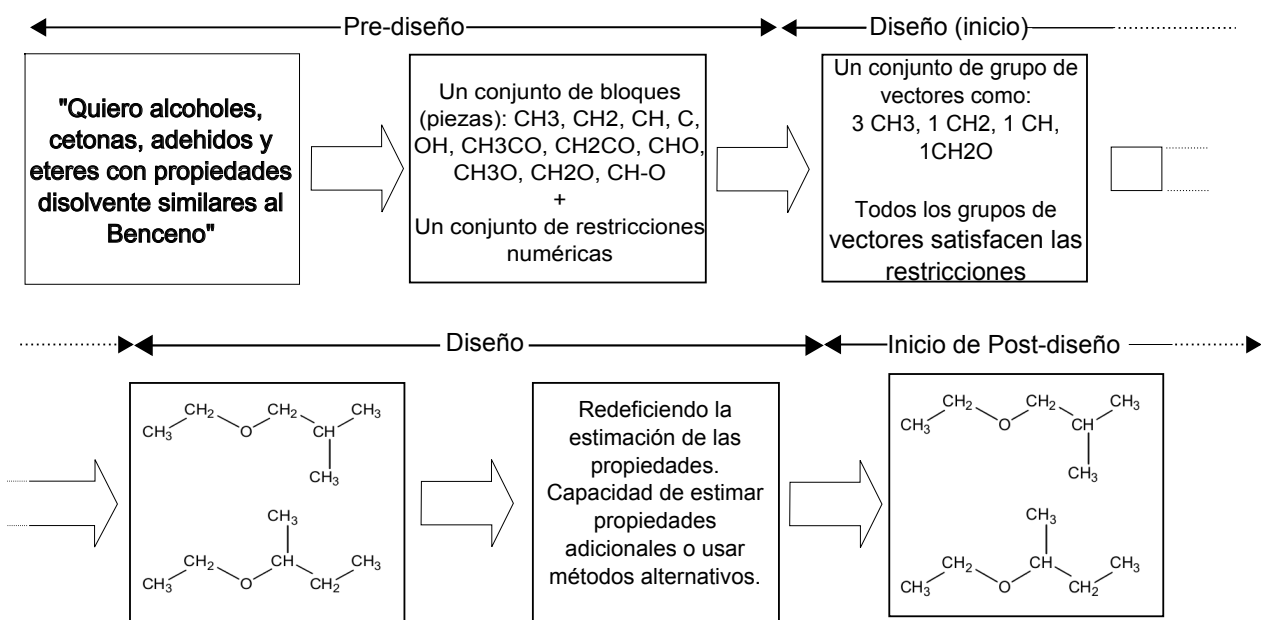


Figura 2.1: Pasos Básicos de CAMD [12].

La aplicación de esta metodología es variada, partiendo de la simple búsqueda de un solvente específico (con ciertas propiedades) en una base de datos [20], en la que no se genera ninguna estructura molecular nueva, hasta la estimación de isotermas de absorción de un absorbente de  $CO_2$  [21], donde finalmente se estiman: costos de operación e inversión de una planta y el VAN proyectado a 15 años.

Existen diversos trabajos relacionados con CAMD. Un ejemplo de esto es la búsqueda del mejor refrigerante [22] en cuanto a ciertas propiedades termodinámicas (como capacidad calorífica, calor de vaporización, interferencia con el ozono, entre otras) y que, además, pueda reemplazar a los refrigerantes que se utilizan actualmente, como es el caso del *R-12* (diclorodifluorometano) que es dañino para la capa de ozono. Estas propiedades objetivo son estimadas en base a diferentes contribuciones por parte de las *piezas* que se utilizan y a través de correlaciones. A través de esto y resolviendo un problema de optimización, se obtiene, para distintos casos de estudio, *la o las mejores sustancias refrigerantes*.

También se puede mencionar el problema de elección del mejor solvente que optimiza la eficiencia de algunas reacciones [23]. Aquí, en un primer paso, se usan diversas técnicas para determinar la constante de reacción de algunos solventes y calcular un modelo continuo de solvatación; en un segundo y tercer paso, se construye un modelo simplificado de reacción y se realiza la optimización basada en CAMD. Dentro del conjunto de resultados obtenidos, se determinó que los solventes que se conocían a priori se encontraban dentro de este conjunto y, además, habían nuevos candidatos que cumplían mejor su cometido. Estos resultados comprueban la potencia predictiva de la metodología pudiéndose determinar una posible mejor solución.

Por último, es interesante destacar el trabajo que permite elegir el mejor solvente con menor impacto ambiental, bajo ciertas propiedades objetivo [24]. Aquí se comparan los



resultados obtenidos a través del uso de CAMD con una guía de elección de solvente existente, en la que cada sustancia es jerarquizada de acuerdo a distintos puntos, tales como: seguridad, residuos, impacto, etc., determinando que los valores predichos por CAMD son muy similares a los obtenidos de la experiencia industrial.

En base a lo anterior, es posible inferir que la metodología CAMD es capaz de predecir propiedades a partir de las *piezas* (o bloques) definidas y, también, de construir nuevos posibles candidatos que logran las propiedades objetivo, llegando así a un compuesto óptimo para su uso en procesos de diferentes tipos. Pero ¿cómo se resuelve un problema CAMD? La respuesta a esta pregunta se resolverá en la Sección 2.3, pero mientras se debe tener claro que para encontrar el óptimo a través de esta metodología, se debe resolver un problema no lineal que mezcla tanto variables continuas (por ejemplo los valores de las propiedades microscópicas) como también variables discretas (por ejemplo las *piezas* definidas para la metodología), es decir, se debe resolver un problema *MINLP* o *Mixed Integer Nonlinear Programming*.

## 2.2. CAMPD

Ahora bien, CAMPD (*Computer Aided Molecular and Process Design*) usa la misma metodología de CAMD en la que se definen propiedades objetivo y luego se realiza una búsqueda de estructuras moleculares que se ajusten a estas propiedades. La gran diferencia radica en que, además de definir como variables a optimizar propiedades del compuesto, se definen variables de diseño de procesos como costos operacionales, emisiones o etapas de separación y reacción, entre otras. También, otra diferencia, es que las *piezas* definidas pueden ser, por ejemplo, operaciones unitarias, reacciones o procesos completos, es decir, ya no solo se consideran elementos moleculares, sino que también variables de procesos.

Del párrafo anterior, se puede definir CAMPD (a partir de la definición de CAMD [12]) como: “Dado un conjunto de piezas o bloques, y un conjunto específico de propiedades y comportamientos de proceso, se determinará la molécula o estructuras que se ajusten a estas características”.

El gran desafío que tiene este método de resolución es la integración del diseño molecular (elección de la mejor molécula que logre las propiedades objetivo) con la optimización del proceso. Algunas soluciones a este problema están relacionadas con la creación de un método multi-paso y multi-objetivo [25], que está basado en la formulación de un problema MINLP donde existe una etapa de pre-procesamiento del problema y un paso donde se resuelve el problema en sí. También se ha usado *simulated annealing* para integrar el diseño de solventes con el de procesos [26].

Una herramienta complementaria a la utilización de CAMPD, es SAFT-VR<sup>4</sup> [21]. Esta metodología se utiliza para la elección del mejor solvente en la separación de  $CO_2$  y metano en una corriente de gas natural en base a ecuaciones de estado de origen molecular. La

---

<sup>4</sup>Método para estimar ecuaciones de estado.

optimización de este problema se realiza en base al índice de evaluación económica, VAN, considerando las siguientes propiedades: tamaño de equipo, condiciones de operación y largo promedio de cadenas del solvente. Finalmente, se obtiene un solvente con un alto grado de pureza y un proceso en que el VAN es óptimo.

CAMPD también se ha utilizado para la elección de las etapas de separación y reacción en la producción de biodiesel y alcoholes grasos, en base a las operaciones que se realizan [27]. La metodología desarrollada se basa en la descomposición del problema de forma jerárquica, para así facilitar el filtrado de alternativas. Las piezas (bloques) que se consideran están relacionadas con las operaciones de separación y reacción, eligiendo finalmente el mejor *flowsheet* en el que se obtienen dos productos (biodiesel y alcoholes) a partir de glicerol.

Otra herramienta complementaria a CAMPD y, también, CAMD es la utilización de QSAR [28] (*Quantitative Structure-Activity Relationship*), la que se utiliza para medir la actividad de diferentes inhibidores y enzimas para producir un tratamiento paliativo del Alzheimer (diseño de drogas). A través de estos resultados, se concluye que esta herramienta es útil para la predicción de actividad biológica y además para generar posibles candidatos en el diseño de drogas.

Se han realizado otros trabajos relacionados con el diseño del mejor solvente para separaciones en *batch* basándose en aspectos económicos [29]. El problema planteado se clasifica como MIDO (*mixed integer dynamic optimization*<sup>5</sup>), en el que se pueden distinguir diversos pasos que ayudan a disminuir los posibles candidatos. A través de esta herramienta se logra desarrollar una base sólida para la elección de solvente, tomando en cuenta el aspecto económico. También puede ser extendida a otros aspectos como el medio ambiente, minimizando el impacto que pueda tener la utilización de ciertas sustancias (solventes) [25].

Finalmente, de lo anterior se deduce que la metodología CAMPD sirve como una herramienta para la selección de compuestos que cumplen una serie de requisitos (propiedades objetivo), ya que es capaz de predecir este comportamiento en base a piezas o bloques que conforman los posibles candidatos que se ajustan a estos requisitos. Además, cabe mencionar que la técnica que se ocuparía a futuro para estimar las propiedades objetivo, estará relacionada con QSPR (*Quantitative Structure-Property Relationship*), la que será explicado en la siguiente Sección. Esta técnica está presente de forma transversal en la mayoría de los trabajos anteriores y es muy similar a QSAR.

## 2.3. Metodología CAMPD

Como se mencionó anteriormente, CAMPD es una metodología para seleccionar compuestos que cumplen ciertas propiedades objetivo a partir del uso de descriptores y QSPRs. Estos últimos, son funciones matemáticas ( $f(\cdot)$ ) que utilizan como variables de entrada ( $\vec{x}$ ) las propiedades moleculares de cada compuesto (como lo su composición y estructura) y cumplen el objetivo de predecir los valores de las propiedades macroscópicas

---

<sup>5</sup>Problema de optimización dinámico en que se utilizan variables enteras.

( $\vec{y}$ ). Por otro lado, los descriptores son propiedades microscópicas que son responsables del comportamiento macroscópico de las moléculas.

No existe una metodología explícita para la resolución del problema, pero se puede desprender de los trabajos anteriormente mencionados en la Sección 2.1 y 2.2 que los pasos generales a seguir son los mencionados a continuación:

1. Identificación del espacio de diseño, es decir, definir las propiedades macroscópicas, el conjunto de elementos sobre el que se va a buscar el óptimo, un tamaño de *piezas*, los descriptores, las restricciones y los QSPRs.
2. Formulación del problema genérico de CAMPD a través del uso de los elementos definidos en la etapa anterior, es decir, plantear un problema de optimización con las variables continuas y discretas del espacio de diseño.
3. Resolución del problema de optimización a través de algún método para problemas *MINLP* (o el tipo de problema resultante).
4. Análisis de los resultados obtenidos.

## 2.4. Estrategia para el problema general

De lo anterior, contextualizado al problema a desarrollar, se propone que la metodología completa para poder responder las preguntas que se plantearon anteriormente acerca de la mejor molécula a reformar, es:

- Seleccionar las propiedades macroscópicas objetivo que se desean optimizar
- Definir descriptores estadísticamente significativos para las propiedades macroscópicas seleccionadas.
- Crear QSPRs que sean capaces de predecir de forma precisa el valor de las propiedades objetivo.
- Plantear el problema de optimización a través de las variables definidas para el problema (QSPRs, descriptores, propiedades moleculares, etc.).
- Resolver el problema de optimización para así encontrar un óptimo en cuanto a las propiedades definidas.

# Capítulo 3

## Metodología

La metodología para el presente trabajo solo abarcará los dos primeros puntos de los listados en el Capítulo 2, Sección 2.4, es decir, desde seleccionar las propiedades macroscópicas a optimizar, hasta definir descriptores estadísticamente significativos para estas propiedades.

En base a lo anterior se planificaron las siguientes tareas:

- Selección de las propiedades macroscópicas a analizar.
- Definición del conjunto de moléculas que serán analizadas.
- Diseño de candidatos a descriptores.
- Elección del tamaño de *piezas* que compondrán las moléculas en análisis.
- Evaluación de la significancia estadística de los candidatos a descriptores.

Cabe mencionar que el presente trabajo es un piloto para la creación de descriptores en el contexto de obtener la mejor molécula reformable. Futuros trabajos, con esta misma estrategia, podrían redefinir los descriptores que se propondrán, usando otras propiedades microscópicas de un conjunto mayor de moléculas.

A continuación se detallarán las actividades en el mismo orden descrito.

### 3.1. Selección de propiedades macroscópicas

La selección de las propiedades macroscópicas es una tarea sumamente importante, ya que son las propiedades objetivo que, en una etapa posterior, serán tomadas como criterios de optimización para la búsqueda de *la mejor molécula a reformar* y, además, serán punto de partida para la metodología CAMPD.

Como se mencionó en las secciones anteriores, los criterios que se utilizarán para la selección de las propiedades macroscópicas buscarán las siguientes características del proceso:

- Bajo costo.
- Alto rendimiento en el proceso.
- Bajas emisiones.

Los criterios antes mencionados están relacionados con lo que se espera de un proceso, ya que engloban aspectos económicos y medio ambientales. Si bien se podrían seleccionar más criterios (como elementos sociales), se decidió realizar este piloto de selección de descriptores con estos tres criterios, para luego, en un trabajo futuro, agregar más.

Por otro lado, como se debe analizar la significancia estadística de los descriptores con respecto a las propiedades macroscópicas, se deberán obtener valores preliminares para cada una de estas propiedades, por lo que se estimarán en base a algún elemento representativo.

Para almacenar estos datos, se utilizará una matriz de  $n$  filas y  $m$  columnas, donde las filas representarán los compuestos y las columnas estarán asociadas a las propiedades macroscópicas. Cabe señalar que la primera columna corresponderá a un número de identificación para cada compuesto, dejando las  $m-1$  columnas restantes para las propiedades que se seleccionen.

## 3.2. Búsqueda de distintas moléculas a reformar

En general, las moléculas reformables tienen en su estructura átomos de hidrógeno y producen, mediante las reacciones de  $SR$ , hidrógeno molecular ( $H_2$ ) que se puede utilizar como sustituto a los combustibles fósiles. Por consiguiente, al reformar se utilizan compuestos orgánicos, donde los átomos que están en mayor presencia son el hidrógeno y el carbono.

Por lo dicho en el párrafo anterior, y agregando la necesidad de tener un conjunto de candidatos a ser *la mejor molécula a reformar*, se requiere crear una base de datos que contenga a todos los compuestos orgánicos que podrían ser reformados y, además, la información de sus propiedades microscópicas, para así, finalmente, poder realizar la evaluación de significancia estadística entre los candidatos a descriptores (variables de entrada al problema) y las propiedades macroscópicas (variables de salida).

Ahora bien, si se agregaran todas las moléculas existentes se lograría el objetivo, pero esto implicaría realizar una extensa búsqueda de todas ellas y de toda la información necesaria para cada una, lo que transformaría a este trabajo en una búsqueda en bases de datos y no en el tema central que acontece. También, al crecer el conjunto, provocaría que los cálculos (al realizar CAMPD o la optimización) se entorpezcan y se eleve el costo computacional.

Dado los argumentos expresados en el párrafo anterior, se impondrán restricciones sobre el conjunto de moléculas candidatas, las que serán mencionadas y explicadas más adelante. Cabe mencionar que estas restricciones o limitaciones están relacionadas con la composición y la estructura molecular de cada una de ellas.

### 3.2.1. Restricción por composición

Una hipótesis que se impone al conjunto de moléculas que será analizado y que se traduce en una limitación, es que solo se considerarán compuestos que tienen en su composición átomos de carbono, hidrógeno u oxígeno ( $C-H-O$ ), descartando a todas las demás moléculas pues no son reformadas corrientemente.

Específicamente, entre las moléculas que tienen oxígeno en su composición (compuestos orgánicos con presencia de oxígeno), solo se considerará al grupo de los alcoholes. Por su parte, debido al elevado número de registros científicos que existen en base a poli-alcoholes (que tienen más de un grupo  $OH$  en su composición) [3, 30, 31], se decidió no aplicar restricción sobre la cantidad de grupos  $OH$  presentes en las moléculas.

### 3.2.2. Restricción por estructura molecular

Otra restricción impuesta en el conjunto de moléculas orgánicas fue el largo máximo de cadena de carbonos aceptada. Se separará, además, el problema en dos: compuestos que no contienen oxígeno (alcanos) y compuestos que sí lo contienen (alcoholes). En ambos casos el largo de cadena podría ser diferente.

Para el caso de los alcanos, se estableció que la cota superior sería la molécula más grande que ha sido reformada según la información obtenida de registros científicos. Dado lo anterior, se encontró una lista de moléculas de gran tamaño que han sido reformadas y que forman parte de un compuesto que busca ser un sustituto del combustible de avión [7]. En la siguiente tabla se muestran tales moléculas.

Tabla 3.1: Características de los hidrocarburos presentes en un sustituto para combustible de avión, obtenida de *Shi et al.* [7].

Nombre	Fórmula molecular	PCI (kJ/mol combustible)
Iso-octano	$C_8H_{18}$	5064
n-Decano	$C_{10}H_{22}$	6293
n-Dodecano	$C_{12}H_{26}$	7512
n-Tetradecano	$C_{14}H_{30}$	8731
n-Hexadecano	$C_{16}H_{34}$	9949
Metilciclohexano	$C_7H_{14}$	4256
Ciclooctano	$C_8H_{16}$	4912
n-Butilbenceno	$C_{10}H_{14}$	5563
1,2,4,5-tetrametilbenceno	$C_{10}H_{14}$	5505
m-Xileno	$C_8H_{10}$	4330
Tetralin	$C_{10}H_{12}$	5356
1-Metilnaftaleno	$C_{11}H_{10}$	5594

De la Tabla 3.1 se obtiene que una de las moléculas reformadas es el *n-hexadecano*, la que

posee 16 carbonos. De lo anterior, y debido a que no se encontró registro de una molécula más larga que haya sido reformada, se consideró a ésta como límite superior. En cuanto al límite inferior, se tomará al metano ( $CH_4$ ) como tal, ya que es una de las moléculas mayormente reformada [5, 32, 33] y, a la vez, la de menor tamaño (largo de cadena 1C).

Ahora, si se consideran moléculas con presencia de oxígeno, específicamente alcoholes, se tomará como límite superior al hexanol ( $C_6H_{13}OH$ ) [31], ya que no se han encontrado registros científicos de moléculas de largo mayor que hayan sido reformadas, al contrario de tamaños inferiores a éste, donde sí existen [1, 2, 34, 35]. Para el tamaño mínimo se considerará la molécula más pequeña que es el metanol ( $CH_3OH$ ), ya que existen variados registros científicos que reportan el reformado de este alcohol con vapor de agua. [1, 36, 37].

### 3.2.3. Otras restricciones e hipótesis

Cuanto más larga sea una molécula orgánica, tanto más probable es la existencia de ramificaciones, lo que produce un incremento del número de compuestos y, en consecuencia, de la base de datos. Por esta razón, se fijó como hipótesis que solo se considerarán compuestos lineales (sin ramificaciones) para simplificar futuros cálculos y el manejo de datos. Cabe mencionar que la inclusión de estos se propondrá como trabajo futuro, ya que agregar más candidatos a la base de datos será sencillo y los pasos siguientes en la metodología serán concebidos de manera general.

Una restricción adicional es que no se tomarán en cuenta las moléculas con doble o triple enlace, pero se dejará también como trabajo futuro, ya que estas moléculas pueden considerarse como reformables, aunque presenten una dificultad en la ruptura de sus enlaces.

Se supuso, además, que las moléculas utilizadas se encuentran en el estado correspondiente a sus condiciones estándar y que de este punto se deben llevar a las condiciones deseadas, es decir, si una especie a  $25[^\circ C]$  y aproximadamente  $1[atm]$  de presión (condiciones estándar) está en estado sólido y se necesita a  $600[^\circ C]$ , se le debe entregar el suficiente calor para provocar el respectivo cambio de entalpía, asociado a calores sensibles y latentes, hasta llegar a la temperatura deseada. Se tomó este supuesto debido a que de esta forma se pueden calcular las principales necesidades energéticas del sistema SR. Como se verá más adelante, este cambio de entalpía para llevar cada componentes desde condiciones estándar hasta la temperatura de reacción, sumado al cambio de entalpía de reacción, permitirá estimar las necesidades energéticas más importantes del sistema SR, y, con ello, los moles de  $CO_2$  equivalente, como seguimiento primero del impacto ambiental.

### 3.2.4. Propiedades

Finalmente, restaría seleccionar las propiedades e información relevantes para cada uno de los elementos del conjunto de compuestos. Estos datos serán escogidos en base a las necesidades para el cálculo estimativo de las propiedades macroscópicas y, también, en función

de los candidatos a descriptores que se definirán más adelante, motivo por el cual se deberá esperar a obtener estos y así poder determinar la información que se necesita. En la Sección 3.4, se reabrirá la explicación de la base de datos y su creación.

### 3.3. Diseño de descriptores

Como se definió en la Sección 2.4, los descriptores son propiedades de un nivel microscópico que influyen de manera significativa en propiedades macroscópicas y que servirán para poder predecir el comportamiento de éstas. Para poder diseñarlos se tienen que responder preguntas relacionadas principalmente al modo en que las propiedades macroscópicas son explicadas desde un punto de vista microscópico. El foco de atención es, entonces, tratar de encontrar alguna característica a nivel molecular, ya sea ángulo de enlace, número de carbonos, u otra, que es responsable de hacer cambiar los valores de esas propiedades en un nivel macroscópico como lo son el costo, el rendimiento, etc.

Por lo anterior, se realizarán preguntas por cada propiedad macroscópica y, así, se obtendrá una lista de posibles descriptores, los que posteriormente serán estudiados estadísticamente para determinar su significancia.

Las preguntas que se realizarán para obtener los candidatos a descriptores y que se responderán con la ayuda de un grupo de expertos [38], son las siguientes:

- ¿Qué propiedades microscópicas se cree que condicionan cada propiedad macroscópica?
- ¿Qué característica de una molécula, mediante su modificación, provocará un cambio en la propiedad en estudio?

### 3.4. Base de datos

Como se mostró anteriormente, existen variados compuestos que se pueden reformar (ver Tabla 1.1 en Sección 1.1.2), entre los que se busca encontrar el mejor en términos económicos y ambientales<sup>1</sup>. Por otro lado, se debe realizar el análisis de significancia estadístico de propiedades microscópicas y macroscópicas para los compuestos seleccionados. Dado los motivos anteriores, se requiere almacenar estos compuestos y además ciertas propiedades e información como entalpía de formación, reacción para cada compuesto, entre otras.

Para poder almacenar estos se utilizará una base de datos, la que se creará en *MATLAB*®<sup>2</sup> [39] a través del uso de una estructura de datos. A pesar de que *MATLAB*® no es un lenguaje

---

<sup>1</sup>Hay que recordar que la búsqueda de la mejor molécula para el *SR* será abordada en futuros trabajos, y que el presente desarrollo se enmarca en la primera etapa: la definición de descriptores (ver objetivos en la Sección 1.2).

<sup>2</sup>Si bien *MATLAB*® es un lenguaje de programación, también posee un ambiente interactivo para computación numérica que provee de diferentes herramientas propias.



desarrollado específicamente para el manejo de bases de datos, se escogió debido a que se presume que el conjunto de moléculas con sus respectivas propiedades e información no será extensa y porque posteriormente se usará este lenguaje para el tratamiento de los datos.

Los datos que se utilizaran para completar esta base de datos se dividen en dos partes: el conjunto de moléculas seleccionadas y las propiedades escogidas para cada uno de ellos. En la Sección 3.2 se entró en detalle de como estos datos fueron escogidos.

### 3.5. Elección del tamaño de *piezas*

Las *piezas*, como se mencionó anteriormente (ver Sección 2.1), comprenden un conjunto de uno o varios elementos que, al unirse entre ellos, formarán una unidad mayor. Un símil es considerar un juego *Lego*, donde los bloques del juego son las *piezas*, y la unidad mayor será la unión de algunos de estos bloques que forman una estructura.

Estos elementos son fundamentales al realizar CAMPD, por lo que su elección es una tarea importante. Las principales consideraciones que se deben tomar son que las *piezas* puedan representar la mayor cantidad de moléculas de la base de datos que se posee, y que durante la generación de estructuras a partir de ellas, no se incurra en un coste computacional alto al tener que filtrar compuestos infactibles fisicoquímicamente, es decir, que el *ruido* generado sea bajo.

Para armar moléculas utilizando las piezas definidas, y posteriormente poder obtener la cantidad de estructuras que se pueden generar a partir éstas, se tomarán algunas consideraciones generales que serán enunciadas a continuación:

- El largo de la cadena para una molécula, será medido a través del número de carbonos de la cadena principal.
- La cantidad de carbonos unidos en una cadena, no superará el número máximo impuesto para alcanos ( $C - H$ ) y alcoholes ( $C - H - O$ ) que se consideró en la Sección 3.2.2, es decir, 16 y 6 respectivamente.
- Si bien puede ser posible formar ramificaciones usando las piezas, éstas no serán consideradas dentro de las estructuras armadas.
- La cantidad de enlaces que acepta cada carbono será de 4, pero se aceptará el caso en que el carbono no tenga sus 4 enlaces ocupados, es decir, se aceptan especies como  $CH_3$ .

Estas indicaciones en el armado, si bien pueden ser consideradas restricciones, tienen como objetivo hacer que el conjunto tenga un límite numerable y no infinito, que es el caso que pasaría si se dejara libre, por ejemplo, el número máximo de carbonos ( $C_\infty H_\infty$ ).

Ahora, para el cálculo del *ruido* que se genera por grupo de piezas, se utilizará la ec. 3.1 donde  $\#Ruido_i$  es el número de estructuras que se pueden generar con los bloques, pero que no se encuentran en la base de datos,  $\#G_i$  es el número de estructuras moleculares que se

pueden formar con un grupo de piezas, y  $\#BD$  es la cardinalidad de moléculas disponibles en la base de datos.

$$\#Ruido_i = \#G_i - \#BD \quad (3.1)$$

### 3.6. Análisis de significancia estadístico

Una vez definidas las propiedades macroscópicas y propiedades microscópicas, estos últimos candidatos a ser descriptores, se debe proceder a comprobar si existe alguna relación entre ambas. Para ello se realizará lo siguiente:

1. Análisis cuantitativo y cualitativo entre propiedades microscópicas y macroscópicas.
2. Análisis de componentes principales (PCA).

Ambos análisis serán explicados en las siguientes secciones. Es importante destacar que comprobar la existencia de alguna relación, ya sea lineal o no lineal, entre las variables en estudio es difícil, porque las pruebas que se utilizan regularmente son específicas para comprobar solamente el caso lineal. Por esto, al obtener un resultado negativo a través del análisis cuantitativo, solo se descartaría la opción de que entre ambas variables exista una dependencia lineal, dejando así dos opciones posibles: que la dependencia sea no lineal, o que realmente no exista relación alguna.

Antes de continuar se definirán de forma general algunos conceptos enlazados con los análisis que se desarrollarán en el presente trabajo.

#### Conceptos [40, 41]

A continuación, se expondrán los conceptos dispuestos en orden alfabético.

#### Covarianza

Es una medida de la asociación lineal entre dos variables, cuya fórmula es:

$$\frac{\sum_{i=1}^N X_i Y_i}{N} - \bar{X}\bar{Y} \quad (3.2)$$

Una desventaja de la covarianza como medida de asociación es que su valor depende de las unidades de medida de las variables de interés, por lo que, para solucionar este problema, se suele usar algún coeficiente de correlación como será explicado más adelante (ver Sección 3.6.1).

**Desviación estándar**

Valor que mide la dispersión de una característica de los individuos alrededor de la media del grupo. Es la raíz cuadrada de la varianza.

**Estadístico**

Es un valor calculado en base a los datos que se obtienen sobre una muestra y corresponde a una estimación de algún parámetro estadístico (como la esperanza o la varianza).

**Hipótesis alternativa**

Aseveración opuesta a la hipótesis nula en un test o contraste de hipótesis.

**Hipótesis nula**

Afirmación establecida por el investigador sobre la población de estudio cuando realiza un test o contraste de hipótesis.

**Media**

Valor numérico en torno al cual se sitúan los valores de una característica de los individuos.

**Muestra**

Grupo de individuos extraídos aleatoriamente de la población de estudio.

**Parámetros**

Es cualquier valor característico de la población como por ejemplo: la media poblacional, la desviación estándar, entre otros.

**p-value**

Valor comprendido entre 0 y 1 que mide el grado de acuerdo entre la información que proporciona la muestra y la hipótesis nula del test de hipótesis

**Rango**

Es un número que mide la amplitud de los valores de un conjunto de datos.

**Test de hipótesis**

Técnica estadística utilizada para decidir si una afirmación establecida por el investigador sobre la población de estudio es compatible con los resultados observados en una muestra.

**Variables Cualitativas**

Característica que recoge una cualidad de los individuos de una muestra.

**Variables Cuantitativas**

Característica de los individuos de una muestra que puede medirse con un instrumento y lleva asociada una unidad de medida.

### **Variables dependientes**

Tipo de variable cuyo su valor dependerá de otras variables.

### **Variables independientes**

Tipo de variable cuyo valor no dependerá de otras variables.

### **Varianza**

Valor que mida la dispersión de una característica de los individuos alrededor de la media del grupo.

Una vez introducidos estos conceptos, se procederá a explicar cada uno de los análisis que se realizarán.

## **3.6.1. Análisis cuantitativo y cualitativo**

Este análisis contempla dos etapas: la primera, relacionada con la determinación del coeficiente de correlación entre descriptores y propiedades macroscópicas, con el fin de determinar si existe una dependencia lineal entre ellas; y la segunda, donde se estudiara gráficamente si dicha relación es correcta y si existe algún tipo de otra dependencia entre las variables restantes [42–48]<sup>3</sup>.

### **Gráfico de dispersión y Coeficiente de correlación**

Cuando se habla de *correlación* o de *relación* se hace referencia al grado en que varían dos o más variables en conjunto. Una primera forma de determinar esto, es a través de un gráfico de dispersión donde se muestran gráficamente las variables que se están analizando, lo que permite determinar la relación existente entre variables de un modo cualitativo. El problema que presenta este método es que no siempre se puede determinar con facilidad si existe una relación entre las variables, ya que generalmente se grafican nubes de puntos las que pueden adquirir distintas formas cuyo patrón no es fácil de identificar.

Otra forma de determinar la relación entre variables es a través de coeficientes de correlación, que sirven para cuantificar el grado de relación lineal entre dos o más variables cuantitativas y para valorar el grado de ajuste de los puntos a una línea recta. La gran limitante de estos coeficientes es que solo sirven para medir si existe una dependencia lineal entre las variables y no si existe alguna de otro tipo. A continuación se explicarán algunos coeficientes de correlación que se utilizan generalmente.

#### ***Coefficiente de correlación de Pearson***

Es uno de los coeficientes más utilizados para estudiar el grado de relación lineal existente entre dos variables. Se denota por  $r$  y adquiere valores entre  $-1$  y  $1$ . El signo de este coeficiente

---

<sup>3</sup>Esta bibliografía sera utilizada a lo largo de esta sección, por lo que se citará solo una vez.

indicará la dirección, mientras que el valor absoluto indicará el grado de relación. Cuando  $r$  toma un valor de 1 se dice que es una relación perfecta positiva y cuando adquiere un valor de  $-1$  es una relación perfecta negativa. Para calcular este coeficiente se utiliza la siguiente ecuación:

$$r_{xy} = \frac{\frac{\sum XY}{N} - \bar{X}\bar{Y}}{S_x S_y} \quad (3.3)$$

En la relación anterior,  $X$  y  $Y$  son los valores de las variables medidas,  $N$  es la cantidad de muestras tomadas,  $\bar{X}$  y  $\bar{Y}$  es la media muestral, y  $S_x$  y  $S_y$  son las desviaciones estándar de las variables.

Existen condiciones de aplicación del coeficiente, de modo que, al no cumplir una de ellas, éste no se podría utilizar. Éstas son:

1. Igualdad de varianzas de  $X$  y de  $Y$ .
2. Los datos deben proceder de una población distribuida normalmente.

Debido a esto último no se podría utilizar este coeficiente, ya que las variables en estudio no provienen de una población normal. Por esta razón se optó por escoger el coeficiente de correlación de Spearman, que se utiliza, en general, cuando no se cumplen las condiciones expuestas.

### *Coefficiente de correlación de Spearman*

Este coeficiente, al igual que el de Pearson, mide el grado de relación lineal existente entre dos o más variables, pero se diferencia en que utiliza rangos para calcular el valor de dicho coeficiente. Se denota por la letra  $\rho$  y toma valores entre  $-1$  y  $1$ , siendo el valor  $1$  una relación perfecta positiva; y el valor de  $-1$ , una relación perfecta negativa. El signo indicará la dirección de la relación y el valor absoluto el grado. La ecuación que se utiliza para su cálculo es la siguiente:

$$\rho = 1 - \frac{6\sum d^2}{N(N^2 - 1)} \quad (3.4)$$

En la fórmula anterior  $d$  es la diferencia entre los estadísticos de orden correspondientes a  $X$  y a  $Y$ .

Una vez obtenido el coeficiente de correlación entre las variables (candidatos a descriptores y propiedades macroscópicas) se realizará una prueba de significancia para éste, con la que se probará si existe realmente una relación entre las variables, o si el resultado obtenido es solo resultado del *azar*.

## Prueba de significancia

La prueba de significancia para el coeficiente de correlación consiste en preguntarse cuán confiable es el valor obtenido a través de este coeficiente, es decir, verificar que las variables analizadas están relacionadas. Con respecto a lo anterior, se pueden plantear dos hipótesis posibles:

- $H_0$  (hipótesis nula): No hay relación entre las variables analizadas, es decir, son independientes.
- $H_1$  (hipótesis alternativa): Existe relación entre las variables.

Para realizar la prueba de significancia, se utilizará el estadístico de referencia:

$$t = r \cdot \sqrt{\frac{N-2}{1-r^2}} \quad (3.5)$$

Este estadístico se distribuye según una función de probabilidad t-Student con  $N-2$  grados de libertad. Cabe notar que para este caso  $r = \rho$ .

Para determinar si se acepta o rechaza la hipótesis nula, se compara el valor obtenido (ver ec. 3.5) con respecto a la función de probabilidad t-Student con un nivel de significancia  $\alpha$  y  $N-2$  grados de libertad, con lo que, finalmente, se pueden dar los siguientes casos:

- $t > t_{(\alpha, N-2)}$ : Se rechaza la hipótesis nula, por lo que se puede decir que las variables están relacionadas.
- $t \leq t_{(\alpha, N-2)}$ : Se acepta la hipótesis nula, es decir, ambas variables no están relacionados (son independientes).

Otra forma de realizar esta comparación, y finalmente decidir entre aceptar o rechazar la hipótesis nula, es obteniendo el valor de la probabilidad asociada al valor de  $t$  ( $P(t < t_{(\alpha, N-2)}) = 1 - \alpha$ ), lo que equivale a obtener el *p-value*. Este último hace referencia a la probabilidad de que los resultados obtenidos sean producto del azar y no de una real relación entre las variables. De esto se desprende que el valor del *p-value* ( $1 - \alpha$ ) debe ser pequeño si dos variables están relacionadas. A partir de esto se tendrán dos casos análogos a los anteriores:

- $p\text{-value} < 1 - \alpha$ : Se rechaza la hipótesis nula, por lo que las variables están realmente relacionadas.
- $p\text{-value} \geq 1 - \alpha$ : Se acepta la hipótesis nula.

Para el caso que se está estudiando se fijará el valor de significancia,  $\alpha$ , como 0,95, lo que equivale a un riesgo de falla de un 5% ( $1 - \alpha = 0,05$ )<sup>4</sup>.

---

<sup>4</sup>Se fijó este valor, ya que es el utilizado en la mayoría de los casos.

## Metodología para el análisis cuantitativo y cualitativo

Como se mencionó anteriormente, el análisis cuantitativo y cualitativo, se divide en dos etapas en las que se realizarán las siguientes actividades:

- **Cálculo del coeficiente de correlación:** Se utilizará el coeficiente de correlación de Spearman para determinar qué candidatos a descriptores están relacionados linealmente con las propiedades macroscópicas. Para ello se realizará lo siguiente:
  - Se tomará una propiedad macroscópica y se calculará su coeficiente con respecto a cada uno de los candidatos a descriptores. Se repetirá esta operación con todas las propiedades macroscópicas seleccionadas.
  - Se realizará la prueba de significancia para cada valor obtenido.
  - Se seleccionarán los descriptores que posean un *p-value* menor a 5% y un coeficiente de correlación mayor a 0,35 (se considerará valor absoluto para este último).
- **Análisis cualitativo:** Se utilizarán gráficos de dispersión para comprobar las relaciones obtenidas anteriormente y para determinar si existe otra relación entre las variables que no haya sido reportada por el análisis cuantitativo. Para esto se graficarán todos los pares de relaciones entre propiedades macroscópicas y candidatos a descriptores, para luego realizar un análisis visual de estos gráficos en función de sus tendencias.

### 3.6.2. Análisis de componentes principales (PCA)

A partir del análisis cuantitativo y cualitativo se debe seleccionar, para cada propiedad macroscópica, un grupo de candidatos a descriptores con los cuales se establece la existencia de alguna relación (lineal), para así, proceder a realizar el análisis de componentes principales (PCA) [42–48].

El PCA es una técnica estadística que consiste principalmente en resumir la información que se posee, o en reducir las dimensiones del problema (disminuir la cantidad de variables). Se usa generalmente cuando se tienen muchas variables y se desea reducir su número perdiendo la menor cantidad de información posible. Las variables resultantes (componentes principales) serán una combinación lineal de las variables originales e independientes entre sí.

Las etapas del análisis de componentes principales son las siguientes:

- **Análisis de la matriz de correlaciones:** Para realizar el PCA se asume que las variables tienen un alto grado de correlación, ya que así se asegura que hay variables que son redundantes, es decir, que se puede representar la información con menos variables de las que se posee.
- **Selección de los factores:** Para seleccionar los factores, se deben ordenar de tal forma que el primero debe recoger la mayor proporción de la variabilidad original, el segundo debe contener la máxima proporción de la variabilidad no recogida por el primero; y

así sucesivamente. Los componentes principales serán entonces los factores que recojan la variabilidad (varianza) que se desee.

- **Análisis de la matriz factorial:** Los resultados obtenidos en la etapa anterior, se representan en una matriz en que las filas son las variables originales y las columnas son los factores ordenados de mayor a menor variabilidad. La matriz contendrá los valores de los coeficientes de las variables (ver ec. 3.6).
- **Interpretación de los factores:** Existen características esperadas para que la interpretación de los factores obtenidos sea fácil, estas son:
  - Los coeficientes deben ser cercanos a 1 para poder determinar la importancia de cada uno en los factores definidos.
  - Una variable debe tener coeficientes elevados solo con un factor, ya que que así no estarán contenidas en más de un factor.
  - No deben existir factores con coeficientes similares, ya que no se podría discriminar la importancia de cada variable.
- **Cálculo de las puntuaciones factoriales:** Son los valores que adquieren las componentes principales para cada caso, que finalmente serán las variables que resumirán la información y reducirán el número de variables. Éstas se calculan mediante la ecuación siguiente:

$$\hat{X}_i = a_{i1}Z_{1j} + \dots + a_{ik}Z_{kj} = \sum_{s=1}^k a_{is} \cdot Z_{sk} \quad (3.6)$$

En la ecuación anterior  $a$  son los coeficientes de la matriz,  $Z$  son los valores estandarizados de las variables  $X$  (grupo de descriptores para este caso) y  $\hat{X}$  son los factores (componentes principales).

## Metodología para el análisis de componentes principales

Las etapas para este análisis serán las siguientes:

- A partir del análisis cualitativo y cuantitativo se agruparán los candidatos a descriptores de acuerdo a la propiedad macroscópica con la que están relacionados<sup>5</sup>.
- Para cada grupo de descriptores, se realizará un PCA para determinar nuevas variables que resuman la información de cada propiedad macroscópica.
- Se eliminarán algunas variables obtenidas de acuerdo a la varianza (determinar componentes principales importantes), es decir, se seleccionarán los factores o variables que resuman la mayor cantidad de información de la propiedad macroscópica con la que están relacionados.
- Se analizarán estas variables, determinando qué candidatos a descriptores son mayormente significativas para el comportamiento de cada propiedad macroscópica.

---

<sup>5</sup>Puede existir el caso en que un mismo candidato a descriptor esté relacionado con dos o más propiedades macroscópicas a la vez.



# Capítulo 4

## Resultados y discusiones

A partir del problema de elección de la *mejor* molécula a reformar y enmarcado al presente trabajo, acerca de la definición de descriptores significativos y elección de tamaño *óptimo* de piezas, se propuso, en el capítulo anterior (ver Capítulo 3), una metodología que abarca la definición de las propiedades objetivo, selección de un conjunto de moléculas reformables y datos para el problema, determinación de candidatos a descriptores, el análisis estadístico de estas últimas y, finalmente, la selección de un tamaño de piezas que represente de forma *óptima* a este conjunto de compuestos.

Dado lo anterior, a continuación se mostrarán en detalle los resultados y discusiones obtenidas a partir de esto.

### 4.1. Propiedades macroscópicas

Las propiedades macroscópicas, como se mencionó en la Sección 3.1, son las características objetivo que se optimizarán a futuro. Según los criterios mencionados en la Sección 3.4, se definieron como propiedades macroscópicas deseables al costo de producción, el rendimiento de hidrógeno, las emisiones y la disponibilidad. A continuación se hará, primero, una breve descripción y la forma en que estas propiedades fueron estimadas, para luego seleccionar algunas que serán abordadas en esta etapa. Se debe dejar claro que las estimaciones fueron necesarias para realizar el estudio de significancia estadístico entre estas propiedades y los candidatos a descriptores, y no hay que confundir con las estimaciones que se obtendrán a través de los QSPRs en trabajos futuros.

- **Costo de producción:** se refiere al costo de realizar el reformado del compuesto orgánico, es decir, cada uno de los gastos que se incurre al llevar a cabo el proceso, incluyendo elementos como materias primas, equipos, tratamiento de riles, etc. Cabe señalar que en esta etapa del trabajo no se tienen definidos los equipos y materias primas que serán necesarios, razón por la que se estimó el costo de producción en

función de la economía potencial, esto es en función del costo del compuesto orgánico que se va a reformar.

- **Rendimiento de hidrógeno:** se refiere al flujo de hidrógeno producido posterior al reformado incluyendo pérdidas en el proceso. Como no se tiene conversión de la reacción, ni tampoco el proceso específico que se llevará a cabo (solo que será un sistema *SR*), se supuso que el flujo de hidrógeno de salida, es el máximo que se puede obtener al reformar la molécula, es decir, se asumió una conversión de 100 % y que, además, no existirían pérdidas.
- **Emisiones:** esta propiedad macroscópica engloba a todos los elementos contaminantes del medio ambiente que son emitidos durante el proceso y, también, incluye la estimación de la huella de carbono de cada uno de los participantes en el reformado. Hay que mencionar que, como se dijo anteriormente, no se tiene definido el proceso completo (eficiencias, equipos, entre otros), motivo por el cual se supuso que las emisiones tendrán dos orígenes principales: dióxido de carbono producto de la reacción de reformado, el que estará dado por la estequiometría de la reacción; y la energía o calor necesario para que ocurra la reacción de reformado (*SR*), la que se obtuvo a partir de la entalpía de reacción. Se consideró este último, ya que para producir el calor necesario para la reacción, se necesita quemar combustible, de acuerdo a la opción industrial actual, para producir este calor, y con esto se liberará dióxido de carbono (o monóxido de carbono).
- **Disponibilidad:** se refiere a la cercanía geográfica de la materia prima con el lugar en donde llevará a cabo el proceso. Se propone esta propiedad, pues no todos los compuestos orgánicos serán igualmente accesibles desde cualquier parte del mundo, por lo que es importante saber la disponibilidad del compuesto según la zona geográfica.

De las propiedades listadas anteriormente, solo se dejaron las tres primeras porque abarcan puntos cruciales desde el punto de vista de las características del proceso industrial de reformado. La disponibilidad no se consideró, ya que es una variable difícil de definir en esta etapa, pero se dejará como trabajo futuro pues se propone que la elección de la mejor molécula para el proceso de reformado podría tener diferentes respuestas, dependiendo de la región del planeta.

#### 4.1.1. Cálculo de emisiones

Tal como se mencionó anteriormente, para estimar las emisiones del proceso de reformado se consideró, como primera aproximación, al  $CO_2$  liberado en la reacción de reformado (*SR*) y al producido al quemar el combustible para generar el calor necesario para que ocurra la reacción.

El primero se obtuvo a partir de la reacción de reformado genérica (ver ec. 4.4) y sus coeficientes estequiométricos (ver Tabla 4.1), suponiendo que la cantidad de moles del compuesto que se estaba reformando era uno. Con esto se obtuvo que la cantidad de  $CO_2$  generado era igual a la cantidad de carbonos que poseía la molécula (o igual al coeficiente estequiométrico del  $CO_2$ ).

Para el segundo caso, se debió calcular el calor utilizado en la reacción (entalpía de reacción) y estimar, con esto, la masa de combustible necesaria para abastecer tal necesidad energética, asociada al reformado de un mol de compuesto orgánico. Se optó por gas natural, pues es el combustible más utilizado en este contexto. Luego, se utilizó el concepto de  $CO_2$  equivalente [49], que es una medida universal que considera la equivalencia en dióxido de carbono para determinar la contribución de cada gas al efecto invernadero, y también los factores de emisión, los cuales son una herramienta que permite conocer la cantidad de  $CO_2$  equivalente que se emite a la atmósfera con la utilización de diversos tipos combustibles [50].

Para obtener la masa estimada de emisiones ( $CO_2$  equivalente) por consumo de gas natural se utilizó la siguiente fórmula [51]:

$$E_{\Delta H_r} = FE \cdot \Delta H_r \quad (4.1)$$

En la ecuación anterior,  $E_{\Delta H_r}$  ( $[\frac{kmolCO_2}{kmolOrg}]$ ) son las emisiones provocadas por la generación del calor necesario para la reacción;  $\Delta H_r$  ( $[\frac{kJ}{kmolOrg}]$ ) es la entalpía de reacción; y  $FE$  ( $[\frac{kmolCO_2}{kJ}]$ ) es el factor de emisión para el combustible utilizado en la generación de energía, gas natural, con un valor de  $FE = 56.100[\frac{kgCO_2}{TJ}] = 1,275 \times 10^{-6}[\frac{kmolCO_2}{kJ}]$  [52]. Así, por ejemplo, si se tiene que la entalpía de reacción para un compuesto es  $400.000[\frac{kJ}{kmolOrg}]$ , entonces las emisiones provocadas por la generación del calor necesario para la reacción, utilizando la ec. 4.1, es  $0,51[\frac{kmolCO_2}{kmolOrg}]$ , valor que estimará las emisiones de  $CO_2$  para este compuesto.

#### 4.1.2. Cálculo del Costo y el rendimiento de hidrógeno

La estimación de estas propiedades macroscópicas resultó más fácil de realizar que las emisiones, ya que se consideraron valores directos para estos. Para el costo de producción, se centró la atención en el precio de las materias primas, ya que no se cuenta con la definición del proceso completo. Los valores utilizados se obtuvieron a partir de la búsqueda en una base de datos en la web [53] y fueron reportados en  $[\frac{CLP}{ml}]$ . Es importante destacar que en la página web vendían la materia prima por ciertas cantidades, por lo que se decidió normalizar esto por  $1[ml]$  del compuesto para poder realizar una comparación entre ellos posteriormente.

Por otro lado, el rendimiento de hidrógeno fue calculado directamente a través de la estequiometría de la reacción, siendo este el correspondiente al hidrógeno que se produce al reaccionar un mol del compuesto reformable y suponiendo una conversión del 100%. Los valores fueron reportados en  $[\frac{kmolH_2}{kmolOrg}]$ . La obtención de los coeficientes estequiométricos será explicado más adelante en la Sección ??.

#### 4.1.3. Almacenamiento de los datos

Los valores de las propiedades macroscópicas fueron almacenados en una matriz, como se mencionó en la Sección 3.4, la que se puede observar en la Figura 4.1. Los valores resultantes

se pueden ver en Anexo A.1. Cabe señalar que la cantidad de compuestos de los que se tienen todas estas propiedades, resultó inferior al total de la base de datos, ya que no se pudieron obtener o estimar todos los datos, como por ejemplo la entalpía de reacción.

Para facilitar las estimaciones de las propiedades para todos los compuestos que se tenían, se crearon diversas funciones en MATLAB®, las que serán explicadas brevemente a continuación (mayor detalle, ver Anexos E.6)

- **PropMacros**: Función que devuelve una matriz con los valores de las propiedades macroscópicas, según la estructura enunciada anteriormente.
- **CostoNorm**: Función que normaliza el costo de la materia prima a  $[\frac{CLP}{ml}]$ .
- **isDHR**<sup>1</sup>: Función que determina si un compuesto tiene todos los datos necesarios para ser agregado a las propiedades macroscópicas.

	CAS	Rendimiento	Emisiones	Costos
Compuestos	173-45-5	9	100	\$20000
	172-5-10	8	240	\$25000
	.	.	.	.
	.	.	.	.
	500-4-8	7	145	\$23000

Figura 4.1: Forma de almacenamiento para los valores de las propiedades macroscópicas medidas. Las filas representan los compuestos y las columnas las propiedades almacenadas.

## 4.2. Conjunto de moléculas reformables

En la Sección 3.2 se ha relatado la metodología de selección de compuestos para la construcción de una base de datos relacionada, que contendrá propiedades de un universo de moléculas que serán candidatas, en la búsqueda de la mejor molécula a reformar.

<sup>1</sup>Esta función también será útil para la definición de los descriptores.

Luego de aplicadas estas restricciones, se obtuvieron alrededor de 40 moléculas que fueron agregadas a la base de datos junto a sus respectivas propiedades definidas en la Sección 3.2.4. Los compuestos seleccionados y la información almacenada se puede ver en las tablas del Anexo B.1.

Como se mencionó también, la información fue almacenada en una estructura de datos en MATLAB® que consta de dos partes (ver Figura 4.2): los campos, que corresponden a los nombres de las propiedades o información que se desea guardar y un arreglo de celdas (bloques de información) asociado a cada campo, los que contienen los valores de los datos que se están almacenando y cuyo largo es igual a la cantidad de compuestos que serán analizados, luego de aplicadas las restricciones vistas en la Sección 3.2.1.

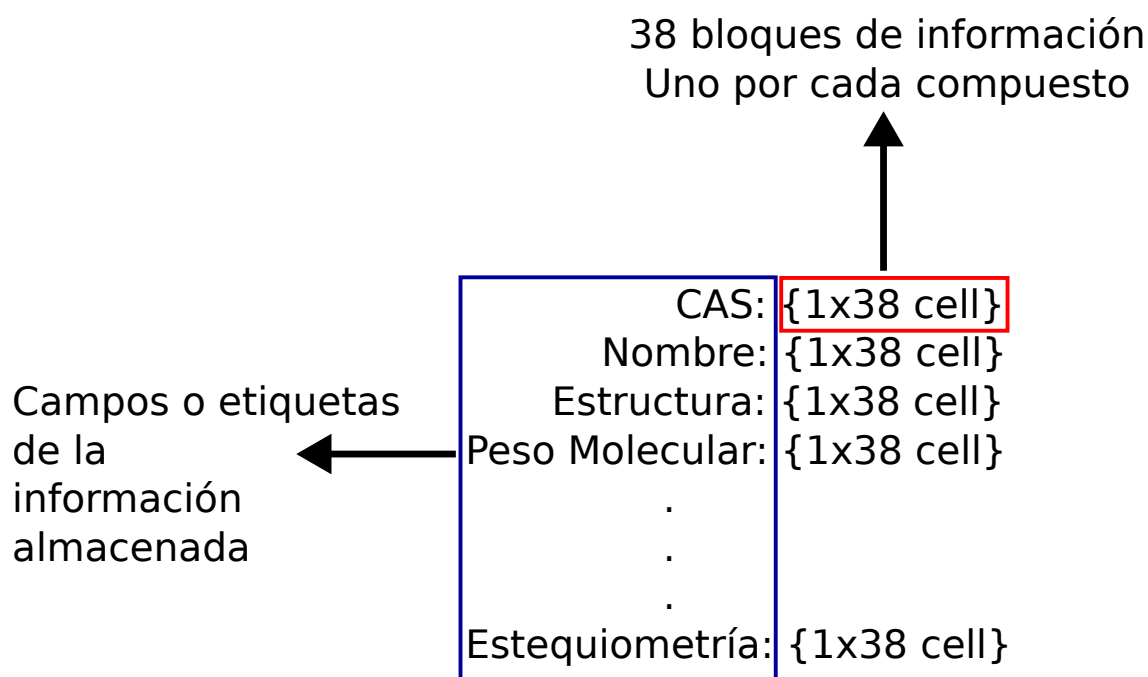


Figura 4.2: Estructura de datos que almacenó la información acerca de los compuestos reformables y algunos datos de ellos.

Dada la configuración anterior, para acceder a la información de un compuesto en esta base de datos, basta con saber la posición  $i$  de dicho compuesto y el nombre de la información (*campo*) que se requiere. Para facilitar la búsqueda de la posición  $i$  se creó la siguiente función en MATLAB®:

- **SearchBD:** Entrega la posición  $i$  de un compuesto en la base de datos, buscando ésta de acuerdo a algún criterio como número de registro (CAS), nombre del compuesto o fórmula química. Para mayor detalle ver Anexos E.1.

### 4.2.1. Propiedades y datos almacenados

La información almacenada en la base de datos fue seleccionada a partir de los datos que se necesitaban para calcular las propiedades microscópicas y macroscópicas (variables de entrada y salida respectivamente), que posteriormente se utilizarán para realizar el estudio de significancia estadístico entre éstas, y, además, para poder identificar los compuestos que estaban en base de datos. A continuación se mostrará la información agregada, el nombre del campo en la estructura de datos (en paréntesis) y una breve descripción de cada uno.

- **CAS** (CAS): Número de registro para compuestos orgánicos e inorgánicos que servirá para identificar cada molécula en la base de datos [54].
- **Nombre** (Name): Nombre del compuesto.
- **Fórmula química** (Structure): Fórmula simbólica del compuesto que servirá para identificar el número de carbonos, hidrógenos y oxígenos presentes.
- **Peso Molecular** (WT): Peso molecular del compuesto en  $[\frac{kg}{kmol}]$  o  $[\frac{g}{mol}]$ .
- **Densidad** (Densityg\_ml): Densidad del compuesto medido a temperatura de referencia (ver campo *Temperatura de densidad*) y presión estándar. Sus unidades son  $[\frac{g}{ml}]$ .
- **Temperatura densidad** (Tdensity\_C): Temperatura a la cual fue medida la densidad en  $[^{\circ}C]$ .
- **Punto de ebullición** (BoilingPoint\_C): Temperatura de ebullición del compuesto en  $[^{\circ}C]$ .
- **Punto de fusión** (MeltingPoint\_C): Temperatura de fusión del compuesto en  $[^{\circ}C]$ .
- $\Delta H_f$  (DHfkJ\_kmol): Entalpía estándar de formación del compuesto medida en  $[\frac{kJ}{kmol}]$ .
- $\Delta H_r$  (DHrkJ\_kmol): Entalpía de reacción del compuesto medida en  $[\frac{kJ}{kmol}]$ . Este dato fue calculado a una temperatura de  $600[^{\circ}C]$ . El detalle de este cálculo se explicará más adelante en esta misma sección.
- $c_{P,sol}$  (Cpsol\_kJ\_kmolC): Calor específico molar del compuesto en estado sólido en  $[\frac{kJ}{kmol \cdot ^{\circ}C}]$ .
- $\Delta H_{fusion}$  (DHfusion\_kJ\_kmol): Entalpía de fusión para la especie medida en  $[\frac{kJ}{kmol}]$ .
- $c_{P,liq}$  (Cpliq\_kJ\_kmolK): Calor específico molar del compuesto en estado líquido en  $[\frac{kJ}{kmol \cdot K}]$ .
- $\Delta H_{vap}$  (DHvap\_kJ\_kmol): Entalpía de vaporización para la especie medida en  $[\frac{kJ}{kmol}]$ .
- $c_{P,gas}$  (Cpgas\_kJ\_kmolK): Calor específico molar del compuesto en estado gaseoso en  $[\frac{kJ}{kmol \cdot K}]$ .
- $T_{crit}$  (Tc\_K): Temperatura crítica de la molécula medida en  $[K]$
- **Precio** (Price): Precio del compuesto en  $[CLP]$ .
- **Cantidad** (Quantity): Cantidad que se vende del compuesto. Las unidades de este varían entre  $[g]$  y  $[ml]$ , por lo que se decidió transformar todas estas cantidades a  $[ml]$  a través de la densidad.
- **Número de carbonos** (CarbonNum): Número de carbonos del compuesto.
- **Número de hidrógenos** (HydrogenNum): Número de hidrógenos que rodean a la cadena de carbonos, sin considerar el que está unido al oxígeno en el grupo hidroxilo (para alcoholes).

- **Número de grupos hidroxilos** (HidroxideNum): Número de grupos hidroxilos que rodean a la cadena de carbonos.
- **Estequiometría** (Stoichiometric): Estequiometría de la reacción general, utilizada para calcular el  $\Delta H_r$ .

La mayoría de los datos mencionados anteriormente fueron obtenidos a partir de bases de datos en la web [53, 55, 56], mientras que los restantes fueron calculados y/o estimados como se explicará a continuación.

## Cálculos y estimaciones

Los cálculos y estimaciones son realizados debido a que, si bien algunos valores de las propiedades y datos de cada compuesto son obtenidos directamente de bases de datos en la web, otros necesitan de algún cálculo para su determinación, como la entalpía de reacción, que dependerá de la temperatura a la que ocurra la reacción y de las moléculas que participan en ella; u otros deben ser estimados, ya que no se encontraron sus valores y eran necesarios para el trabajo, como por ejemplo el calor específico molar para el estado líquido, que se utiliza para calcular la entalpía de reacción para algunas moléculas y con ello para obtener las emisiones generadas.

Los principales cálculos realizados fueron la obtención del número de átomos de hidrógeno, carbono y grupos  $OH$  presentes en cada molécula, los coeficientes estequiométricos de la reacción de reformado con vapor de agua y la determinación de la entalpía de reacción a  $600[^\circ C]$ . Se escogió esta última, ya que, si bien la temperatura óptima de reacción depende del uso de catalizador y del tipo de reformado que se realice (además de otros factores), existen registros científicos que coinciden en el uso de temperaturas similares a la elegida [32, 57].

Para obtener el número de átomos y grupos  $OH$  presentes en cada molécula, se creó una función en MATLAB® (*getAtomMol*), que toma el campo *Structure* (estructura atómica) de la base de datos y obtiene el número de hidrógenos, carbonos y grupos  $OH$  para cada compuesto. Esto se logró, ya que se ocupó el hecho de que los datos almacenados en este campo son *strings* que tienen la siguiente forma:

C	n	H	m	O	z
---	---	---	---	---	---

En este registro, cada átomo presente en una molécula aparece escrito con su distinción simbólica ( $C$ : carbono;  $H$ : hidrógeno;  $O$ : oxígeno) y lo sigue un número que indica la cantidad total de átomos presentes en la molécula ( $n$ , para carbono;  $m$ , para hidrógeno; y  $z$ , para oxígeno). Para el caso en que hay más de 9 átomos de una especie se considerarán 2 caracteres dentro del *string*; para el caso en que no hay presencia de algún átomo, no aparecerá la identificación simbólica ni la cantidad; y si solo existe un átomo de una especie, no aparecerá escrita el número, pero sí la identificación simbólica. Por ejemplo si se tiene el compuesto etanol, que consta de dos carbonos, seis hidrógenos y un oxígeno, su notación simbólica sería  $C2H6O$ ; para el glicerol (tres carbonos, ocho hidrógenos y tres oxígenos) sería  $C3H8O3$ ; y para el metanol (un carbono y cuatro hidrógenos)  $CH4$ .

Por lo dicho en anteriormente, y sumado al hecho que todas las moléculas siempre poseerán carbono e hidrógeno, la función antes mencionada se basa en la búsqueda de los caracteres  $H$  y  $O$  en la cadena, distinguiendo con ello la posición de los datos a extraer (número de átomos). Para mayor detalle, ver Anexo E.2.

Por otro lado, para calcular la entalpía de reacción a una determinada temperatura ( $600[^\circ C]$ ) se usó la siguiente ecuación [58]:

$$\Delta H_r(600[^\circ C]) = \sum_i \alpha_i \Delta H_{f,prod,i}(600[^\circ C]) - \sum_i \beta_i \Delta H_{f,react,i}(600[^\circ C]) \quad (4.2)$$

En la ecuación anterior,  $\Delta H_{f,prod,i}(600[^\circ C])$  y  $\Delta H_{f,react,i}(600[^\circ C])$  son las entalpías de formación a  $600[^\circ C]$  para los productos y reactantes respectivamente; y  $\alpha_i$  y  $\beta_i$ , sus coeficientes estequiométricos en la reacción.

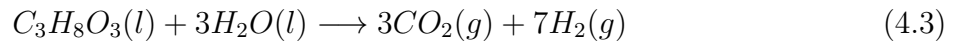
En general, para calcular la entalpía de reacción a una cierta temperatura, se necesita la reacción general balanceada de reformado, sus coeficientes estequiométricos, y, además, la entalpía de formación ( $\Delta H_f$ ) de cada compuesto a la temperatura deseada (ver ec. 4.5), para así, a través de la ec. 4.2, obtener su valor.

De lo anterior se desprende que los datos necesarios son los siguientes: la reacción genérica de reformado, los punto de ebullición y fusión, y el calor específico de cada especie para el estado sólido, líquido y gaseoso, según corresponda.

A continuación se detallará la obtención de la reacción genérica de reformado y, posteriormente, cómo se obtuvieron los datos termodinámicos necesarios.

### ***Reacción***

De las tres maneras de reformar presentadas en la Sección 1.1.2, se escogió el reformado por vapor (*Steam Reforming*, abreviado SR), que utiliza agua como agente reformador y requiere una fuente de calor para que ocurra la reacción.



Para obtener la reacción genérica se utilizó como punto de partida la reacción de reformado para el glicerol [15] (ver ec. 4.3) que permite determinar las especies participantes en ésta. De lo anterior, se obtuvo la reacción general de reformado mostrada en la ec. 4.4, donde  $A$ ,  $B$ ,  $D$  y  $E$  son los coeficientes estequiométricos.

Para determinar el valor de estos coeficientes se realizó un balance por átomo y se consideró que el valor de  $A$  era uno para normalizar los cálculos y comparaciones entre los compuestos



reformables. Para mayor detalle de los cálculos realizados, ver Anexo B.2.

Tabla 4.1: Coeficientes estequiométricos de la ecuación 4.4

Coef. Est.	Valor
$A$	1
$B$	$2n - z$
$D$	$n$
$E$	$2n + \frac{m}{2} - z$

En la Tabla 4.1 se observan los valores obtenidos para los coeficientes. Se debe notar que estos solamente dependen de los valores de  $n$ ,  $m$  y  $z$ , los cuales son propios para cada compuesto. Con esto se podrán determinar los coeficientes estequiométricos a partir del compuesto que se desea reformar.

Para agregar estos coeficientes a la base de datos, se creó una función, *reaction* (ver Anexo E.3), que calcula estos valores en función de lo obtenido anteriormente (ver Tabla 4.1) y los almacena en un vector numérico.

### *Datos termodinámicos*

Debido a que se necesita calcular la entalpía de formación a  $600[^\circ C]$  para cada compuesto (para obtener posteriormente la entalpía de reacción) y que además se está considerando el supuesto de que todas las especies inicialmente se encuentran en el estado de la materia correspondiente a sus condiciones estándar, se debe calcular el calor necesario para llevar cada compuesto desde su estado a  $25[^\circ C]$  (condiciones estándar) hasta el estado correspondiente a  $600[^\circ C]$ .

De lo expuesto en el párrafo anterior, se tienen tres posibles escenarios que se observan en la Figura 4.3 y que serán enunciados a continuación junto a los datos termodinámicos necesarios para cada caso. Los escenarios posibles son:

1. La molécula está en estado gaseoso a condiciones estándar, por lo que se necesitarían los datos de punto de ebullición y calor específico del gas.
2. La molécula está en estado líquido a condiciones estándar. La información necesaria consta del calor específico para el estado líquido y gaseoso, entalpía de vaporización, y puntos de fusión y ebullición.
3. La molécula está en estado sólido a condiciones estándar, por lo que se necesitarían los datos de calor específico para el estado sólido, líquido y gaseoso, entalpías de vaporización y fusión, y puntos de fusión y ebullición

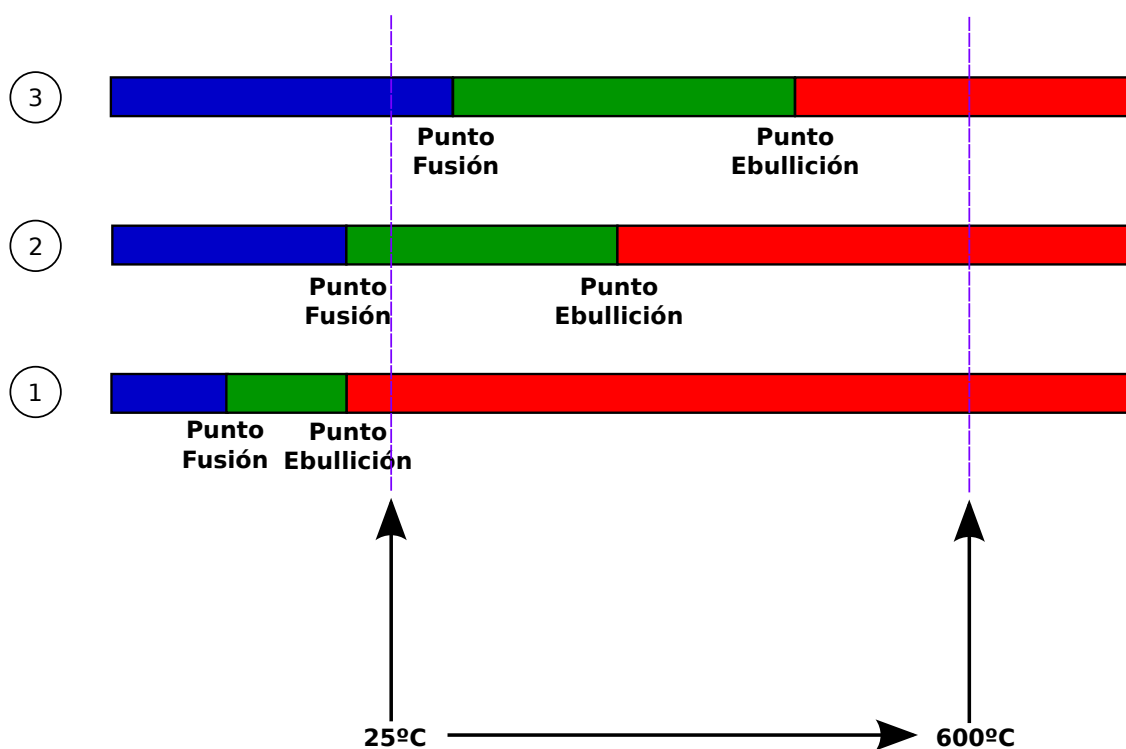


Figura 4.3: Escenarios posibles para las moléculas en la base de datos. Las flechas indican el punto de partida ( $25^{\circ}\text{C}$ ) y el punto final ( $600^{\circ}\text{C}$ ). En rojo, se representa el estado gaseoso del compuesto; en verde, el estado líquido; y en azul, el estado sólido.

Los datos necesarios, mencionados para cada escenario, fueron obtenidos a partir de distintas bases de datos consultadas [53, 55, 56]. Muchos de ellos no se pudieron obtener, ya que no se encontraron en la literatura. Debido a esto la cantidad de compuestos de los cuales se poseía toda información necesaria era muy baja, alrededor de cinco. Ante este evento, se propusieron dos soluciones: la primera, que consistía en cambiar el supuesto tomado (de que las moléculas están en el estado correspondiente a sus condiciones estándar); y la segunda, que considera la estimación de las propiedades faltantes de los compuestos.

Cabe mencionar el calor que se está calculando servirá para obtener la entalpía de reacción y, con ello, poder determinar cuánta energía se debe aportar a la reacción para que ocurra. Así, cualquier distorsión de ésta, hará que las estimaciones posteriores se alejen de la realidad.

Para tomar esta decisión, se realizó un cuadro comparativo de las ventajas y desventajas en cuanto a tomar la primera alternativa (ver Tabla 4.2). De este se obtuvo que si se elegía ingresar todas las especies en estado líquido o gaseoso (diferente al estado natural a condiciones estándar), provocaría que la información se distorsione, ya que gran parte de las especies no se encuentran en estos estados a condiciones estándar (por ejemplo, los compuestos sólidos a  $25^{\circ}\text{C}$  y  $1[\text{atm}]$ ). Esto conllevaría que al calcular la energía para la reacción, el resultado no se acerque a la realidad, ya que para las especies, que se ingresan a diferente estado de la material del natural, no se consideraría la energía que se necesita para elevar la temperatura hasta los  $600^{\circ}\text{C}$  (temperatura en que ocurre la reacción) o la asociada a los cambios de estado que los compuestos deben ser sometidos para que reaccionen (se

necesitan en estado gaseoso). Por otro lado, si se escogiera que las especies ingresen en estado sólido, o en su defecto, a condiciones estándar, el conjunto de moléculas, de los que se poseen todos los datos termodinámicos, sería pequeño. Por estas razones, se desechó esta opción y se optó por la segunda alternativa propuesta, en que los valores faltantes son estimados.

Tabla 4.2: Ventajas y desventajas para los diferentes casos si se cambia supuesto

Supuestos	Ventajas	Desventajas
Todos los compuestos están inicialmente en estado gaseoso y solo se considerarán las moléculas que posean todos los datos termodinámicos pertinentes.	Mayor número de compuestos.	Gran distorsión de datos.
Todos los compuestos que en condiciones estándar se encuentren como sólidos serán considerados como líquidos. Además solo se aceptarán las moléculas que posean los datos necesarios para el cálculo.	Cantidad de compuestos intermedios.	Gran distorsión de datos.
Compuestos inicialmente en su estado natural correspondiente a las condiciones estándar, considerando solamente moléculas que posean los datos necesarios.	Baja distorsión de datos, mayor cercanía a la realidad.	Baja cantidad de compuestos.

Las estimaciones fueron realizadas en base a razones entre valores que se poseían. Por ejemplo se tomó la razón entre el  $c_{P,liq}$  y el  $c_{P,gas}$  para cada especie que se poseían los datos, luego se determinó el promedio entre los valores calculados y se obtuvo así un factor  $F_1$ , que servirá para obtener  $c_{P,liq}$  o  $c_{P,gas}$ , si no se posee uno de ellos para alguna especie.

$$c_{P,liq,i} = c_{P,gas,i} \cdot F_1$$

Los factores obtenidos, que fueron utilizados para estimar los datos faltantes en la base de datos, se muestran en la Tabla 4.3. Se crearon funciones para calcular dichos valores ( $CpgCplCpsDHvDHf$ ,  $DHf\_DHvFactor$ ,  $Cps\_CpgFactor$  y  $Cpl\_CpgFactor$ ) y para agregar los datos a la base de datos. En Anexo E.4 se pueden ver éstas y la función creada para completar la estructura de datos.

Tabla 4.3: Factores promedios para la estimación de datos

Factor	Valor promedio
$F_1 = \left( \frac{c_{P,liq}}{c_{P,gas}} \right)$	10.8
$F_2 = \left( \frac{c_{P,sol}}{c_{P,gas}} \right)$	5.20
$F_3 = \left( \frac{\Delta H_{fusion}}{\Delta H_{vap}} \right)$	0.436

### *Entalpía de reacción*

Una vez obtenidos los datos termodinámicos necesarios, se procedió a calcular la entalpía de formación para cada especie. Para ello, se utilizó la ec. 4.5, donde  $\Delta H_f(600^\circ C)$  y  $\Delta H_f(25^\circ C)$  son las entalpías de formación a  $600[^\circ C]$  y  $25[^\circ C]$ ,  $c_{P,sol}$ ,  $c_{P,liq}$  y  $c_{P,gas}$  son los calores específicos para el estado sólido, líquido y gaseoso, y  $T_{fus}$  y  $T_{eb}$  son las temperaturas de fusión y ebullición respectivamente. Es importante mencionar que si la especie en condiciones estándar está en estado gaseoso, los términos  $c_{P,sol}$ ,  $\Delta H_{fus}$ ,  $c_{P,liq}$  y  $H_{vap}$  serán 0, al igual que para el caso en que la especie inicialmente está en estado líquido, donde  $c_{P,sol}$  y  $\Delta H_{fus}$  tomarán este valor.

$$\Delta H_f(600^\circ C) = \Delta H_f(25^\circ C) + c_{P,sol}(T_{fus} - 25^\circ C) + \Delta H_{fus} + c_{P,liq}(T_{eb} - T_{fus}) + \Delta H_{vap} + c_{P,gas}(600^\circ C - T_{eb}) \quad (4.5)$$

Una vez obtenidas estas entalpías y utilizando la ec. 4.2, los datos estequiométricos de la Tabla 4.1 y la ecuación de la reacción 4.4, se obtuvieron las entalpías de reacción para cada compuesto orgánico de la lista (ver lista de compuestos orgánicos en Anexo B.1). Cabe señalar que además de la entalpía de formación para cada compuesto, se necesitó el valor de este para el agua, el dióxido de carbono y el hidrógeno a  $600[^\circ C]$  los que se obtuvieron de *Perry's Chemical Engineers' Handbook* [58]. Para realizar estos cálculos se utilizaron algunas funciones creadas en MATLAB® ( $DHr$ ,  $DHf_{-600}$ ,  $DHfw$ ,  $DHfH2$  y  $DHfCO2$ ), que permiten, además, agregar esta información a la base de datos (mayor detalle, ver Anexo E.5).

## 4.3. Candidatos a descriptores

- ¿Qué propiedades microscópicas se cree que condicionan cada propiedad macroscópica?
- ¿Qué característica de una molécula, mediante su modificación, provocará un cambio en la propiedad en estudio?

En la Sección 3.3 se plantearon estas interrogantes para poder definir los candidatos a descriptores que están relacionados con la elección de la mejor molécula a reformar. Cabe señalar que el listado de candidatos resultante, fue realizado con la ayuda de un grupo de

colaboradores [38]. Las respuestas obtenidas son mostradas a continuación junto a una breve explicación del criterio utilizado para la elección de éstas.

- **Rendimiento de hidrógeno:**

- **Número de hidrógenos:** Se escogió esta característica, ya que el hidrógeno molecular producido es obtenido a partir de los átomos de hidrógeno presente en la molécula que es reformada.
- **Número de grupos OH:** Se eligió ésta porque el grupo funcional, *OH*, presenta un átomo de hidrógeno y, además, porque los alcoholes (que tienen presente este grupo) son atractivos para ser reformados.
- **Restricciones de acceso al carbono por presencia de OH y H:** Se optó por ésta porque, para que la reacción se lleve a cabo, se deben efectuar una serie de mecanismos de reacción y, así, poder acceder al átomo que tiene unido los átomos de hidrógeno.
- **Ángulos de enlace:** Esta característica fue seleccionada, ya que los ángulos son responsables de la ubicación de los átomos enlazados al carbono (átomo central), por lo que podrían interferir al efectuarse los mecanismos de reacción.

- **Costo de producción:**

- **Número de carbonos:** Se eligió esta característica, ya que sintetizar moléculas más largas es una tarea compleja, por lo que haría cambiar el costo de la materia prima.
- **Número de grupos OH:** Se escogió esta característica porque se presume que el presentar grupos *OH* en la cadena, provocará una variación en el costo de realizar la reacción.
- **Número de hidrógenos:** Se optó por esta característica, ya que la cantidad de átomos de hidrógeno podría hacer variar el costo asociado al procesamiento de las moléculas.
- **Ángulos de enlace:** Esta característica fue seleccionada, como se dijo anteriormente, ya que los ángulos son responsables de la ubicación de los átomos enlazados al carbono, por lo que podría determinar la facilidad de sintetizar las moléculas a reformar y con esto el costo de la materia prima.

- **Emisiones:**

- **Número de carbonos:** Esta característica fue escogida porque el  $CO_2$  producido a través de la reacción, está relacionado con el número de carbonos de la molécula que es reformada, ya que es el único producto que presenta *C* en su composición.
- **Número de grupos OH:** Se optó por ésta porque el grupo hidroxilo tiene *O* en su composición y el único producto de la reacción que presenta este átomo es el dióxido de carbono, el que es considerado como emisión.
- **Ángulos de enlace:** Esta característica fue seleccionada, ya que, como es responsable en parte de la ubicación de los átomos enlazados al carbono, podría interferir al efectuarse los mecanismos de reacción y así en la obtención de  $CO_2$ , ya sea facilitando o dificultando su generación.

De los candidatos a descriptores anteriores, se puede apreciar que la mayoría de ellos se repiten para cada propiedad, por lo que se realizó una lista general de candidatos. Además,

debido a la dificultad para poder obtener algunos de estos descriptores moleculares, fueron filtrados. La accesibilidad al carbono por presencia de  $OH$  y  $H$  fue uno de ellos, ya que para obtener esta propiedad, se necesitaría saber el volumen de cada átomo, tener una representación espacial y, además, determinar una manera de cuantificar la *accesibilidad al carbono* de forma sencilla y rápida a través de estos datos. Por estas razones, se decidió descartar este candidato a descriptor y dejarlo como trabajo futuro.

En cuanto al ángulo de enlace, descriptor molecular que se repite en las tres propiedades macroscópicas, fue filtrada, ya que, como el caso anterior, los datos para su cálculo resultaban ser muy complejos de obtener porque, para cuantificar de forma fácil y rápida los ángulos asociados a cada compuesto, se necesita tener una representación espacial de las moléculas y saber qué átomos están enlazados entre sí.

Por otro lado, el número de carbonos, hidrógenos y grupos  $OH$ , no fueron descartados, ya que estas propiedades moleculares eran fáciles de obtener y de cuantificar. Para obtener el número de carbonos e hidrógenos, basta con tener la fórmula química de cada compuesto (por ejemplo,  $CH_4$ ), y de aquí obtener el número total de  $C$  y  $H$ . En cuanto al grupo  $OH$  en los alcoholes (por ejemplo,  $CH_4O$ ), se cuenta adicionalmente el número de oxígenos presentes y se resta este valor a los hidrógenos (el grupo hidroxilo tienen  $H$  en su composición), para determinar, así, el número total de átomos de cada especie.

Cabe mencionar que, si bien, se podrían haber definido otros descriptores que explican el comportamiento de las propiedades macroscópicas (como calor específico, entalpía de reacción, proporción de productos, entre otros), éstas fueron descartadas por el tamaño de *piezas* que se seleccionó (se hablará en la Sección 4.4 sobre esta elección), ya que hablar a nivel atómico, por ejemplo, de entalpía de reacción, no tendría sentido, como lo es hablar del número de átomos de hidrógeno presentes en cada compuesto, donde sí se podría cuantificar a través de las *piezas* elegidas.

A continuación, se muestra la lista final de candidatos a descriptores moleculares.

- Número de carbonos.
- Número de hidrógenos.
- Número de grupos  $OH$ .

Cabe notar que se seleccionó el número de grupos  $OH$  y no de  $O$ , ya que los compuestos que presentan oxígeno en su composición (según la base de datos seleccionada) son solamente alcoholes y son justamente estos los que se desea estudiar su comportamiento (por presencia de grupos  $OH$ ).

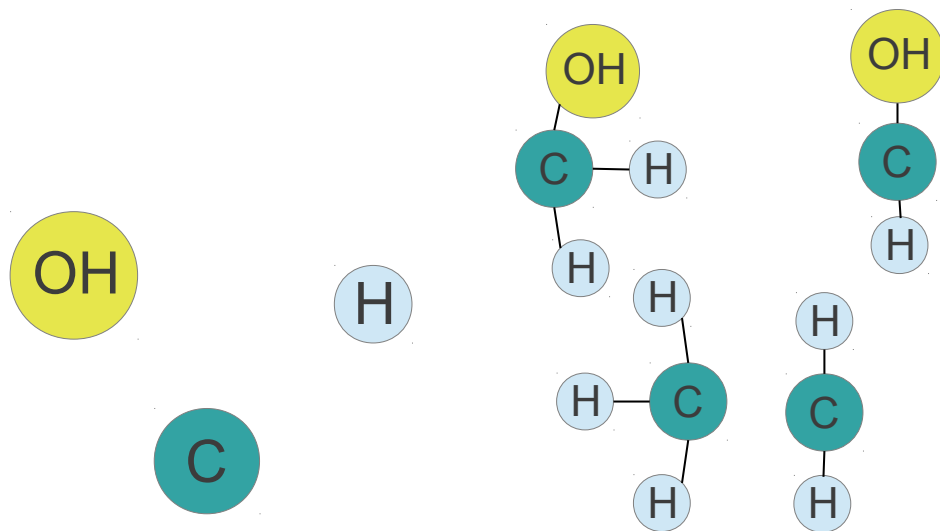
Los datos de los descriptores definidos fueron almacenados en una matriz, en que las filas representan los compuestos; y las columnas, los candidatos a descriptores. Se decidió agregar una columna con el número de registro, CAS, para poder identificarlos fácilmente. Para llenar esta matriz se utilizó una función en MATLAB® la que recopila los datos de la base de datos creada anteriormente. Esta función se puede ver en Anexo E.6.3. Además en la Tabla C.1 de Anexo C.1 se observa dicha matriz y los datos almacenados.

## 4.4. Piezas

Se definieron los siguientes tamaños de piezas para el problema:

1. **Atómica:** Este tamaño considera a los átomos como piezas. Al unirse entre sí, formarán una estructura molecular (molécula). Para el presente problema, dado que en la base de datos solo existen compuestos con *C*, *H* y *O*, las piezas corresponderían a tales átomos (ver Figura 4.4a). Cabe señalar que el oxígeno, por encontrarse siempre en un grupo *OH*, será representado como una pieza para tal efecto.
2. **Grupo de átomos:** Las piezas para este caso, considera a grupos de átomos de carbono, oxígeno e hidrógeno (ver Figura 4.4b).

Cabe mencionar que si bien se pensó en otros tamaños de piezas, como *grupo funcional*, estos no fueron considerados posteriormente, ya que no lograban el objetivo de representar fácilmente las moléculas de la base de datos que se construyó porque se tenían que modificar los grupos de *piezas*, ya sea agregado nuevas o alterando las ya consideradas. Por ejemplo para el *grupo funcional*, las *piezas* eran *CH* y *OH*, pero con ellas no se podían armar estructuras moleculares factibles, ya que faltaba una pieza de *H* para completar los enlaces vacíos. Ahora bien, si se agregaba esta pieza, se obtendría un grupo similar al *atómico*.



(a) Piezas atómicas.

(b) Piezas por grupo de átomos.

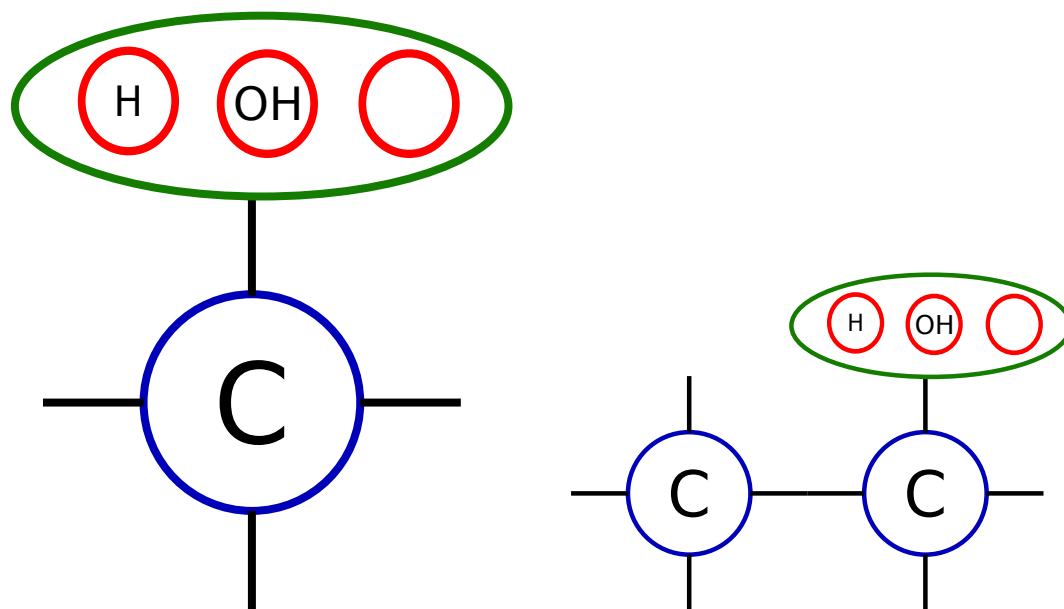
Figura 4.4: Tamaños de piezas propuestos.

Para decidir cuál era el mejor tamaño de pieza se tomaron los criterios de selección propuestos en la Sección 3.5, donde el primer parámetro medido era la representatividad de las moléculas de la base de datos. Con ello se obtuvo que ambas opciones lograban dicho

objetivo<sup>2</sup>. Dado lo anterior, se procedió a medir el nivel de *ruido* (ver ec. 3.1) que se genera al armar todos los compuestos de la base de datos a través de las piezas definidas. Los cálculos y consideraciones que se tuvieron para ello serán explicadas a continuación.

#### 4.4.1. Cardinalidad del universo de moléculas representables cuando el tamaño de piezas es atómico

Para determinar la cardinalidad del conjunto, primero se calculó el número total de estructuras moleculares que se puede formar con las piezas definidas en base a las consideraciones expuestas en la Sección 3.5. Para ello, se determinaron los posibles candidatos que se pueden unir a los enlaces del carbono, definiendo, así, tres alternativas para cada enlace libre: no unir átomo alguno, unir un átomo de hidrógeno o uno de oxígeno (ver Figura 4.5a). Traduciendo esto a combinatoria, se llega a que para una cadena lineal de tamaño uno, se tienen  $3^4$  estructuras moleculares que se pueden formar.



(a) Cadena lineal de tamaño 1 carbono. Aquí se muestran los posibles candidatos para rellenar uno de los enlaces disponibles.

(b) Cadena lineal de tamaño 2 carbonos. Aquí se muestran los posibles candidatos para rellenar uno de los enlaces disponibles.

Figura 4.5: Cardinalidad de moléculas para caso de 1 carbono y 2 carbonos.

Hay que dejar claro que se consideró la alternativa de no unir átomo alguno al carbono, ya que, si bien esto puede generar estructuras moleculares infactibles fisicoquímicamente, estas serán filtradas o se establecerán restricciones al momento de construir las moléculas factibles en una etapa posterior (ver etapas de CAMD 2.1). Por lo anterior, en esta fase solo se están

<sup>2</sup>El tamaño de piezas *grupo de átomos* no representa dos compuestos de la base de datos, metano y metanol, pero se adicionaron como se explicará más adelante.



contando las estructuras totales que se podrían armar con las *piezas* para poder cuantificar el *ruido* asociado a cada tamaño, sin considerar las restricciones que se deberán imponer para generar estructuras factibles. Así, si minimiza el *ruido*, se tendrá un conjunto afín con el planteamiento y operación del problema.

Ahora, si se considera una cadena de dos carbonos se tendrá el caso de la Figura 4.5b, donde habrán seis enlaces disponibles y por cada uno tres opciones de candidatos, obteniendo con esto,  $3^6$  estructuras moleculares que se pueden generar con las *piezas*.

Generalizando, si se toman  $n$  carbonos en la cadena principal se tendrán  $2(n + 1)$  enlaces totales, por lo que las posibles estructuras que se pueden formar son  $3^{2(n+1)}$  (ver Figura 4.6).

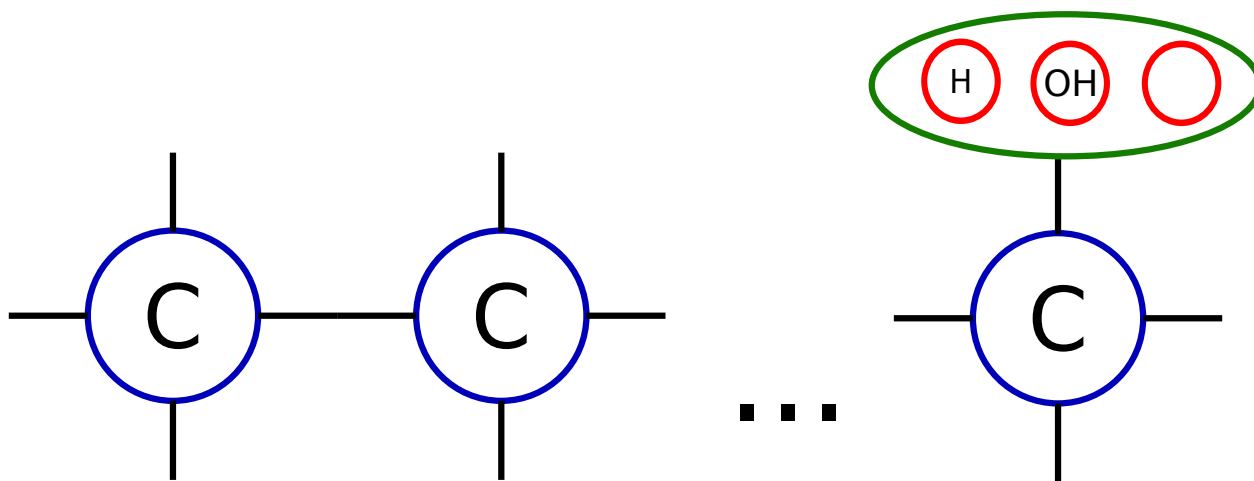


Figura 4.6: Cadena lineal de  $n$  carbonos, y los posibles candidatos para completar uno de sus enlaces.

Cabe mencionar que, tal y como se explica en la Sección 3.2.2, para los alcoholes el valor máximo de  $n$  es de 6, mientras que para los alcanos este valor es de 16. Por otro lado, los candidatos que se pueden unir a los enlaces del carbono son 3 para los alcoholes y 2 para alcanos, con lo que la fórmula generalizada para obtener la cantidad de compuestos dado una cierta cantidad de carbonos es:  $3^{2(1+n)}$ , si  $n \leq 6$  y es  $2^{2(1+n)}$ , si  $n > 6$ <sup>3</sup>.

Finalmente, calculando la cardinalidad para el conjunto completo, se obtiene lo siguiente:

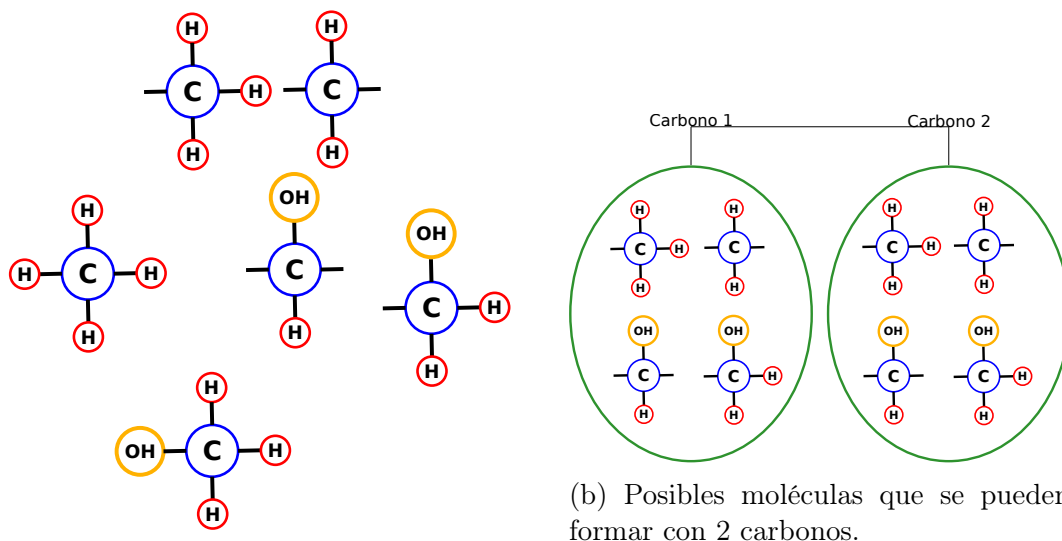
$$\#G1 = \sum_{i=1}^6 3^{2(1+i)} + \sum_{i=7}^{16} 2^{2(1+i)} = 5.380.830 + 22.906.470.400 = 22.911.851.230 \quad (4.6)$$

<sup>3</sup>Cuando  $n \leq 6$  se están contando la cantidad de estructuras moleculares posibles tanto para alcanos como para alcoholes, mientras que para  $n > 6$  solo se consideran los alcanos debido a los largos máximos de cadena.

#### 4.4.2. Cardinalidad del universo de moléculas representables cuando el tamaño de piezas corresponde a grupo de átomos

Al igual que para calcular la cardinalidad del anterior universo de moléculas representables, primero se comenzó contando la cantidad de estructuras moleculares que se pueden formar con un carbono. Si bien este conjunto es diferente, ya que son piezas más grandes (grupos de átomos), se consideraron para este caso todas las piezas por sí solas (ya que todas tenían solamente un carbono). Cabe notar que a través de estas piezas (ver Figura 4.4b) no se pueden formar metano ni metanol, por lo que la base de datos no estaría completa. Por esta razón, se decidió agregar un par de piezas adicionales que permitieran representar estas dos moléculas, resultando, finalmente, un total de seis piezas (ver Figura 4.7a). Cabe notar que, al igual que para el grupo atómico, se consideró la opción de que los carbonos no tengan unido algún átomo a sus enlaces libres, ya que se está cuantificando el *ruido* que se genera a través de este tamaño de piezas, dejando para la siguiente etapa, la tarea de filtrar o restringir las moléculas que se construirán.

Si se considera un largo de dos carbonos se tendrán cuatro alternativas de piezas por cada carbono (ver Figura 4.7b). No se consideraron las piezas que representan al metano y metanol, ya que debido a las indicaciones de armado (ver Sección 3.5), solo se pueden unir piezas que tengan un enlace libre y para éstas, los cuatro enlaces del carbono están ocupados. Por lo anterior, la cantidad de estructuras moleculares posibles será  $4^2$ .



(a) Piezas para el grupo molecular incluyendo las 2 nuevas que se mencionan (metano y metanol).

(b) Posibles moléculas que se pueden formar con 2 carbonos.

Figura 4.7: Piezas y cardinalidad para piezas de *grupo atómico*.

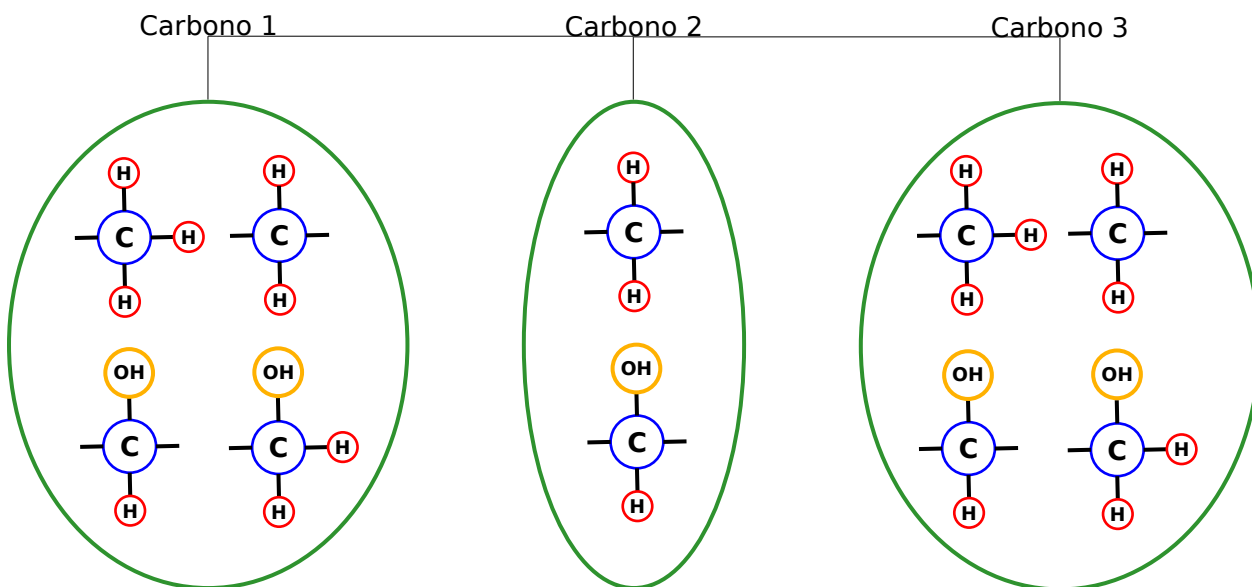


Figura 4.8: Posibles moléculas que se pueden formar con 3 carbonos.

Ahora, si se calcula la cantidad de estructuras que se forman con tres carbonos, se tendrá que para los extremos de la cadena lineal se tienen 4 opciones de adicionar piezas, mientras que para el carbono de entre medio, solo tendrá 2 alternativas como se puede apreciar en la Figura 4.8. Esto es debido a las indicaciones de armado, en que al unir dos carbonos, estos deben tener un enlace libre para poder hacerlo (no superar el máximo de cuatro enlaces), y para que una pieza pueda estar en el centro, debe tener al menos 2 enlaces libres, lo cual solo lo cumplen 2 de las piezas.

Generalizando el cálculo de la cantidad de estructuras que se puede formar con  $n$  carbonos, se tiene que son  $2^{(n-2)} + 4^2$  si  $3 \leq n \leq 6$  (para alcanos y alcoholes) y 5 si  $n > 6$  (para alcanos solamente). Hay que destacar que esta última fórmula no incluye el largo de cadena  $n$ , ya que el número de estructuras moleculares que se pueden formar depende de los extremos porque entre medio solo podrá ir una única pieza (ver Figura 4.9).

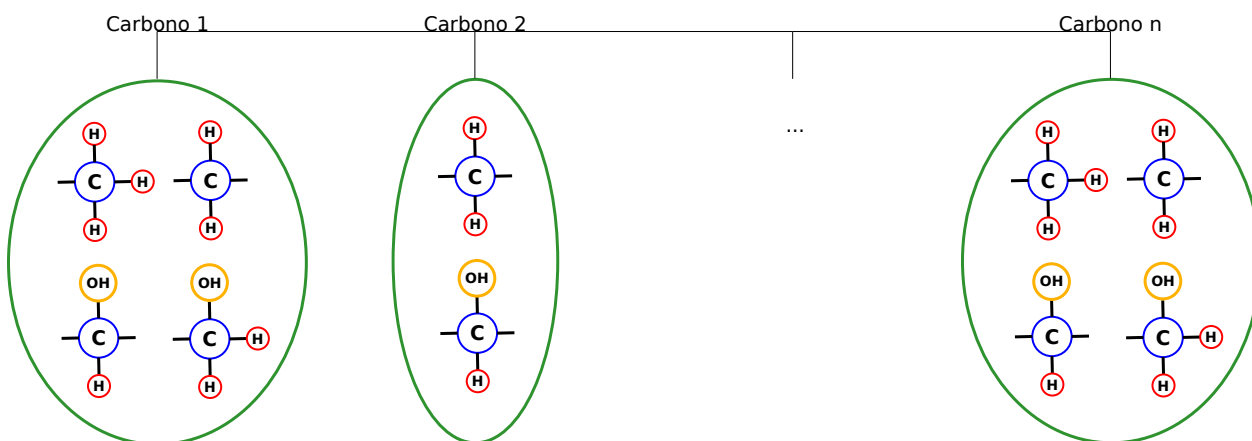


Figura 4.9: Posibles moléculas que se pueden formar con  $n$  carbonos.

De lo expuesto anteriormente, y considerando que se tomó el caso separado para alcanos y alcoholes (a diferencia con el caso anterior, que se tomó el caso en conjunto), se obtiene que la cardinalidad total de moléculas que se pueden formar es:

$$\#G2 = 6 + 4^2 + \sum_{i=3}^6 (2^{(i-2)} + 4) + \sum_{i=7}^{16} 5 = 166 \quad (4.7)$$

### 4.4.3. Selección de piezas

Como se mencionó, la elección del conjunto de piezas es una tarea fundamental para CAMPD, ya que a partir de ellas se podrá construir, en una etapa posterior, la o las moléculas que cumplan con las propiedades objetivo. Para seleccionar un conjunto de piezas se utilizaron los criterios vistos en la Sección 3.5. Con el primero de ellos, se obtuvo que ambos grupos definidos cumplen con representar todos los componentes de la base de datos, mientras que con el segundo, que contempla el cálculo del ruido a través de la ec. 3.1, resultó que las piezas por *grupo de átomos* generan un menor ruido, con una diferencia del orden de  $10^8$  aproximadamente (ver Tabla 4.4).

Tabla 4.4: Resumen de parámetros cuantificados para la selección de piezas.

	Representación	#Ruido
G1	✓	22.911.851.204
G2	✓	140

A partir de esto y la tabla anterior, finalmente se obtiene que el tamaño elegido es el de *grupo de átomos*, ya que es capaz de representar la totalidad de los compuestos que están en el conjunto de moléculas a analizar y, a la vez, generar un bajo nivel de ruido (del orden de  $10^2$  estructuras inafectibles). A través de esto se disminuirá considerablemente el costo computacional que se efectuará al filtrar o establecer restricciones en el armado de estructuras a futuro (al realizar CAMPD) en comparación con el grupo atómico, que genera un ruido del del orden de  $10^{10}$  estructuras inafectibles fisicoquímicamente.

Las gran desventajas de este tamaño de piezas, es que no da la posibilidad de expandir el grupo de moléculas a analizar, como por ejemplo si se consideraran ramificaciones o compuestos con dobles o triples enlaces, ya que la forma de las piezas (enlaces disponibles) no lo permiten (ver Figura 4.7a). Una solución a este problema es agregar nuevas piezas a este conjunto como se hizo para que pudieran representar el metanol y el metano, aunque implicaría la generación de una mayor cantidad de ruido, por lo que se debería realizar una comparación entre ruido generado y moléculas representadas. Otra solución podría ser el generar inteligentemente nuevas piezas que cubran el nuevo conjunto.

Como se mencionó, el tamaño de *piezas* escogido para el conjunto de moléculas a analizar es el de grupos atómicos, si bien esta elección implicaría que no se puedan adicionar *expansiones* al conjunto, se puede destacar que se desarrollo una estrategia para poder determinar el mejor

tamaño a través del nivel de *ruido* y la representatividad del conjunto de compuesto, por lo que a futuro, si se llegaran a definir nuevas *piezas*, se podrá obtener qué tamaños serán los mejores candidatos.

#### 4.4.4. Datos y supuestos

A continuación se discutirán alguno de los supuestos y datos más importantes que fueron considerados en las secciones anteriores.

#### Propiedades macroscópicas

La elección de las propiedades macroscópicas fue realizada según las características que se deberían obtener al realizar un proceso óptimo de reformado, por lo que se abarcaron tópicos como costos, rendimiento y emisiones. La elección de realizar o no cualquier tipo de proyecto (cualquier escala) no solo va en torno a estos tres aspectos mencionados, si no que también en cuanto al tema social, a través de una evaluación social o alguna herramienta de evaluación cualitativa como el HAIN [59], etc., por lo que se deja abierta la posibilidad de expandir las propiedades macroscópicas objetivo involucrando estas variables de salida.

Las estimaciones de las propiedades macroscópicas, que sirvieron posteriormente para realizar el análisis estadístico, no fueron totalmente representativas, ya que por ejemplo, el valor utilizado para el costo, fue el precio del compuesto orgánico que se iba a reformar, siendo éste solo una parte del total. Por esta razón podrían no obtenerse algunas relaciones con las propiedades microscópicas (candidatos a descriptores), ya que pueden haberse subestimado al no considerar los costos por separación o algún proceso adicional, ya que estos podrían marcar una gran diferencia entre cada molécula y, así, implicar una relación con los candidatos a descriptores.

Cabe mencionar que, como se discutirá más adelante, las emisiones calculadas solo corresponden a valores parciales ( $CO_2$  generado por el reformado y  $CO_2$  asociado a la combustión que provee energía para la reacción), ya que solo consideran parte de las emisiones totales que se pueden generar en un proceso, por lo que no se podría asegurar que las relaciones obtenidas para esta propiedad macroscópica sean totalmente acertadas. Sin embargo, se debe rescatar la realización de la estrategia que se hizo para el análisis de significancia estadístico (ver Sección 3.6), ya que así, solamente se deberán modificar los datos y aplicar ésta para determinar la existencia total de relaciones entre candidatos a descriptores y propiedades macroscópicas.

#### Temperatura de reformado

Se escogió la temperatura de reformado a  $600^\circ C$ , ya que, como se explicó anteriormente (ver Sección ??), existen varios registros científicos en que usan esta temperatura para

reformar. El hecho de escoger esta temperatura, provoca que las necesidades energéticas (específicamente la entalpía de reacción) sean distintas en comparación con si se escogiera una menor, ya que se necesita llevar todos los compuestos participantes en la reacción, desde los  $25^{\circ}\text{C}$  hasta los  $600^{\circ}\text{C}$ . Por otra parte, para que ocurra la reacción, se necesita que las especies a esta temperatura estén en estado gaseoso para que puedan ser reformadas, pero no todas ellas a condiciones estándar están en este estado, si no que solo una parte (el resto está en estado líquido o sólido). Ahora bien, esta diferencia de entalpía de reacción que existirá entre ambos casos (entre reformar a  $600^{\circ}\text{C}$  o a otra temperatura), se presume que sería bajo, ya que en los dos escenarios, las especies que no se encuentran en estado gaseoso, deberán ser sometidos a cambios de estado, el que, normalmente, se le adjudica el mayor requerimiento energético para su realización. Es importante notar que, a pesar de que se fijó una temperatura de reformado, se deja abierta la posibilidad de encontrar una óptima.

## Reacción

Con respecto al producto de la reacción, se supuso que las únicas especies que se obtenían eran  $\text{CO}_2$  y  $\text{H}_2$ , aproximación que en la realidad no es tan cierta, ya que además, dependiendo de las condiciones de operación, se podría obtener coque,  $\text{CO}$  u otro elemento intermedio. No se consideraron para la realización de este trabajo, ya que no se quería complejizar aun más el problema involucrando reacciones intermedias o algún mecanismo de reacción, aunque se debe mencionar que esto no influiría en el resultado final, ya que si se ajustan estas condiciones de operación se podría llegar cerca de lo planteado.

## Entalpía de reacción

La energía de reacción fue una de las propiedades que limitó el conjunto de moléculas finales a analizar, ya que no se poseían todos los datos necesarios para su cálculo para cada una de las moléculas en la base de datos, y aún con las estimaciones de algunos de ellos, no se logró abarcar la totalidad de los compuestos reformables encontrados. Este hecho provocó la generación de un conjunto pequeño de individuos al momento de realizar el análisis estadístico entre las propiedades microscópicas y macroscópicas y, con ello, no tener más casos de estudio para comprobar la existencia de alguna relación entre las propiedades mencionadas.

Por otro lado, el hecho de que la entalpía de reacción fuera calculada a través de los datos estimados de  $c_P$ ,  $\Delta H_{fus}$ ,  $\Delta H_v$  y  $\Delta H_r$ , provocó que no se pudiera asegurar que su valor sea el correcto, como se mencionó anteriormente, por lo que las emisiones obtenidas podrían no ser acertadas.

## Reformado de vapor de agua

El supuesto tomado con relación a la utilización de  $SR$  como manera para reformar, se traduce en el aumento de las necesidades energéticas ( $\Delta H_r$ ) y, con ello, en las emisiones que

se generan (propiedad que se quiere optimizar). Si se cambiara a otra forma de reformar (ver Sección 1.1.2) se obtendría una cantidad menor de hidrógeno, ya que las demás tecnologías utilizan este último para generar el calor necesario para que ocurra la reacción. Es por eso que se decidió comenzar, en una primera aproximación, con *SR*, donde el calor se obtiene de una fuente externa, pero se deja abierta la posibilidad de cambiar la tecnología según como se estime conveniente.

## Sobre las estimaciones realizadas

Como se explicó, la estimación de los datos termodinámicos para los compuestos fueron realizadas debido a que sin ellos, el número de elementos del conjunto de moléculas a analizar habría sido muy baja (alrededor de 5), por lo que su estimación se hizo necesaria. Esta fue hecha a través de la tendencia de las propiedades y el uso de razones entre los promedios de los datos que se poseían, por lo que no se podría asegurar la existencia real de alguna relación entre las propiedades macroscópicas y los candidatos a descriptores, ya que no son los valores reales. Sin embargo, a través de esto, se pudo desarrollar una estrategia para realizar el estudio de significancia estadístico entre las propiedades macroscópicas y microscópicas. Por lo que a futuro, cuando se tengan los valores correctos de los datos estimados, solo bastará con modificar estos en la base de datos para poder determinar la existencia de alguna relación.

## 4.5. Significancia estadística

### 4.5.1. Análisis cuantitativo y cualitativo

A través del análisis cuantitativo y cualitativo, como se explicó en la Sección 3.6, se espera establecer la existencia de relación entre los candidatos a descriptores (variables de entrada) y las propiedades macroscópicas (variables de salida); para así determinar los descriptores moleculares que son estadísticamente significativos para el problema de la elección de la mejor molécula a reformar. Además, a partir de los criterios definidos en la Sección 3.6.2 y los resultados de este estudio, se seleccionarán las propiedades a las que se le aplicará el análisis de componentes principales.

Como se mencionó en la Sección 3.6.1, para realizar el análisis cuantitativo, se utilizó el coeficiente de correlación de Spearman para determinar el grado de relación lineal entre cada uno de los candidatos a descriptores y las propiedades macroscópicas definidas. Se debe recordar, que si bien algunos de estos valores podría indicar que no existe relación entre las variables (cercano a 0), no se puede descartar tajantemente que exista alguna dependencia entre ellas, ya que podría ser que ésta fuera no lineal.

Para obtener los valores de los coeficientes de correlación y el *p-value* asociado, se creó una función en MATLAB®, *correlationSpearman*, que calcula estos y los guarda en una matriz, donde la celda (*i, j*) representa el valor del coeficiente y del *p-value* para la propiedad

macroscópica  $j$  y el candidato a descriptor  $i$  (ver Figura 4.10). Para mayor detalle, ver en Anexo E.7.

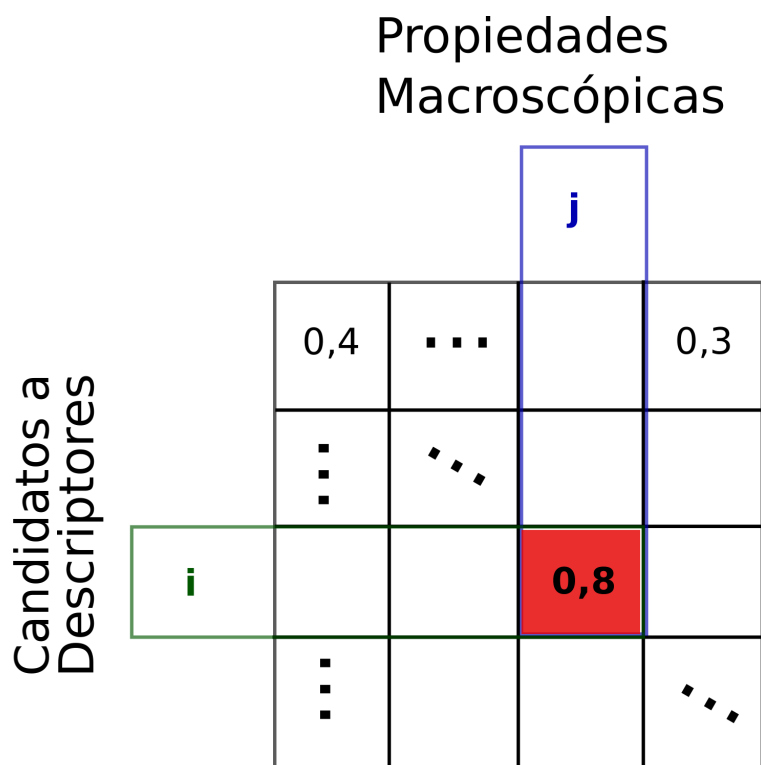


Figura 4.10: Esquema de la matriz que almacena los valores del coeficiente de correlación y  $p$ -value entre descriptores y propiedades macroscópicas. El valor de la celda  $(i, j)$ , que tiene valor 0,8 (para el ejemplo), indica que el coeficiente de correlación o el  $p$ -value (dependiendo del caso), asociado al candidato a descriptor  $i$  y propiedad macroscópica  $j$ , obtiene dicho valor.

Los coeficientes obtenidos y el  $p$ -value asociado a cada uno de estos valores se pueden observar en las Tablas 4.5 y 4.6 respectivamente, donde las columnas representan a las propiedades macroscópicas (rendimiento de hidrógeno, emisiones y costos), las filas a los candidatos a descriptores (número de carbonos, número de hidrógenos y número de grupos  $OH$ ) y las entradas de la matriz, dichos valores. Se espera que los valores obtenidos para los coeficientes de correlación sean cercanos a 1 en módulo, ya que indicaría que existe una relación lineal entre las variables. En cuanto al  $p$ -value, se espera que sean inferiores a 0,05, para poder decir que los coeficientes de correlación son confiables.

Tabla 4.5: Valores obtenidos para el coeficiente de correlación en el estudio de relación entre propiedades macroscópicas y candidatos a descriptores. Un valor cercano a 1 en módulo significará que existe relación lineal marcada entre las dos variables.

	Rend. de $H_2$	Emisiones	Costos
N° de carbonos	0,995	0,996	0,442
N° de hidrógenos	0,999	0,998	0,470
N° de grupos $OH$	-0,621	-0,614	-0,398



Tabla 4.6: Valores obtenidos para cada test de significancia entre propiedades macroscópicas y cada uno de los candidatos a descriptores (*p-value*). Un valor inferior a  $5 \times 10^{-2}$  indica que se puede confiar en que el valor para el coeficiente de correlación asociado es correcto.

	Rend. de $H_2$	Emisiones	Costos
N° de carbonos	2,48e-26	3,03e-27	2,38e-2
N° de hidrógenos	6,96e-36	4,10e-31	1,52e-2
N° de grupos <i>OH</i>	6,96e-4	8,37e-4	4,38e-2

Por otro lado, como se mencionó en la Sección 3.6.1, para determinar la existencia de alguna relación no lineal entre las variables, se realizaron gráficos entre las propiedades macroscópicas y los candidatos a descriptores. Un ejemplo de estos gráficos se puede ver en la Figura 4.11, donde se muestra la dependencia del rendimiento de hidrógeno (eje y) con respecto al número de hidrógenos de la molécula a reformar (eje x). A partir de este, se espera que exista una relación lineal, ya que la producción de  $H_2$  está ligada directamente con el número de hidrógenos que se puedan obtener de cada compuesto. El resto de los gráficos se pueden ver en Anexo D.1.

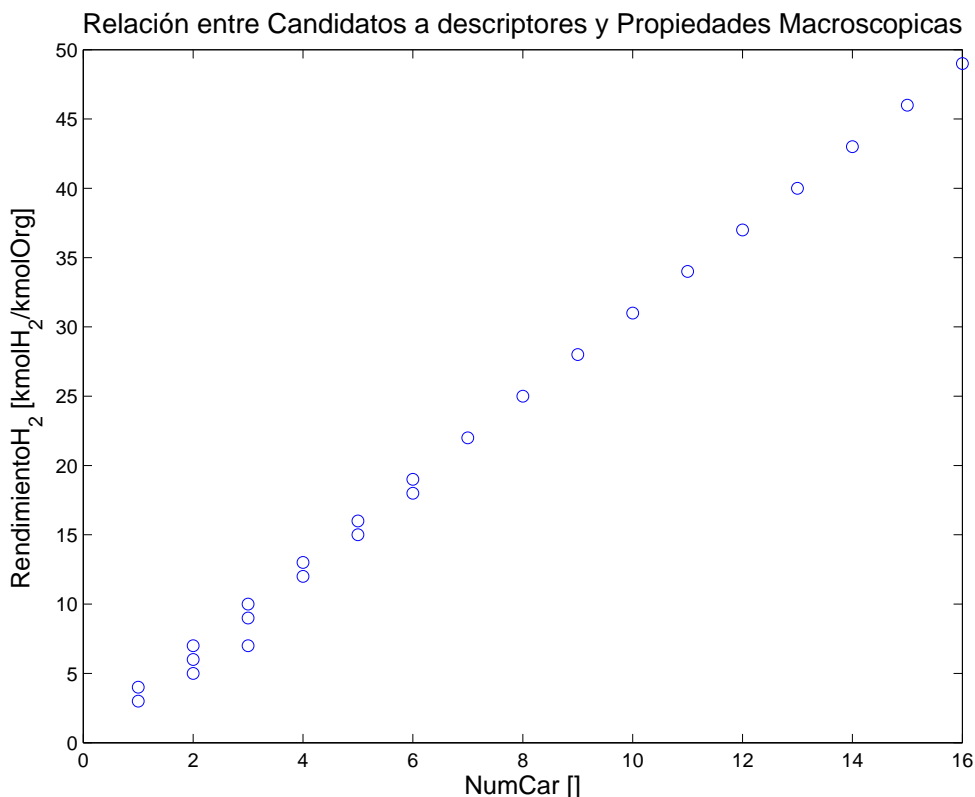


Figura 4.11: Gráfico de relación entre Número de hidrógenos y Rendimiento de hidrógeno.

Para facilitar la obtención de estos gráficos, se construyó una función en MATLAB® llamada *GrafCuali*, la que guarda en un archivo de imagen cada gráfico generado (más detalle en Anexo E.8).

De la tabla de *p-value* se aprecia que la totalidad de los candidatos a descriptores presentan un valor menor a 0,05 (95% de significancia estadística) tanto para el rendimiento de hidrógeno como las emisiones y el costo del proceso. Por esta razón se puede decir que los resultados obtenidos a través del coeficiente de correlación indican correctamente el grado de dependencia lineal entre las variables analizadas.

Si se observa la tabla de coeficientes de correlación, se aprecia que los valores obtenidos tanto para el rendimiento como para las emisiones son muy similares, encontrando una diferencia entre ellos del 0,5%. En cambio, para el costo de producción, se obtuvo una diferencia cercana al 52% con respecto a las otras dos. Ahora bien, si se grafica el rendimiento y las emisiones (ver Figura 4.12) se observa que efectivamente existe una relación lineal muy marcada entre ambas, observándose una pendiente aproximadamente de  $m = 2$  y  $n = 0^4$ , lo que indica que estas propiedades se comportarán similarmente. Esta relación existente se debe a que ambas propiedades son estimadas en base a la estequiometría de la reacción (ver Sección 4.1), específicamente, al número de carbonos del compuesto, donde únicamente la segunda (las emisiones) adiciona un término relacionado con las emisiones generadas por las necesidades energéticas, que si bien no está ligado directamente con la estequiometría, pero utiliza ésta indirectamente para su cálculo a través de la entalpía de reacción (ver ec. 4.1 y 4.2).

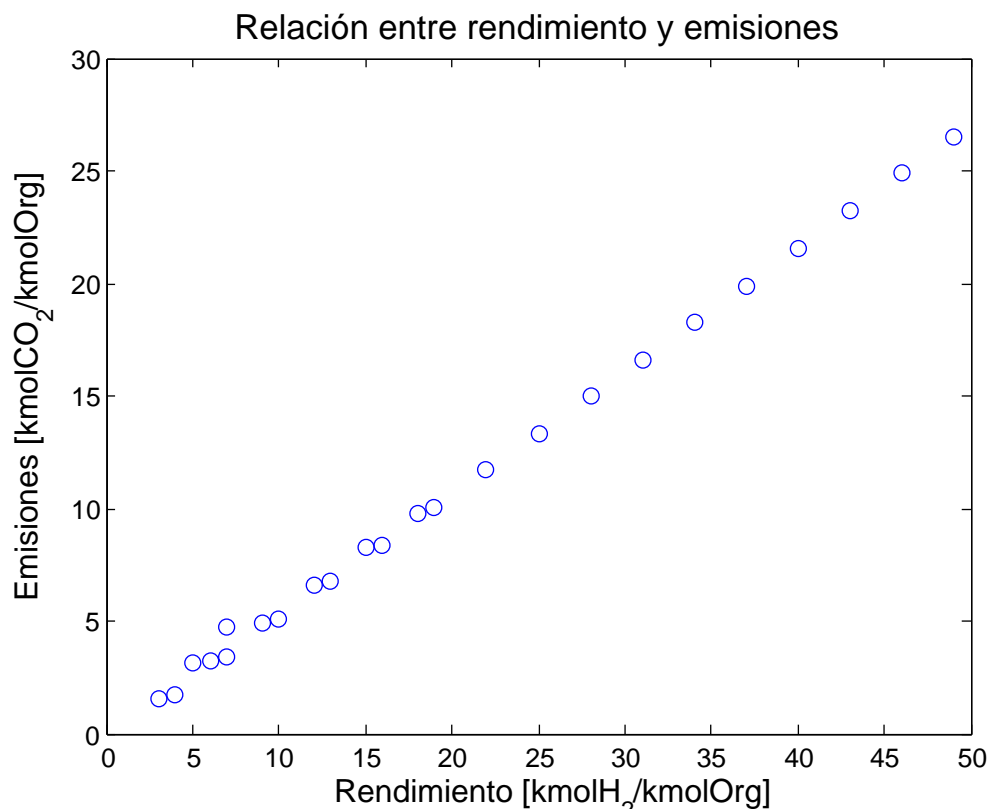


Figura 4.12: Dependencia entre rendimiento de hidrógeno y emisiones.

<sup>4</sup>Ecuación de la recta  $y = mx + n$ , en que  $m$  es la pendiente y  $n$  es el valor de  $y$  donde la recta cruza al eje de las abscisas.

También, se observa que los candidatos a descriptores que poseen una mayor dependencia lineal con respecto al rendimiento de hidrógeno y a las emisiones son el número de carbonos e hidrógenos de cada compuesto. Estas propiedades macroscópicas, como se explicó anteriormente, son calculadas en base a la estequiometría y, más específicamente, al número de carbonos, hidrógenos y oxígenos de cada compuesto (ver Tabla 4.1), razón por la que existe una alta relación con respecto a estas variables (coeficiente de correlación cercano a 1). Si bien, el valor del coeficiente de correlación para el número de grupos *OH* debería ser cercano a 1, ya que también participa en la estequiometría (ver Tabla 4.1) al igual que el número de hidrógenos y de carbonos, este fue cercano al 0,6 para ambas propiedades macroscópicas, lo que indica que existe una dependencia lineal, pero no es tan marcada como las otras dos. Esto se debe a que no se posee una variedad de compuestos (en el conjunto de moléculas a analizar) que posean más de tres grupos *OH*, por lo que la intensidad de la relación no se puede comprobar con seguridad y, por ende, es baja.

Si se observan los gráficos correspondientes a las relaciones mencionadas anteriormente (ver Figuras D.1, D.2, D.4 y D.5), se puede comprobar que los valores obtenidos del análisis cuantitativo (coeficientes de correlación), son coherente a los comportamientos observados de los datos, ya que al aumentar el valor del número de carbonos o hidrógenos, tanto el rendimiento como las emisiones aumentarán de forma proporcional, lo que da cuenta de que tiene un comportamiento lineal.

En cuanto al número de grupos *OH* del compuesto (ver Figuras D.3 y D.6), se observa que, para ambos casos (rendimiento y emisiones), los puntos están concentrados en torno a ciertos valores puntuales. En este caso, la dependencia no es clara, debido a que para un mismo número de carbonos o hidrógenos, se tienen diferentes valores para el rendimiento y las emisiones. Esto se debió a al variar el número de grupos *OH*, no se asegura la variación del largo de cadena, por lo que para un cierto valor de este candidato a descriptor, tendrá asociado más de un resultado en cuanto a emisiones y rendimiento de hidrógeno. Ahora bien, si se consideran solamente los últimos datos, es decir, para un número de grupos *OH* de 2 y 3, se deduce que existe una relación creciente, ya que al aumentar el valor de esta propiedad, aumenta el valor de ambas propiedades macroscópicas. Sin embargo, no se puede decir nada con respecto al tipo de tendencia que se observa, ya que solo se están analizando dos puntos.

Analizando los coeficientes de correlación obtenidos para el costo de producción, los descriptores que obtuvieron un valor mayor a 0,35 son el número de carbonos, el número de hidrógenos y el número de grupos *OH*. Cabe destacar que estos valores no son tan altos como los obtenidos para las demás propiedades macroscópicas, ya que el máximo observado es de 0,442 a diferencia de las otras dos que incluso obtienen un valor de 0,999. Los valores obtenidos, tanto para el número de carbonos como de hidrógeno se deben a que mientras más larga la molécula a reformar (mayor número de carbonos e hidrógenos), más complejo será obtener dicho compuesto, ya que la síntesis será más compleja y, con ello, su precio se elevará.

Para el número de grupos *OH* en el compuesto, se aprecia, a partir del coeficiente de correlación, que existe una dependencia negativa y baja ( $\sim -0,4$ ), lo que indicaría que mientras más grupos hidroxilos tenga la molécula, menor será el precio, pero, como se mencionó anteriormente, se posee más de un valor de emisiones para este candidato a

descriptor, por lo que dicho resultado no se puede asegurar acertado.

Para los gráficos<sup>5</sup> asociados a los descriptores mencionados en el párrafo anterior (ver Figuras D.7, D.8 y D.9), se observa que no existe una tendencia marcada para ninguno de ellos, razón por la que los coeficientes de correlación obtenidos no fueron altos en comparación las demás propiedades macroscópicas y sus descriptores asociados. Específicamente, para el número de hidrógenos y carbonos, se observa que la nube de puntos, presenta una tendencia a crecer levemente a medida que se aumentan los valores de estos descriptores (despreciando lo dispersado que están los datos). Por otro lado, nuevamente, para el número de grupos *OH*, se observa que los datos se encuentran concentrados en torno a valores puntuales debido a que no existen variadas moléculas con más de tres grupos *OH*, por lo que para un mismo número de hidroxilos, se tendrán diferentes valores de costos. Si se toman solamente los últimos puntos, esta vez, no se aprecia un aumento significativo al variar de 2 a 3 grupos *OH* en un compuesto.

Por último, como se dijo anteriormente, no hay que olvidar que el valor de las emisiones y de algunos de los candidatos a descriptores, corresponden solamente a valores parciales, ya que se realizaron estimaciones a los datos utilizados en su cálculo o se consideraron parte de los valores totales, por lo que no se podría asegurar que las relaciones obtenidas sean correctas completamente, pero se debe destacar que se desarrolló una estrategia para estudiar las relaciones que existen entre las variables de entrada y salida del presente problema.

A partir de los resultados obtenidos anteriormente, se escogieron los posibles descriptores para cada una de las propiedades macroscópicas definidas, a las que se les efectuará un análisis de componentes principales. Bajo los criterios mencionados en la Sección 3.6.1 ( $p - value < 0,05$  y  $\rho > 0,35$ ) se obtuvo que los tres descriptores moleculares (número de carbonos, número de hidrógenos y número de grupos *OH*) definidos, estaban relacionados con el rendimiento de hidrógeno, las emisiones generadas y el costo de producción.

#### 4.5.2. PCA

A partir del análisis de componentes principales se esperan obtener nuevas variables que agrupen a los descriptores más importantes y que estén relacionados con el comportamiento de las propiedades macroscópicas definidas. Para ello, primero, se agruparon los descriptores por propiedad macroscópica de acuerdo a los criterios vistos en la Sección 3.6.2, realizando posteriormente el análisis de componentes por cada uno de estos grupos.

Para facilitar este cálculo se creó una función que agrupaba los descriptores en tres matrices (*AddMatrix*), una por cada propiedad macroscópica (para más detalle, ver Anexo E.9). Además, como MATLAB® cuenta con una herramienta para realizar el análisis de componentes principales, la que entrega una matriz con las nuevas variables ordenadas de mayor a menor varianza (ver Sección 3.6.2 para mayor especificación), se utilizó ésta para este resultado.

---

<sup>5</sup>Estos gráficos fueron realizados en escala logarítmica, ya que los valores que se tenían eran muy distintos y no se podían analizar correctamente.

En las Tablas D.1, D.2 y D.3 en Anexos, se aprecian las matrices resultantes de realizar el análisis de componentes principales al conjunto de descriptores. En éstas las columnas representan las nuevas variables ( $v_1$ ,  $v_2$  y  $v_3$ ), las filas los descriptores (número de carbonos, número de hidrógenos y número de grupos  $OH$ ) y las entradas de la cada matriz, los coeficientes  $a_{i,j}$  de la ec. 3.6. A través de esto, se obtuvo que los valores obtenidos para el rendimiento, las emisiones y los costos son iguales, ya que este análisis solo considera el valor de los descriptores y no la propiedad macroscópica con la que están relacionados. Por lo tanto, al obtener el mismo conjunto de propiedades microscópicas para las tres, resulta la misma matriz de componentes principales.

A partir de los valores obtenidos para los PCAs, se deduce que los descriptores: número de hidrógenos y número de grupos  $OH$ ; se pueden agrupar en una nueva variable, la que puede resumir ambos comportamientos. Esto se aprecia en que si se observan los coeficientes obtenidos tanto para la variable  $v_2$  como  $v_3$ , los valores asociados a estos descriptores, no son despreciables ( $\gg 0$ ), a diferencia del valor resultante para el número de carbonos en esta variable<sup>6</sup>. En cuanto a la variable  $v_1$ , se observa que solamente resume el número de carbonos, ya que los coeficientes asociado al resto de los descriptores, son despreciables.

A continuación se muestran las nuevas variables definidas, despreciando los valores que son muy pequeños.

$$\begin{aligned}v_1 &= Z_1 \\v_2 &= 0,420\dot{Z}_2 + 0,903\dot{Z}_3 \\v_3 &= 0,903\dot{Z}_2 - 0,420\dot{Z}_3\end{aligned}$$

Ahora bien, ya que se tienen dos variables ( $v_2$  y  $v_3$ ) que resumen el comportamiento de estos descriptores (numero de hidrógenos y número de grupos  $OH$ ) a través de la combinación lineal de ambos ( $Z_2$  y  $Z_3$ ), se calculó el porcentaje de varianza que acumula cada nueva variable a través de los PCAs realizados (ver Tabla 4.7) para determinar si se puede elegir una sola de ellas.

A partir de esto, se obtiene que la primera variable ( $v_1$ ) resume el comportamiento de cada propiedad macroscópica, lo que es parcialmente acertado. Esto se debe a que el número de hidrógenos, dado las características de los compuestos seleccionados a analizar (ver Sección 3.2), se puede determinar únicamente a través del número de carbonos, si se supiera, *a priori*, qué los compuestos son alcanos<sup>7</sup>. Pero para determinar el número de grupos hidroxilo, es más difícil, ya que no existe alguna fórmula que relacione el número de carbonos y oxígeno en los alcoholes (para el número de hidrógeno en este grupo sigue sirviendo la fórmula  $2n + 2$ ). Por este motivo, la variable  $v_1$ , no puede resumir toda la información, ya que además, el coeficiente asociado al número de grupos  $OH$  ( $a_{1,3}$ ), es muy pequeño.

<sup>6</sup>Para este caso, se despreció el valor de  $a_{2,1}$  y  $a_{3,1}$ , ya que son muy pequeños ( $-2x10^{-5}$  y  $1,2x10^{-7}$ )

<sup>7</sup>En los alcanos, para calcular el número de hidrógenos, se usa la fórmula  $2n + 2$ , siendo  $n$  el número de carbonos.

(a)	
	$\Delta Var$
$v_1$	1
$v_2$	1
$v_3$	1

(b)	
	$\Delta Var$
$v_1$	1
$v_2$	1
$v_3$	1

(c)	
	$\Delta Var$
$v_1$	1
$v_2$	1
$v_3$	1

Tabla 4.7: Varianza acumulada para cada nueva variable del PCA realizado con respecto al 4.7a rendimiento de hidrógeno, 4.7b las emisiones y 4.7c costo de producción.

Por lo expuesto en el párrafo anterior, se aceptaron las variables  $v_1$  y  $v_2$ , las que, además de resumir toda la información (ver Tabla 4.7) de los descriptores, explican el comportamiento de cada propiedad macroscópica definida. A continuación, se muestra las ecuaciones de dichas variables en que  $Z_1$ , representa el número de carbonos;  $Z_2$ , el número de hidrógenos; y  $Z_3$  el número de grupos  $OH$ :

$$v_1 = Z_1 \tag{4.8}$$

$$v_2 = 0,420\dot{Z}_2 + 0,903\dot{Z}_3 \tag{4.9}$$

A partir de lo anterior, se puede decir que el número de hidrógenos y grupos  $OH$  en una molécula, están relacionados, ya que se puede resumir el comportamiento de ambas, en una sola variable. Finalmente, se debe destacar que a través del análisis de componentes principales, se logró resumir el comportamiento de los descriptores a partir de dos variables nuevas.

# Capítulo 5

## Conclusiones

*Computer Aided Molecular and Process Design* o CAMPD, corresponde a una metodología diseñada para poder resolver problemas inversos de predicción de propiedades a través la definición de descriptores y QSPRs; y, así, facilitar las decisiones relacionadas con la elección de moléculas y procesos de acuerdo a ciertas propiedades macroscópicas.

En particular, el presente trabajo busca determinar los descriptores que son estadísticamente significativos para la elección de la mejor molécula que se pueda reformar y proponer un tamaño de piezas *óptimo* que permita representar el conjunto de moléculas a analizar. Para ello, se definieron las propiedades macroscópicas que se quieren optimizar, se seleccionó el conjunto de moléculas a analizar, se seleccionó un tamaño de *piezas* óptimo, se determinaron los candidatos a descriptores, se estudió la relación existente entre variables de entrada y salida (propiedades microscópicas y macroscópicas) y se realizó un análisis de componentes principales para determinar las variables importantes para el sistema.

A partir de esto, se logró cumplir con los objetivos generales planteados: definir descriptores estadísticamente significativos para el desempeño de moléculas en el proceso de reformado y proponer un tamaño de *piezas* que permita representar el conjunto de moléculas a analizar de manera óptima.

El mayor problema que se tuvo en la realización del presente trabajo, fue el relacionado con la obtención de los datos termodinámicos de cada compuestos, porque muchos de estos no se pudieron conseguir de las bases de datos químicas, por lo que se debieron aproximar, ya que eran necesarios para desarrollar una estrategia para determinar la existencia de relación entre variables de entrada y salida al problema.

Las propiedades macroscópicas definidas, fueron el rendimiento de hidrógeno, las emisiones generadas en el proceso y el costo de producción que conlleva reformar cada compuesto orgánico.

En cuanto a sus valores, estos fueron calculadas en base a estimaciones, ya que solo consideraban una parte del valor total y, además, muchos de los datos utilizados para su

obtención, no se pudieron encontrar, por lo que tuvieron que ser aproximados a través de relaciones.

Por otro lado, se creó una base de datos con el universo de moléculas a analizar y la información pertinente para realizar el posterior análisis estadístico. Cabe señalar que este conjunto fue restringido de acuerdo a ciertos criterios de composición, estructura molecular y según los datos de las propiedades que se pudieron encontrar o estimar.

Se definió que el tamaño de *piezas* era el de grupos atómicos, ya que cumplía con los requisitos principales: representar toda la base de datos y generar el menor nivel de ruido. Si bien el tamaño de piezas atómicos puede representar también estos compuestos y, además, algunos que no se consideraron como moléculas con doble o triple enlace, el nivel de ruido que genera es mucho mayor, por lo que provocará que el gasto computacional futuro al desarrollar CAMPD, sea elevado, a diferencia del seleccionado.

El análisis estadístico fue exitoso, ya que, a pesar de que solo se pudo estudiar si existía una relación lineal a través del análisis cuantitativo, se pudo complementar por medio del análisis cualitativo, determinando así, la existencia de alguna relación no lineal entre las variables.

Específicamente, a partir del análisis estadístico cuantitativo, se obtuvo que los coeficientes de correlación indicaban que los candidatos a descriptores (número de carbonos, hidrógenos y grupos *OH*), estaban relacionados con las propiedades macroscópicas definidas. Si bien este análisis está basado en la suposición de linealidad en la relación, se realizó uno cualitativo (a través de gráficos), que confirmó efectivamente la existencia de la linealidad.

Del análisis de componentes principales, se obtienen dos variables que resumen el comportamiento de las propiedades macroscópicas: la primera, que contempla solo el número de carbonos, y la segunda que es una combinación lineal entre el número de hidrógenos y grupos *OH*.

Ahora, si bien en presente trabajo se logra obtener que los descriptores moleculares estadísticamente significativos (tanto para el rendimiento de hidrógeno como las emisiones generadas y los costos asociados al proceso) son el número de carbonos, hidrógenos y grupos hidroxilos que están presentes en el compuesto a reformar; estos no son los únicos que se pueden identificar, ya que, durante el trabajo, se mencionaron algunos otros (como el ángulo de enlace, la accesibilidad al carbono por presencia de *OH* y *H*, entre otros más), pero dada la complejidad de su cálculo, no se optó por incluirlos. Además, se debe recalcar que el análisis cuantitativo, si bien se utilizó para comprobar que los candidatos a descriptores eran estadísticamente significativos, este solo determina si existe alguna dependencia lineal entre las variables y no de otro tipo, lo que podría generar un sesgo en los resultados obtenidos.

Por último, en cuanto al tamaño óptimo de *piezas* para representar el conjunto de moléculas reformables seleccionadas, se eligió el *grupo de átomos*, el que tiene la ventaja, además de representar a este conjunto, de generar un bajo número de estructuras no factibles fisicoquímicamente, por lo que a futuro, en la siguientes etapas de CAMPD, solamente se deberán ingresar algunas restricciones para filtrar un conjunto pequeño en comparación con la elección de otro tamaño de *piezas*.



## 5.1. Recomendaciones y trabajos futuros

### Recomendaciones

En base al trabajo realizado, se pueden plantear algunas recomendaciones para afinar y verificar los resultados obtenidos. Estas son:

- Buscar en otras bases de datos los valores de las propiedades microscópicas que fueron estimadas, para así poder compararlas y modificar las propiedades que presenten mayor error. Además, adicionar los compuestos que no fueron considerados, ya que no se poseían todos los datos necesarios.
- Comparar los resultados que se obtendrían con este nuevo conjunto de compuestos y los anteriores, para así determinar la veracidad de las aproximaciones realizadas y comprobar los descriptores obtenidos.
- Estudiar alguna relación no lineal (a través del análisis cualitativo) que se pueda registrar entre las propiedades microscópicas y macroscópicas, ya que, para utilizar el análisis cuantitativo, se supone que existe alguna relación lineal entre las variables.

### Trabajos futuros

A partir de los descriptores obtenidos para las propiedades macroscópicas y las *piezas* definidas, se pueden plantear los siguientes trabajos futuros para así poder realizar CAMPD y, finalmente, resolver las preguntas planteadas a lo largo de este trabajo:

- Creación de QSPRs para poder obtener los valores de las propiedades macroscópicas que se desean estimar.
- Validar estadísticamente los QSPRs a través de compuestos que han sido reformados, es decir, buscar en literatura valores para el rendimiento de hidrógeno, emisiones y costo de producción para compuestos que han sido reformados y compararlos con los valores obtenidos a través de los QSPRs.
- Plantear y resolver el problema de optimización.

# Bibliografía

- [1] GLISENTI, A., GALENDA, A., NATILE, M. Steam reforming and oxidative steam reforming of methanol and ethanol: The behaviour of LaCo<sub>0.7</sub>Cu<sub>0.3</sub>O<sub>3</sub>. Applied Catalysis A: General, 453(0): 102 – 112, 2013.
- [2] YERMÁN, L., HOMS, N., DE LA PISCINA, P.R. Hydrogen production from oxidative steam-reforming of n-propanol over Ni/Y<sub>2</sub>O<sub>3</sub>-ZrO<sub>2</sub> catalysts. International Journal of Hydrogen Energy, 37(8): 7094 – 7100, 2012. III Iberian Symposium on Hydrogen, Fuel Cells and Advanced Batteries, HYCELTEC-2011.
- [3] WU, G., LI, S., ZHANG, C., WANG, T., GONG, J. Glycerol steam reforming over perovskite-derived nickel-based catalysts. Applied Catalysis B: Environmental, 144(0): 277 – 285, 2014.
- [4] HUANG, X., REIMERT, R. Kinetics of steam reforming of ethane on Ni/YSZ (yttria-stabilised zirconia) catalyst. Fuel, 106(0): 380 – 387, 2013.
- [5] YAN, Y., ZHANG, J., ZHANG, L. Properties of thermodynamic equilibrium-based methane autothermal reforming to generate hydrogen. International Journal of Hydrogen Energy, 38(35): 15744 – 15750, 2013.
- [6] SHI, L., BAYLESS, D.J. Analysis of jet fuel reforming for solid oxide fuel cell applications in auxiliary power units. International Journal of Hydrogen Energy, 33(3): 1067 – 1075, 2008.
- [7] SHI, L., BAYLESS, D.J. Analysis of jet fuel reforming for solid oxide fuel cell applications in auxiliary power units. International Journal of Hydrogen Energy, 33(3): 1067–1075, feb. 2008.
- [8] GREENE, D.L., HOPSON, J.L., LI, J. Have we run out of oil yet? Oil peaking analysis from an optimist's perspective. Energy Policy, 34(5): 515 – 531, 2006. Hong Kong Editorial Board meeting presentations.
- [9] RAMSDEN, T.S., D. ZUBOY, J. Analyzing the Levelized Cost of Centralized and Distributed Hydrogen Production Using the H<sub>2</sub>A Production Model, Version 2.

National Renewable Energy Laboratory, 2009.

- [10] ESLICK, J.C., SHULDA, S.M., SPENCER, P., CAMARDA, K.V. Optimization-Based Approaches to Computational Molecular Design, págs. 173–193. Wiley-VCH Verlag GmbH & Co. KGaA, 2011.
- [11] ROUGHTON, B.C., CHRISTIAN, B., WHITE, J., CAMARDA, K.V., GANI, R. Simultaneous design of ionic liquid entrainers and energy efficient azeotropic separation processes. Computers & Chemical Engineering, 42(0): 248 – 262, 2012. European Symposium of Computer Aided Process Engineering - 21.
- [12] GANI, R. CAMD: Computer Aided Molecular Design - Examples of Applications. CAMD: COmputer Aided Molecular Design - Examples of Applications, 1: 18, 2004.
- [13] ÖRÜCÜ, E., KARAKAYA, M., AVCI, A.K., ÖNSAN, Z.I. Investigation of ethanol conversion for hydrogen fuel cells using computer simulations. Journal of Chemical Technology & Biotechnology, 80(10): 1103–1110, 2005.
- [14] SRINIVASAN, S. Fuel Cells: From Fundamentals to Applications. Materials Today, 9(10): 57 –, 2006.
- [15] DÍAZ ALVARADO, F.A. Reformado de alcoholes para la producción de hidrógeno: análisis termodinámico y propuesta de mecanismos de reacción mediante Grafos-P. Tesis de Doctorado, Facultad de Ciencias Físicas y Matemáticas - Universidad de Chile, 2012.
- [16] XU, X., LI, P., SHEN, Y. Small-scale reforming of diesel and jet fuels to make hydrogen and syngas for fuel cells: A review. Applied Energy, 108(0): 202 – 217, 2013.
- [17] HARPER, P.M., GANI, R., KOLAR, P., ISHIKAWA, T. Computer-aided molecular design with combined molecular modeling and group contribution. Fluid Phase Equilibria, 158–160(0): 337 – 347, 1999.
- [18] CAMARDA, K.V., MARANAS, C.D. .optimisation in polymer design using connectivity indices. Industrial & Engineering Chemistry Research, 38: 1884–1892, 1999.
- [19] HARPER, P.M., GANI, R. A multi-step and multi-level approach for computer aided molecular design. Computers & Chemical Engineering, 24(2–7): 677 – 683, 2000.
- [20] CABEZAS, H. Designing green solvents. Chemical Engineering, 107: 107–109, March 2000.
- [21] PEREIRA, F., KESKES, E., GALINDO, A., JACKSON, G., ADJIMAN, C. Integrated solvent and process design using a SAFT-VR thermodynamic description: High-pressure separation of carbon dioxide and methane. Computers & Chemical Engineering, 35(3): 474 – 491, 2011.

- [22] DUVEDI, A.P., ACHENIE, L.E. Designing environmentally safe refrigerants using mathematical programming. Chemical Engineering Science, 51(15): 3727 – 3739, 1996.
- [23] STRÜBING, H., KARAMERTZANIS, P.G., PISTIKOPOULOS, E.N., GALINDO, A., ADJIMAN, C.S. Solvent design for a Menshutkin reaction by using CAMD and DFT calculations. En: S. PIERUCCI, G.B. FERRARIS (Eds.) 20th European Symposium on Computer Aided Process Engineering, Computer Aided Chemical Engineering, tomo 28, págs. 1291 – 1296. Elsevier, 2010.
- [24] WEIS, D.C., VISCO, D.P. Computer-aided molecular design using the Signature molecular descriptor: Application to solvent selection. Computers & Chemical Engineering, 34(7): 1018 – 1029, 2010. Process Modeling and Control in Drug Development and Manufacturing.
- [25] PISTIKOPOULOS, E., STEFANIS, S. Optimal solvent design for environmental impact minimization. Computers & Chemical Engineering, 22(6): 717 – 733, 1998. Design of chemical compounds.
- [26] MARCOULAKI, E., KOKOSSIS, A. On the development of novel chemicals using a systematic optimisation approach. Part II. Solvent design. Chemical Engineering Science, 55(13): 2547 – 2561, 2000.
- [27] SIMASATITKUL, L., ARPORNWICHANOP, A., GANI, R. Design methodology for bio-based processing: Biodiesel and fatty alcohol production. Computers & Chemical Engineering, 0(0): –, 2013.
- [28] SIPPL, W., CONTRERAS, J.M., PARROT, I., RIVAL, Y.M., WERMUTH, C.G. Structure-based 3D QSAR and design of novel acetylcholinesterase inhibitors. Journal of computer-aided molecular design, 15(5): 395–410, 2001.
- [29] GIOVANOGLU, A., BARLATIER, J., ADJIMAN, C.S., PISTIKOPOULOS, E.N., CORDINER, J.L. Optimal solvent design for batch separation based on economic performance. AIChE Journal, 49(12): 3095–3109, 2003.
- [30] DE VLIAGER, D., MOJET, B., LEFFERTS, L., SESHAN, K. Aqueous Phase Reforming of ethylene glycol – Role of intermediates in catalyst performance. Journal of Catalysis, 292(0): 239 – 245, 2012.
- [31] JUN, J.W., SUH, Y.W., SUH, D.J., LEE, Y.K. Thermodynamic analysis of steam and aqueous reforming of hydroxylated {C6} aliphatic compounds. Journal of Industrial and Engineering Chemistry, 19(6): 2072 – 2078, 2013.
- [32] NIEVA, M.A., VILLAVERDE, M.M., MONZÓN, A., GARETTO, T.F., MARCHI, A.J. Steam-methane reforming at low temperature on nickel-based catalysts. Chemical Engineering Journal, 235(0): 158 – 166, 2014.
- [33] PÉREZ-MORENO, L., SOLER, J., HERGUIDO, J., MENÉNDEZ, M. Stable hydrogen

- production by methane steam reforming in a two zone fluidized bed reactor: Experimental assessment. Journal of Power Sources, 243(0): 233 – 241, 2013.
- [34] CAI, W., DE LA PISCINA, P.R., HOMS, N. Hydrogen production from the steam reforming of bio-butanol over novel supported Co-based bimetallic catalysts. Bioresource Technology, 107(0): 482 – 486, 2012.
- [35] HUANG, L., ZHOU, J., HSU, A.T., CHEN, R. Catalytic partial oxidation of n-butanol for hydrogen production over LDH-derived Ni-based catalysts. International Journal of Hydrogen Energy, 38(34): 14550 – 14558, 2013.
- [36] IULIANELLI, A., RIBEIRINHA, P., MENDES, A., BASILE, A. Methanol steam reforming for hydrogen generation via conventional and membrane reactors: A review. Renewable and Sustainable Energy Reviews, 29(0): 355 – 368, 2014.
- [37] MAYR, L., LORENZ, H., ARMBRÜSTER, M., VILLASECA, S.A., LUO, Y., CARDOSO, R., BURKHARDT, U., ZEMLYANOV, D., HAEVECKER, M., BLUME, R., KNOP-GERICKE, A., KLÖTZER, B., PENNER, S. The catalytic properties of thin film Pd-rich GaPd<sub>2</sub> in methanol steam reforming. Journal of Catalysis, 309(0): 231 – 240, 2014.
- [38] DÍAZ, F., ADÁN, D., DÍAZ, E., FREDES, N., MARTÍNEZ, G., MIRANDA, J., PROBOSTE, V., SOTO, E., SCHLATTER, R., TAPIA, M., ZAPATA, I. Colaboración en trabajos de título DIQBT. Octubre 2013. [Entrevista].
- [39] MATLAB. versión 7.14.0 (R2012a). The MathWorks Inc., Natick, Massachusetts, 2012.
- [40] DIVESTADÍSTICA. Diccionario estadístico. [en línea] <[www.divestadistica.es/es/diccionario\\_estadistico.html#T](http://www.divestadistica.es/es/diccionario_estadistico.html#T)>. [Consulta: 30 de Diciembre de 2013].
- [41] INEI. Glosario básico de términos estadísticos. [en línea] <[www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib0900/Libro.pdf](http://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib0900/Libro.pdf)>. [Consulta: 30 de Diciembre de 2013].
- [42] UCM. Análisis de correlación lineal: Los procedimientos Correlaciones bivariada y Correlaciones parciales. [en línea] <[pendientedemigracion.ucm.es/info/socivmyt/paginas/D\\_departamento/materiales/analisis\\_datosyMultivariable/17corlin\\_SPSS.pdf](http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/17corlin_SPSS.pdf)>. [Consulta: 30 de Diciembre de 2013].
- [43] UHU. Práctica 9: Correlación. [en línea] <[www.uhu.es/45110/Ficheros%20de%20datos/cursos%202007%202008/spss/PRACTICA%209.pdf](http://www.uhu.es/45110/Ficheros%20de%20datos/cursos%202007%202008/spss/PRACTICA%209.pdf)>. [Consulta: 30 de Diciembre de 2013].
- [44] Correlación, Pearson y Spearman. [en línea] <[www.e-pfb.com/ebiometria/pfb\\_teb/tecnicas\\_y\\_casos/te11.pdf](http://www.e-pfb.com/ebiometria/pfb_teb/tecnicas_y_casos/te11.pdf)>. [Consulta: 30 de Diciembre de 2013].
- [45] UC3M. Coeficiente de correlación lineal de Pearson. [en línea] <[kovachi.sel.inf](http://kovachi.sel.inf)>.

- uc3m.es/@api/deki/files/141/=correlacion.pdf>. [Consulta: 30 de Diciembre de 2013].
- [46] BUESA, E. Tema 18: Contraste de dos variables cuantitativas. [en línea] <[www.eduardobuesa.es/Tema18.pdf](http://www.eduardobuesa.es/Tema18.pdf)>. [Consulta: 30 de Diciembre de 2013].
- [47] TERRÁDEZ GURREA, M. Analisis de componentes principales. [en línea] <[www.uoc.edu/in3/emath/docs/Componentes\\_principales.pdf](http://www.uoc.edu/in3/emath/docs/Componentes_principales.pdf)>. [Consulta: 30 de Diciembre de 2013].
- [48] DE LA FUENTE FERNÁNDEZ, S. Componentes Principales. [en línea] <[www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/ACP/ACP.pdf](http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/ACP/ACP.pdf)>. [Consulta: 30 de Diciembre de 2013].
- [49] DEL CAMBIO CLIMÁTICO, O.C. Metodología de Inventarios de Gases de Efecto Invernadero (GEI). [en línea] <[www20.gencat.cat/docs/canviclimatic/Home/Politiques/Inventaris%20demissions/Mes%20informacio%20sobre%20metodologies/metodologia\\_annex\\_ES.pdf](http://www20.gencat.cat/docs/canviclimatic/Home/Politiques/Inventaris%20demissions/Mes%20informacio%20sobre%20metodologies/metodologia_annex_ES.pdf)>. [Consulta: 13 de Noviembre de 2013].
- [50] MINENERGÍA. Inventarios de Emisiones de GEI para PyMEs. [en línea] <[huelladecarbono.minenergia.cl](http://huelladecarbono.minenergia.cl)>. [Consulta: 13 de Noviembre de 2013].
- [51] DEL CANVI CLIMÀTIC, O.C. Guía práctica para el cálculo de emisiones de gases de efecto invernadero (GEI). Generalitat de Catalunya Comisión Interdepartamental del Cambio Climático, marzo 2012 ed<sup>ón</sup>., Marzo 2012.
- [52] MINENERGÍA. Factores de emisión de CO2 de los combustibles utilizados en Chile. [en línea] <[huelladecarbono.minenergia.cl/combustible-chile](http://huelladecarbono.minenergia.cl/combustible-chile)>. [Consulta: 13 de Noviembre de 2013].
- [53] SIGMA-ALDRICH. SIGMA-ALDRICH. [en línea] <[www.sigmaaldrich.com/chile.html](http://www.sigmaaldrich.com/chile.html)>. [Consulta: 23 de Septiembre de 2013].
- [54] CAS. CAS number. [en línea] <[www.cas.org/content/chemical-substances/](http://www.cas.org/content/chemical-substances/)>. [Consulta: 4 de Noviembre de 2013].
- [55] CHERIC. Korea Thermophysical Properties Data Bank. [en línea] <[www.cheric.org/research/kdb/](http://www.cheric.org/research/kdb/)>. [Consulta: 23 de Septiembre de 2013].
- [56] NIST. NIST Chemistry WebBook. [en línea] <[webbook.nist.gov/](http://webbook.nist.gov/)>. [Consulta: 23 de Septiembre de 2013].
- [57] MATSUMURA, Y. Development of durable copper catalyst for hydrogen production by high temperature methanol steam reforming. International Journal of Hydrogen Energy, 38(32): 13950 – 13960, 2013.

- [58] ROBERT H. PERRY, D.W.G. Perry's Chemical Engineers' HandBook. McGraw-Hill, 2008, 2-485 págs.
- [59] MARTÍNEZ CARO, R. Dilemas éticos en la ingeniería. Tesis de Magister, Facultad de Ciencias Físicas y Matemáticas - Universidad de Chile, 2012.

# Anexos





# Anexo A

## Propiedades Macroscópicas

### A.1. Matriz de propiedades macroscópicas

Tabla A.1: Matriz que contiene en sus entradas las propiedades macroscópicas para cada uno de los compuesto a analizar. En la primera columna se encuentra el número CAS, en la segunda Eficiencia de hidrógeno ( $[\frac{kmolH_2}{kmolOrg}]$ ), en la tercera emisiones ( $[\frac{kmolCO_2}{kmolOrg}]$ ) y en la cuarta costos ( $[\frac{CLP}{ml}]$ ). En las filas se representan cada uno de las moléculas del conjunto.

74828	4	1,73	15,42
74840	7	3,39	4272,73
74986	10	5,06	1993,1
106978	13	6,74	1252,94
109660	16	8,37	77
110543	19	10,03	450
142825	22	11,68	610
111659	25	13,34	450
111842	28	14,98	880
124185	31	16,61	97
1120214	34	18,26	840
112403	37	19,87	1120
629505	40	21,52	1160
629594	43	23,22	8400
629629	46	24,88	840
544763	49	26,54	610
67561	3	1,54	270
64175	6	3,24	77
71238	9	4,9	360
67630	9	4,92	340
71363	12	6,56	340
78922	12	6,58	390
71410	15	8,23	230
111273	18	9,82	24000
107211	5	3,11	360
56815	7	4,73	350

# Anexo B

## Base de Datos

### B.1. Base de datos

Tabla B.1: Compuestos y datos almacenados en la base de datos.

CAS	Name	Structure	WT	Densityg_ml	Tdensity_C
74-82-8	METHANE	CH <sub>4</sub>	16,0426	0,716	25
74-84-0	ETHANE	C <sub>2</sub> H <sub>6</sub>	30,0694	0,362	20
74-98-6	PROPANE	C <sub>3</sub> H <sub>8</sub>	44,0962	1,55	25
106-97-8	N-BUTANE	C <sub>4</sub> H <sub>10</sub>	58,123	0,579	20
109-66-0	N-PENTANE	C <sub>5</sub> H <sub>12</sub>	72,1498	0,626	25
110-54-3	N-HEXANE	C <sub>6</sub> H <sub>14</sub>	86,1766	0,659	25
142-82-5	N-HEPTANE	C <sub>7</sub> H <sub>16</sub>	100,203	0,684	25
111-65-9	N-OCTANE	C <sub>8</sub> H <sub>18</sub>	114,23	0,703	25
111-84-2	N-NONANE	C <sub>9</sub> H <sub>20</sub>	128,257	0,718	25
124-18-5	N-DECANE	C <sub>10</sub> H <sub>22</sub>	142,284	0,73	25
1120-21-4	N-UNDECANE	C <sub>11</sub> H <sub>24</sub>	156,311	0,74	25
112-40-3	N-DODECANE	C <sub>12</sub> H <sub>26</sub>	170,337	0,75	25
629-50-5	N-TRIDECANE	C <sub>13</sub> H <sub>28</sub>	184,364	0,756	25
629-59-4	N-TETRADECANE	C <sub>14</sub> H <sub>30</sub>	198,391	0,762	20
629-62-9	N-PENTADECANE	C <sub>15</sub> H <sub>32</sub>	212,418	0,769	25
544-76-3	N-HEXADECANE	C <sub>16</sub> H <sub>34</sub>	226,445	0,773	25
67-56-1	METHANOL	CH <sub>4</sub> O	32,042	0,791	25
64-17-5	ETHANOL	C <sub>2</sub> H <sub>6</sub> O	46,0688	0,789	25
71-23-8	N-PROPANOL	C <sub>3</sub> H <sub>8</sub> O	60,0956	0,804	25
67-63-0	ISOPROPANOL	C <sub>3</sub> H <sub>8</sub> O	60,0956	0,785	25
71-36-3	N-BUTANOL	C <sub>4</sub> H <sub>10</sub> O	74,1224	0,81	25
78-83-1	ISOBUTANOL	C <sub>4</sub> H <sub>10</sub> O	74,1224	0,803	25
78-92-2	SEC-BUTANOL	C <sub>4</sub> H <sub>10</sub> O	74,1224	0,808	25
71-41-0	N-PENTANOL	C <sub>5</sub> H <sub>12</sub> O	88,1492	0,811	25
6032-29-7	2-PENTANOL	C <sub>5</sub> H <sub>12</sub> O	88,1492	0,812	25

584-02-1	3-PENTANOL	C5H12O	88,1492	0,815	25
111-27-3	1-HEXANOL	C6H14O	102,176	0,814	25
626-93-7	2-HEXANOL	C6H14O	102,176	0,81	25
107-21-1	ETHYLENE_GLYCOL	C2H6O2	62,0682	1,113	25
504-63-2	1,3-PROPYLENE_GLYCOL	C3H8O2	76,095	1,053	25
56-81-5	GLYCEROL	C3H8O3	92,0944	1,25	25
110-63-4	1,4-BUTANEDIOL	C4H10O2	90,1218	1,017	25
107-88-0	1,3-BUTANEDIOL	C4H10O2	90,1218	1,005	25
513-85-9	2,3-BUTANEDIOL	C4H10O2	90,1218	1,002	20
149-32-6	ERYTHRITOL	C4H10O4	122,121	1,44	20
111-29-5	1,5-PENTANEDIOL	C5H12O2	104,149	0,994	25
629-11-8	1,6-HEXANEDIOL	C6H14O2	118,175	0,967	0
69-65-8	MANNITOL	C6H14O6	182,173	1,52	20

Tabla B.2: Compuestos y datos almacenados en la base de datos.

CAS	Name	BoilingPoint_C	MealtingPoint_C
74-82-8	METHANE	-161	-183
74-84-0	ETHANE	-88	-172
74-98-6	PROPANE	-42,1	-188
106-97-8	N-BUTANE	-0,5	-138
109-66-0	N-PENTANE	35	-130
110-54-3	N-HEXANE	69	-95
142-82-5	N-HEPTANE	98	-91
111-65-9	N-OCTANE	126	-57
111-84-2	N-NONANE	151	-53
124-18-5	N-DECANE	174	-30
1120-21-4	N-UNDECANE	196	-26
112-40-3	N-DODECANE	216	-9,6
629-50-5	N-TRIDECANE	234	-5
629-59-4	N-TETRADECANE	253	5,5
629-62-9	N-PENTADECANE	270	9
544-76-3	N-HEXADECANE	287	18
67-56-1	METHANOL	64,7	-98
64-17-5	ETHANOL	78	-114
71-23-8	N-PROPANOL	97	-127
67-63-0	ISOPROPANOL	82	-89,5
71-36-3	N-BUTANOL	117	-90
78-83-1	ISOBUTANOL	108	-108
78-92-2	SEC-BUTANOL	98	-115
71-41-0	N-PENTANOL	137	-78
6032-29-7	2-PENTANOL	118	0
584-02-1	3-PENTANOL	114	0

111-27-3	1-HEXANOL	156	-52
626-93-7	2-HEXANOL	136	0
107-21-1	ETHYLENE_GLYCOL	195	-13
504-63-2	1,3-PROPYLENE_GLYCOL	214	-27
56-81-5	GLYCEROL	182	20
110-63-4	1,4-BUTANEDIOL	230	16
107-88-0	1,3-BUTANEDIOL	203	121
513-85-9	2,3-BUTANEDIOL	183	25
149-32-6	ERYTHRITOL	330	119
111-29-5	1,5-PENTANEDIOL	242	0
629-11-8	1,6-HEXANEDIOL	250	40
69-65-8	MANNITOL	293	168

Tabla B.3: Compuestos y datos almacenados en la base de datos.

CAS	Name	DHfkJ_kmol	Cpsol_kJ_kmolC
74-82-8	METHANE	-74900	15,557471776
74-84-0	ETHANE	-84740	0
74-98-6	PROPANE	-103900	0
106-97-8	N-BUTANE	-126200	0
109-66-0	N-PENTANE	-146500	0
110-54-3	N-HEXANE	-167300	0
142-82-5	N-HEPTANE	-187900	0
111-65-9	N-OCTANE	-208600	0
111-84-2	N-NONANE	-229200	0
124-18-5	N-DECANE	-249800	0
1120-21-4	N-UNDECANE	-270500	0
112-40-3	N-DODECANE	-291100	0
629-50-5	N-TRIDECANE	-311700	0
629-59-4	N-TETRADECANE	-332300	0
629-62-9	N-PENTADECANE	-353000	0
544-76-3	N-HEXADECANE	-373600	468,5373495
67-56-1	METHANOL	-201300	0
64-17-5	ETHANOL	-235000	0
71-23-8	N-PROPANOL	-256600	0
67-63-0	ISOPROPANOL	-272600	0
71-36-3	N-BUTANOL	-274900	0
78-83-1	ISOBUTANOL	-283400	0
78-92-2	SEC-BUTANOL	-292800	0
71-41-0	N-PENTANOL	-298900	0
6032-29-7	2-PENTANOL	-312700	0
584-02-1	3-PENTANOL	-317200	0
111-27-3	1-HEXANOL	-317800	0

626-93-7	2-HEXANOL	-333500	0
107-21-1	ETHYLENE_GLYCOL	-389600	94,956897816
504-63-2	1,3-PROPYLENE_GLYCOL	-409100	0
56-81-5	GLYCEROL	-585300	127,03501536
110-63-4	1,4-BUTANEDIOL	-426000	0
107-88-0	1,3-BUTANEDIOL	-447100	0
513-85-9	2,3-BUTANEDIOL	-544800	0
149-32-6	ERYTHRITOL	-910480	179,17348878
111-29-5	1,5-PENTANEDIOL	-444500	0
629-11-8	1,6-HEXANEDIOL	-462100	0
69-65-8	MANNITOL	-1337500	0

Tabla B.4: Compuestos y datos almacenados en la base de datos.

CAS	Name	DHfusion_kJ_kmol	Cpliq_kJ_kmolK
74-82-8	METHANE	940,82469404	82,56
74-84-0	ETHANE	2854,673369504	17,46
74-98-6	PROPANE	3520,5524156	61,36
106-97-8	N-BUTANE	4657,4308638	0
109-66-0	N-PENTANE	8405,20639068	159,08
110-54-3	N-HEXANE	13018,28531432	172,12
142-82-5	N-HEPTANE	14035,6145754	0
111-65-9	N-OCTANE	20612,872038	224,83
111-84-2	N-NONANE	22087,907512	383,08
124-18-5	N-DECANE	28726,285896	278,62
1120-21-4	N-UNDECANE	22280,257318	293,98
112-40-3	N-DODECANE	36526,044258	508,21
629-50-5	N-TRIDECANE	0	350,18
629-59-4	N-TETRADECANE	0	353,14
629-62-9	N-PENTADECANE	0	356,91
544-76-3	N-HEXADECANE	0	370,35
67-56-1	METHANOL	0	112,37
64-17-5	ETHANOL	0	102,64
71-23-8	N-PROPANOL	0	158,76
67-63-0	ISOPROPANOL	0	0
71-36-3	N-BUTANOL	0	191,2
78-83-1	ISOBUTANOL	0	185,75
78-92-2	SEC-BUTANOL	0	0
71-41-0	N-PENTANOL	0	201,2
6032-29-7	2-PENTANOL	0	0
584-02-1	3-PENTANOL	0	0
111-27-3	1-HEXANOL	0	484,66
626-93-7	2-HEXANOL	0	0

107-21-1	ETHYLENE_GLYCOL	0	35,54
504-63-2	1,3-PROPYLENE_GLYCOL	0	0
56-81-5	GLYCEROL	18281,49357408	0
110-63-4	1,4-BUTANEDIOL	0	0
107-88-0	1,3-BUTANEDIOL	0	0
513-85-9	2,3-BUTANEDIOL	0	0
149-32-6	ERYTHRITOL	0	0
111-29-5	1,5-PENTANEDIOL	0	0
629-11-8	1,6-HEXANEDIOL	0	0
69-65-8	MANNITOL	0	0

Tabla B.5: Compuestos y datos almacenados en la base de datos.

CAS	Name	DHvap_kJ_kmol	Cpgas_kJ_kmolK
74-82-8	METHANE	8180	36,15
74-84-0	ETHANE	14707	33,31
74-98-6	PROPANE	18774	29,6
106-97-8	N-BUTANE	22393	24,26
109-66-0	N-PENTANE	25773	33,78
110-54-3	N-HEXANE	28853	32,37
142-82-5	N-HEPTANE	31698	37,66
111-65-9	N-OCTANE	34413	37,21
111-84-2	N-NONANE	36915	33,01
124-18-5	N-DECANE	39279	121,19
1120-21-4	N-UNDECANE	41505	130,8
112-40-3	N-DODECANE	43639	140,41
629-50-5	N-TRIDECANE	45647	254,89
629-59-4	N-TETRADECANE	47614	159,63
629-62-9	N-PENTADECANE	49455	169,02
544-76-3	N-HEXADECANE	51212	178,86
67-56-1	METHANOL	35254	34,37
64-17-5	ETHANOL	38744	19,96
71-23-8	N-PROPANOL	41756	18,88
67-63-0	ISOPROPANOL	39832	24,21
71-36-3	N-BUTANOL	43095	11,71
78-83-1	ISOBUTANOL	0	0
78-92-2	SEC-BUTANOL	40794	28,78
71-41-0	N-PENTANOL	44350	3,87
6032-29-7	2-PENTANOL	0	0
584-02-1	3-PENTANOL	0	0
111-27-3	1-HEXANOL	48116	49,07
626-93-7	2-HEXANOL	0	0
107-21-1	ETHYLENE_GLYCOL	52509	35,7

504-63-2	1,3-PROPYLENE_GLYCOL	0	8,27
56-81-5	GLYCEROL	0	8,42
110-63-4	1,4-BUTANEDIOL	0	0
107-88-0	1,3-BUTANEDIOL	0	0
513-85-9	2,3-BUTANEDIOL	0	0
149-32-6	ERYTHRITOL	0	0
111-29-5	1,5-PENTANEDIOL	0	0
629-11-8	1,6-HEXANEDIOL	0	0
69-65-8	MANNITOL	0	0

Tabla B.6: Compuestos y datos almacenados en la base de datos.

CAS	Name	Tc_K	Price	UnitP	Quantity	UnitQ
74-82-8	METHANE	190	185000	CLP	12000	ml
74-84-0	ETHANE	305	470000	CLP	110	g
74-98-6	PROPANE	369	289000	CLP	145	g
106-97-8	N-BUTANE	425	213000	CLP	170	g
109-66-0	N-PENTANE	469	77000	CLP	1000	ml
110-54-3	N-HEXANE	507	45000	CLP	100	ml
142-82-5	N-HEPTANE	504	61000	CLP	100	ml
111-65-9	N-OCTANE	568	45000	CLP	100	ml
111-84-2	N-NONANE	594	88000	CLP	100	ml
124-18-5	N-DECANE	617	97000	CLP	1000	ml
1120-21-4	N-UNDECANE	639	21000	CLP	25	ml
112-40-3	N-DODECANE	658	112000	CLP	100	ml
629-50-5	N-TRIDECANE	675	29000	CLP	25	g
629-59-4	N-TETRADECANE	693	42000	CLP	5	ml
629-62-9	N-PENTADECANE	708	21000	CLP	25	g
544-76-3	N-HEXADECANE	723	61000	CLP	100	ml
67-56-1	METHANOL	512	27000	CLP	100	ml
64-17-5	ETHANOL	514	77000	CLP	1000	ml
71-23-8	N-PROPANOL	536	36000	CLP	100	m
67-63-0	ISOPROPANOL	508	34000	CLP	100	ml
71-36-3	N-BUTANOL	563	34000	CLP	100	ml
78-83-1	ISOBUTANOL	547	62000	CLP	1000	g
78-92-2	SEC-BUTANOL	536	39000	CLP	100	ml
71-41-0	N-PENTANOL	588	23000	CLP	100	ml
6032-29-7	2-PENTANOL	560	29000	CLP	100	ml
584-02-1	3-PENTANOL	559	25000	CLP	25	ml
111-27-3	1-HEXANOL	632	24000	CLP	1	ml
626-93-7	2-HEXANOL	608	48000	CLP	25	g
107-21-1	ETHYLENE_GLYCOL	718	36000	CLP	100	ml
504-63-2	1,3-PROPYLENE_GLYCOL	724	30000	CLP	100	g

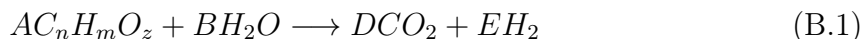


56-81-5	GLYCEROL	726	35000	CLP	100	ml
110-63-4	1,4-BUTANEDIOL	0	30000	CLP	1000	ml
107-88-0	1,3-BUTANEDIOL	0	34000	CLP	100	ml
513-85-9	2,3-BUTANEDIOL	0	107000	CLP	100	g
149-32-6	ERYTHRITOL	0	66000	CLP	25	g
111-29-5	1,5-PENTANEDIOL	0	36000	CLP	100	g
629-11-8	1,6-HEXANEDIOL	0	33000	CLP	50	g
69-65-8	MANNITOL	0	20000	CLP	100	g

## B.2. Cálculo de la reacción general y sus coeficientes estequiométricos

Para obtener la ecuación general de reformado (SR), se utilizó la ec. 4.3 como base y se procedió de la siguiente forma:

- Se fijó la siguiente ecuación<sup>1</sup>:



- Debido a que si se hace el balance por especie ( $C$ ,  $H$  y  $O$ ) se tendrán 3 ecuaciones, y ya se tienen 5 incógnitas ( $A$ ,  $B$ ,  $D$  y  $E$ ), se tuvieron que realizar las siguientes simplificaciones o suposiciones:
  - $A = 1$  para facilitar los cálculos.
- Luego, teniendo la misma cantidad de ecuaciones que de incógnitas, se procedió a realizar el balance por especie.

$$\text{Balance para C: } 1 \cdot n = D \cdot 1 \quad (\text{B.2})$$

$$\text{Balance para H: } 1 \cdot m + B \cdot 2 = F \cdot 2 \quad (\text{B.3})$$

$$\text{Balance para O: } 1 \cdot z + B = D \cdot 2 \quad (\text{B.4})$$

- De la ec. B.2 se desprende que:

$$D = n \quad (\text{B.5})$$

---

<sup>1</sup>Los valores de  $n$ ,  $m$  y  $z$  se consideran conocidos, ya que se obtienen a partir del compuesto a reformar

- De la ec. B.3 se despeja  $B = \frac{2E-m}{2}$ , reemplazando en la ec. (B.4) y despejando  $E$  se sigue:

$$z + \frac{2E - m}{2} = 2n$$

$$E = 2n + \frac{m}{2} - z \tag{B.6}$$

- Despejando finalmente  $B$

$$B = 2n - z \tag{B.7}$$

- Con lo que se obtienen los valores estequiométricos.



# Anexo C

## Candidatos a Descriptores

### C.1. Matriz de Descriptores

Tabla C.1: Matriz que contiene en sus entradas la información de los descriptores definidos para cada compuesto. Las filas representan las moléculas y las columnas los candidatos a descriptores. El orden de estas últimas es el siguiente: CAS, Número de carbonos, Número de hidrógenos, Número de grupos *OH*.

74828	1	4	0
74840	2	6	0
74986	3	8	0
106978	4	10	0
109660	5	12	0
110543	6	14	0
142825	7	16	0
111659	8	18	0
111842	9	20	0
124185	10	22	0
1120214	11	24	0
112403	12	26	0
629505	13	28	0
629594	14	30	0
629629	15	32	0
544763	16	34	0
67561	1	3	1
64175	2	5	1
71238	3	7	1
67630	3	7	1
71363	4	9	1
78922	4	9	1
71410	5	11	1
111273	6	13	1
107211	2	4	2
56815	3	5	3

# Anexo D

## Gráficos y tablas: significancia estadística

### D.1. Gráficos: Análisis cualitativo

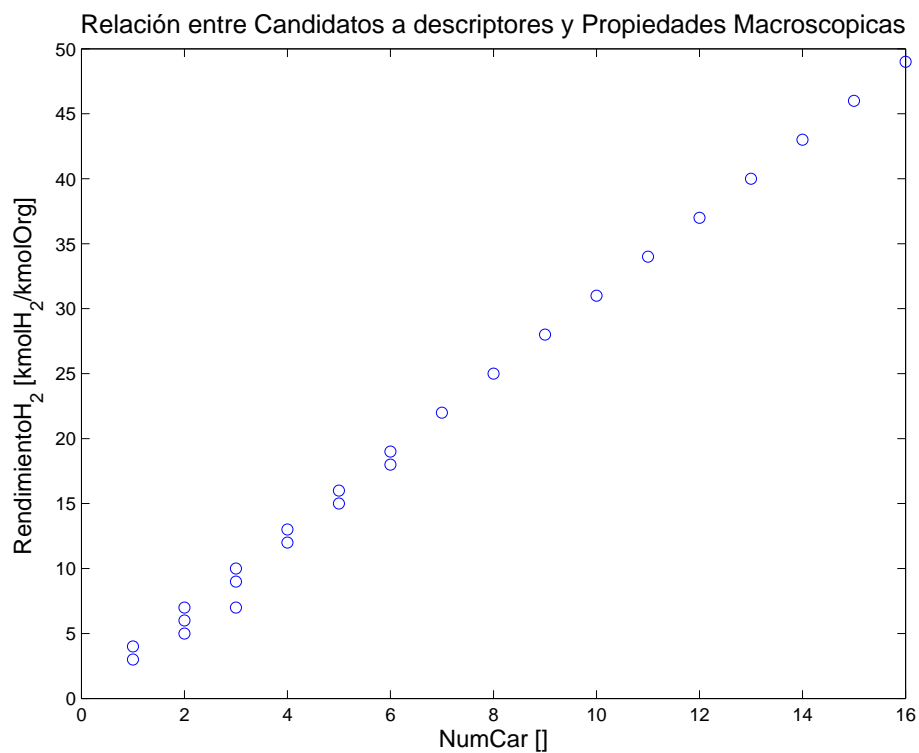


Figura D.1: Gráfico de relación entre Número de carbonos y Rendimiento de hidrógeno.

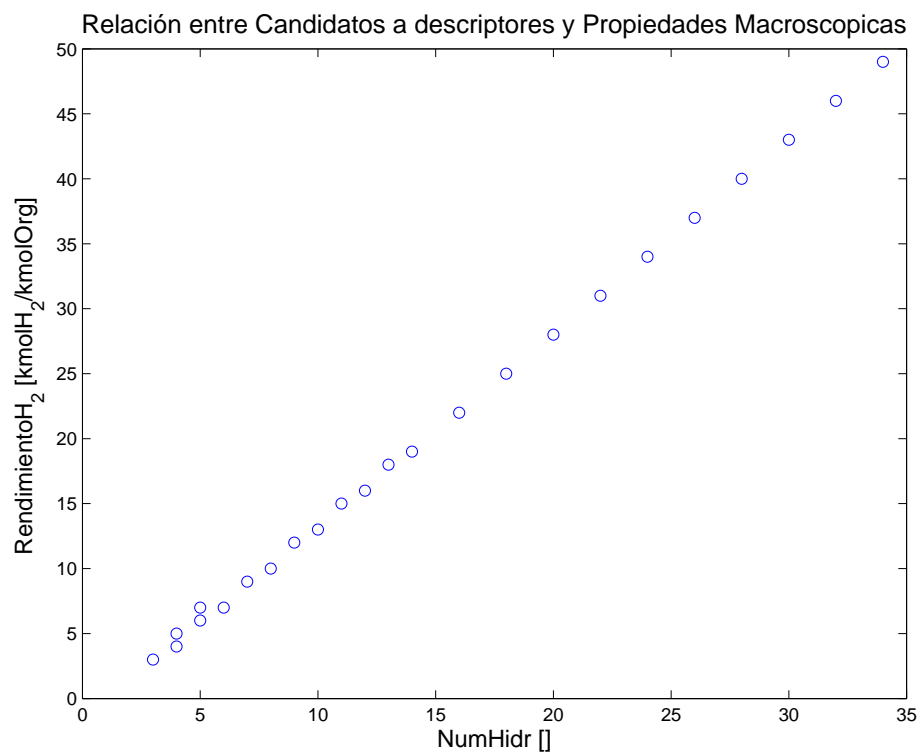


Figura D.2: Gráfico de relación entre Número de hidrógenos y Rendimiento de hidrógeno.

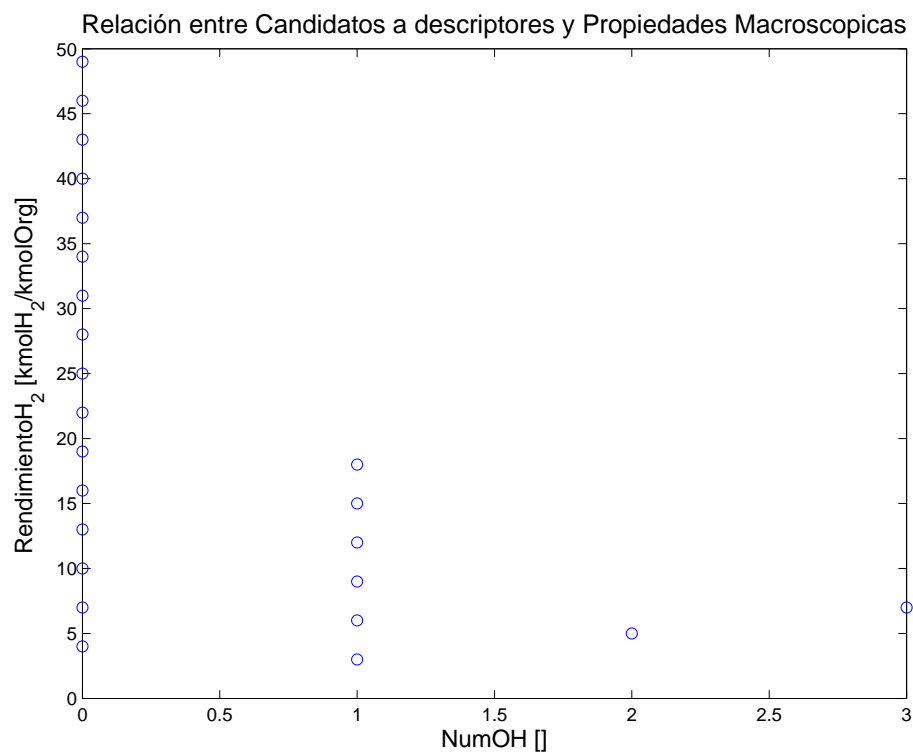


Figura D.3: Gráfico de relación entre Número de grupos *OH* y Rendimiento de hidrógeno.

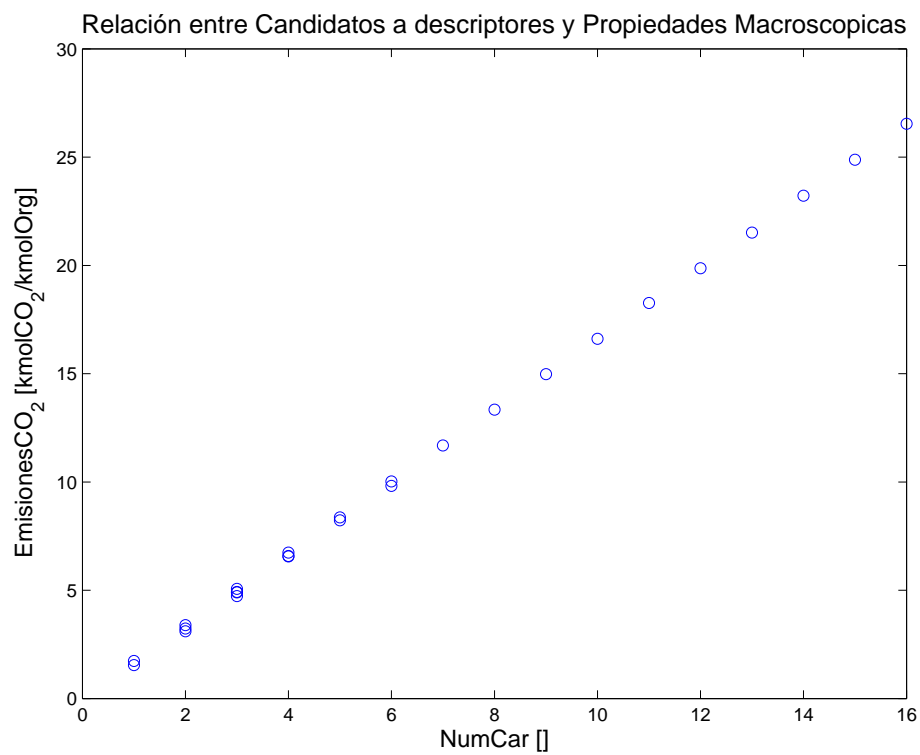


Figura D.4: Gráfico de relación entre Número de carbonos y Emisiones.

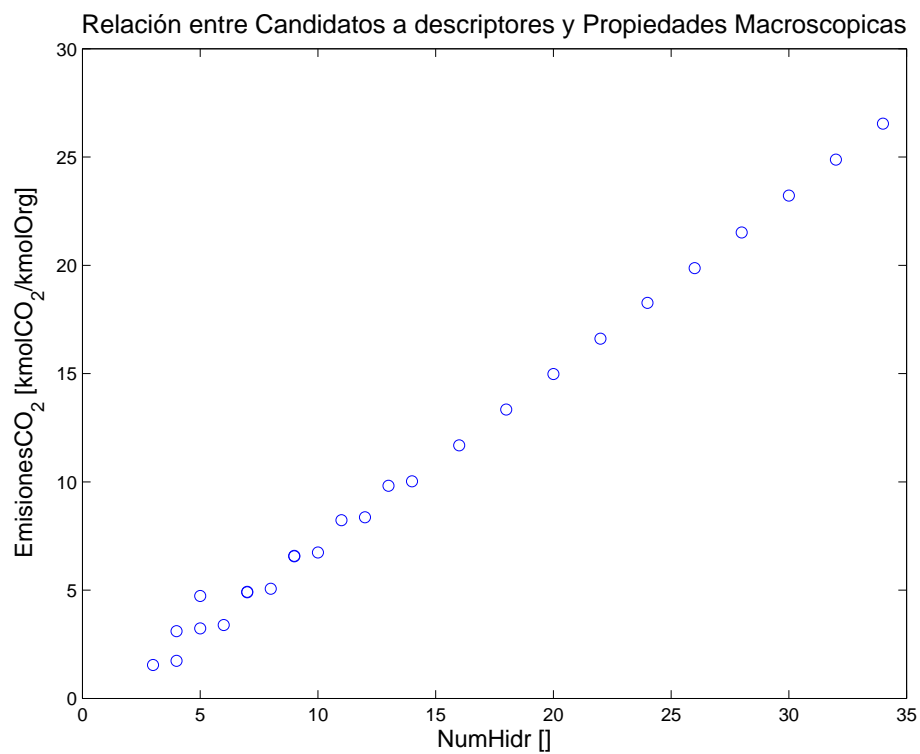


Figura D.5: Gráfico de relación entre Número de hidrógenos y Emisiones.

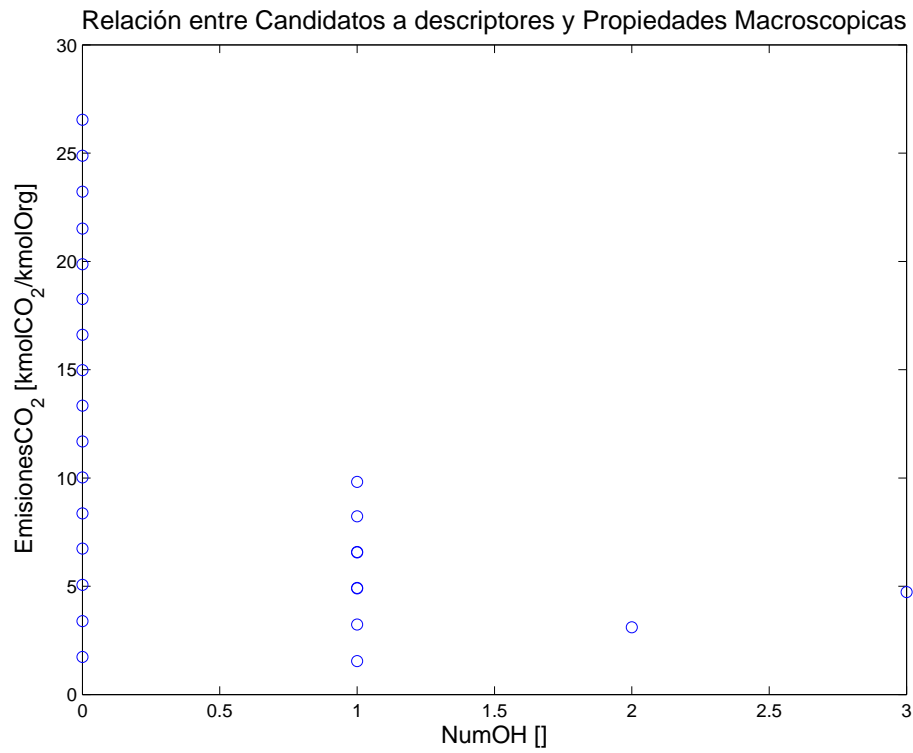


Figura D.6: Gráfico de relación entre Número de grupos *OH* y Emisiones.

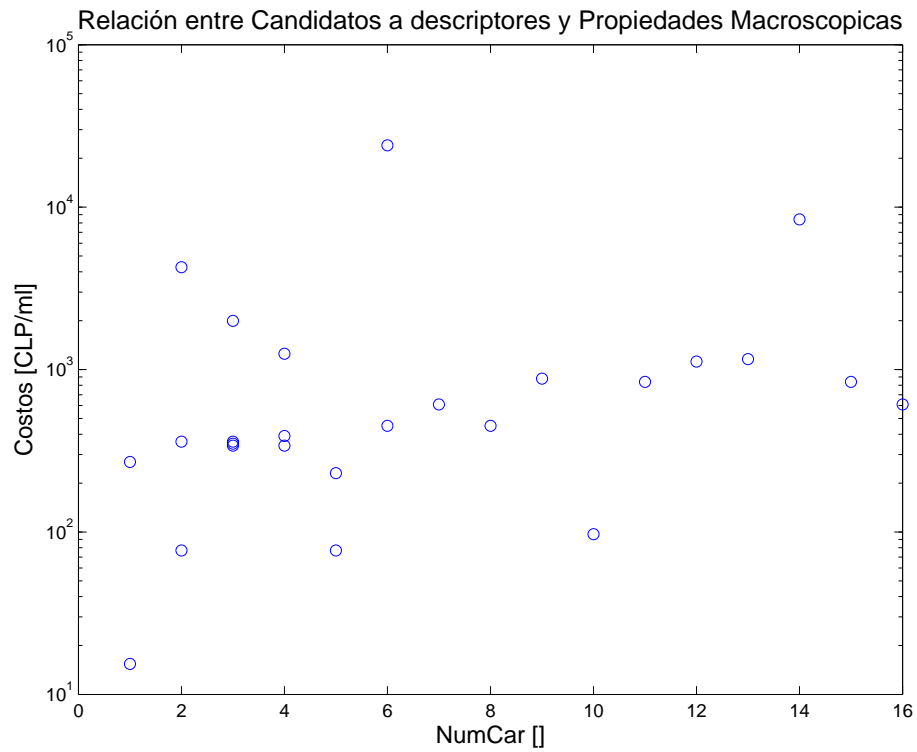


Figura D.7: Gráfico de relación entre Número de carbonos y Costos.



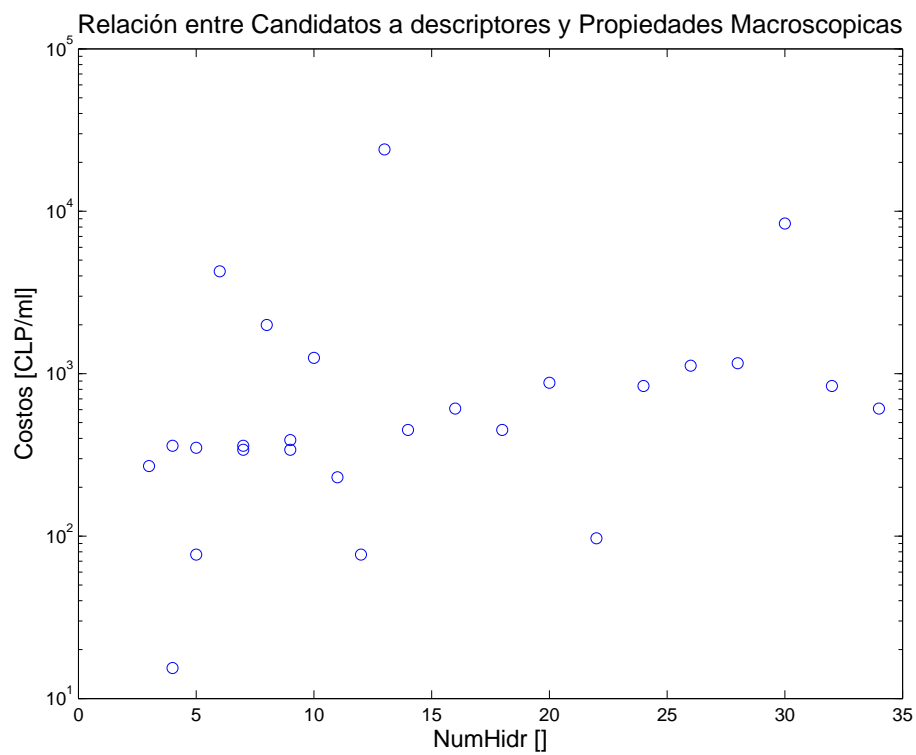


Figura D.8: Gráfico de relación entre Número de hidrógenos y Costos.

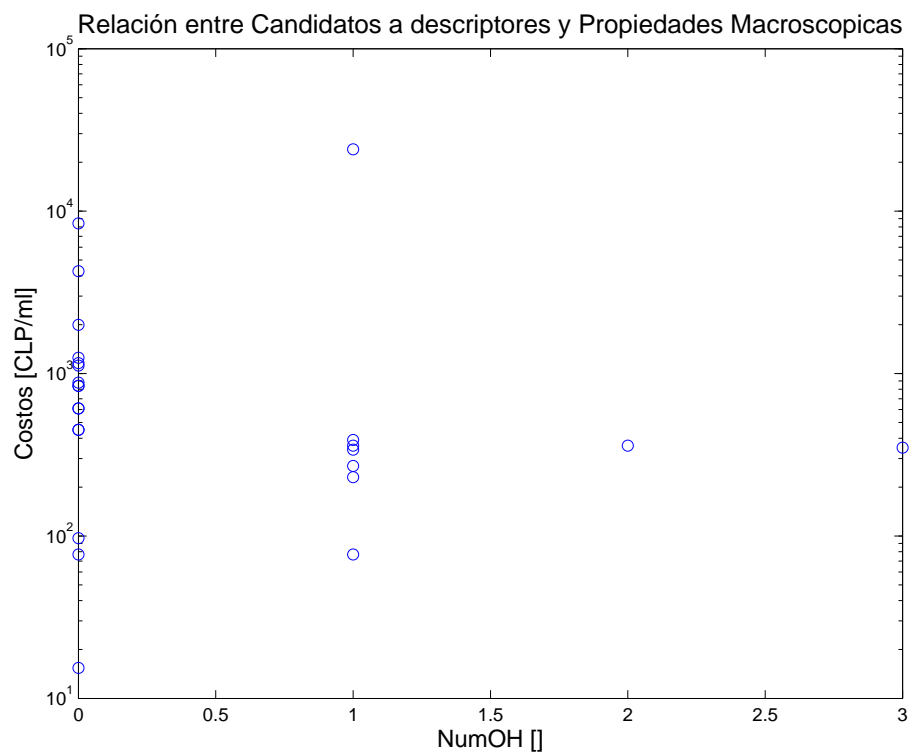


Figura D.9: Gráfico de relación entre Número de grupos  $OH$  y Costos.

## D.2. Resultados del PCA

Tabla D.1: Matriz que contiene en sus entradas los valores de los coeficientes de las componentes principales (ver Sección 3.6.2) para el grupo de descriptores relacionados con la eficiencia de hidrógeno. Las columnas representes a las nuevas variables, ordenadas de mayor a menor varianza, y las filas los candidatos a descriptores.

	$v_1$	$v_2$	$v_3$
$N^\circ$ de carbonos	0,999	-2,87E-05	-1,28E-07
$N^\circ$ de hidrógenos	1,24E-05	0,429	0,903
$N^\circ$ de grupos $OH$	2,58E-05	0,903	-0,429

Tabla D.2: Matriz que contiene en sus entradas los valores de los coeficientes de las componentes principales (ver Sección 3.6.2) el grupo de descriptores relacionados con las emisiones. Las columnas representes a las nuevas variables, ordenadas de mayor a menor varianza, y las filas los candidatos a descriptores.

	$v_1$	$v_2$	$v_3$
$N^\circ$ de carbonos	0,999	-2,87E-05	-1,28E-07
$N^\circ$ de hidrógenos	1,24E-05	0,429	0,903
$N^\circ$ de grupos $OH$	2,58E-05	0,903	-0,429

Tabla D.3: Matriz que contiene en sus entradas los valores de los coeficientes de las componentes principales (ver Sección 3.6.2) para el grupo de descriptores relacionados con el costo. Las columnas representes a las nuevas variables, ordenadas de mayor a menor varianza, y las filas los candidatos a descriptores.

	$v_1$	$v_2$	$v_3$
$N^\circ$ de carbonos	0,999	-2,87E-05	-1,28E-07
$N^\circ$ de hidrógenos	1,24E-05	0,429	0,903
$N^\circ$ de grupos $OH$	2,58E-05	0,903	-0,429

# Anexo E

## Funciones Matlab

### E.1. Búsqueda en base de datos

```
1 function r = SearchBD(BD, categ, value)
2 % Busca la posicion de un compuesto en base de datos segun alguna categoria.
3 % Si
4 % no existe el compuesto, devuelve 0. Devuelve la posicion del compuesto.
5 % INPUT:
6 % BD: estructura de datos en la que se esta buscando.
7 % categ: categoria de busqueda
8 % 1: numero CAS
9 % 2: nombre
10 % 3: Estructura
11 % value: numero CAS, nombre o estructura del compuestos
12 % OUTPUT:
13 % r: posicion del compuesto en l base de datos (dentro del vector de celdas)
14
15 % verificar si categoria de busqueda existe
16 [n,m] = size(BD.CAS);
17 if categ~=1 && categ~=2 && categ~=3
18     r = 0;
19     disp('Bad category');
20     return;
21 end
22
23 for k=1:m
24     % busqueda por numero CAS
25     if categ==1 && strcmpi(value, BD.CAS{k})
26         r = k;
27         return;
28     end
29     % busqueda por nombre
30     if categ==2 && strcmpi(value, BD.Name{k})
31         r = k;
32         return;
33     end
34     % busqueda por estructura
```

```

    if categ==3 && strcmpi(value, BD.Structure{k})
35         r = k;
           return;
37     end
end
39 r = 0;
disp('No exist');

```

## E.2. Conteo de átomos de una molécula

```

1 function BD = getAtomMol(BD)
   % Funcion que me entrega el numero de carbonos (Cn), hidrogenos (Hm) y
3   % Oxigenos (Oz) en una estructura molecular de una base de datos
   % Siempre habra Carbono e hidrogeno, pero no necesariamente Oxigeno
5   %INPUT
   %     BD: Base de datos de los compuestos
7   %OUIPUT
   %     BD: Base de datos modificada
9
   [n,m] = size(BD.Structure);
11  Cn = zeros(m,1);
   Hm = zeros(m,1);
13  Oz = zeros(m,1);

15  for k=1:m
       strc = BD.Structure{k};
17       l = length(strc);
       n = strfind(strc, 'C');
19       m = strfind(strc, 'H');
       z = strfind(strc, 'O');
21
       % Verificar si hay mas de un carbono (siempre hay carbono)
23       if n == m-1
           Cn(k) = 1;
25       else
           Cn(k) = str2num(strc(n+1:m-1));
27       end

29       % Caso de Oxigeno, puede haber presencia de este o no
       if isempty(z)
31           Oz(k) = 0;
           if m == 1
33               Hm = 1;
           else
35               Hm(k) = str2num(strc(m+1:l));
           end
37       else
           % Verificar presencia de oxigeno
39           if z == 1
               Oz(k) = 1;
41           else
               Oz(k) = str2num(strc(z+1:l));
43           end

```

```

45     % Verificar si hay mas de un hidrogeno (siempre hay hidrogeno)
46     if m == z-1
47         Hm = 1;
48     else
49         Hm(k) = str2num(strc(m+1:z-1));
50     end
51 end
52 BD.CarbonNum{k} = Cn(k);
53 BD.HydrogenNum{k} = Hm(k)-Oz(k); % Por cada oxigeno hay un hidrogeno que
    lo acompaña
54 BD.HydroxideNum{k} = Oz(k);
55 end

```

### E.3. Cálculo coeficientes estequiométricos

```

1 function BD = reaction(BD)
2 % Funcion que entrega los coeficientes estequiometricos de la reaccion
3 % general ACnHmOz + BH2O -> DCO2 + ECO + FH2
4 % INPUT
5 %     BD: Base de datos de los compuestos
6 % OUTPUT
7 %     BD: Base de datos modificada
8
9 [n,m] = size(BD.CAS);
10 A = zeros(m,1);
11 B = zeros(m,1);
12 D = zeros(m,1);
13 E = zeros(m,1);
14 F = zeros(m,1);
15
16 for k=1:m
17     A(k) = 1; % Se calculo en base a la molecula a reformar
18     B(k) = 2*BD.CarbonNum{k}-BD.HydroxideNum{k};
19     D(k) = BD.CarbonNum{k};
20     E(k) = 0; % Se supuso que solo se producía CO2
21     F(k) = 2*BD.CarbonNum{k}+(BD.HydroxideNum{k}+BD.HydrogenNum{k})/2-BD.
        HydroxideNum{k};
22     BD.Stoichiometric{k} = [A(k) B(k) D(k) E(k) F(k)];
23 end
24
25 end

```

### E.4. Estimación de datos

```

1 function BD = CpgCplCpsDHvDHf(BD)
2 % Funcion que llena los datos de la base de datos de calores especificos y
3 % entalpias de vaporizacion y fusion a traves de factores
4 % INPUT

```

```

5 %      BD: Base de datos de los compuestos
6 %OUTPUT
7 %      BD: Base de datos modificada

9 r1 = Cpl_CpgFactor(BD); % factor Cpl/Cpg
10 r2 = Cps_CpgFactor(BD); % Factor Cps/Cpg
11 r3 = DHf_DHvFactor(BD); % Factor DHf/DHv

13 [n,m] = size(BD.CAS);

15 for i=1:m

17     % Se hacen 2 ciclos para completar datos si o si
18     for k=1:2
19         % Completar Cpsol
20         if str2num(BD.Cpsol_kJ_kmolC{i})==0
21             BD.Cpsol_kJ_kmolC{i} = num2str(r1*str2num(BD.Cpgas_kJ_kmolK{i}));
22         end
23         if str2num(BD.Cpgas_kJ_kmolK{i})==0
24             BD.Cpgas_kJ_kmolK{i} = num2str(str2num(BD.Cpsol_kJ_kmolC{i})/r1);
25         end

27         % Completar Cpliq
28         if str2num(BD.Cpliq_kJ_kmolK{i})==0
29             BD.Cpliq_kJ_kmolK{i} = num2str(r2*str2num(BD.Cpgas_kJ_kmolK{i}));
30         end
31         if str2num(BD.Cpgas_kJ_kmolK{i})==0
32             BD.Cpgas_kJ_kmolK{i} = num2str(str2num(BD.Cpliq_kJ_kmolK{i})/r2);
33         end
34     end

35     % Completar DHfus
36     if str2num(BD.DHfusion_kJ_kmol{i})==0
37         BD.DHfusion_kJ_kmol{i} = num2str(r3*str2num(BD.DHvap_kJ_kmol{i}));
38     end
39     if str2num(BD.DHvap_kJ_kmol{i})==0
40         BD.DHvap_kJ_kmol{i} = num2str(str2num(BD.DHfusion_kJ_kmol{i})/r3);
41     end
42 end
43 end
44 end
45 end

```

#### E.4.1. Estimación de entalpías de fusión y vaporización

```

1 function r = DHf_DHvFactor(BD)
2 % Funcion que determina el promedio de las razones entre DHv y DHs de
3 % los distintos compuestos
4 %INPUT:
5 %      BD: Base de datos donde estan los Cpliq y Cpgas
6 %OUTPUT:
7 %      r: Promedio de las razones DHfus/DHvap

9 DHf = BD.DHfusion_kJ_kmol;

```

```

DHv = BD.DHvap_kJ_kmol;
11 [n,m] = size(BD.CAS);
13
15 sum = 0;
15 n_comp = 0;
17 for i=1:m
17     if str2num(DHf{i})~=0 && str2num(DHv{i})~=0
19         n_comp = n_comp + 1;
19         sum = sum + (str2num(DHf{i})/str2num(DHv{i}));
21     end
21 end
23 r = sum/n_comp;
23 end

```

### E.4.2. Estimación de calor específico (sólido-gas)

```

1 function r = Cps_CpgFactor(BD)
2 % Funcion que determina el promedio de las razones entre Cpsol y Cpgas de
3 % los distintos compuestos
4 %INPUT:
5 %     BD: Base de datos donde estan los Cpliq y Cpgas
6 %OUTPUT:
7 %     r: Promedio de las razones Cpsol/Cpgas
8
9 Cps = BD.Cpsol_kJ_kmolC;
10 Cpg = BD.Cpgas_kJ_kmolK;
11
12 [n,m] = size(BD.CAS);
13
14 sum = 0;
15 n_comp = 0;
17 for i=1:m
17     if str2num(Cps{i})~=0 && str2num(Cpg{i})~=0
19         n_comp = n_comp + 1;
19         sum = sum + (str2num(Cps{i})/str2num(Cpg{i}));
21     end
21 end
23 r = sum/n_comp;
23 end

```

### E.4.3. Estimación de calor específico (líquido-gas)

```

1 function r = Cpl_CpgFactor(BD)
2 % Funcion que determina el promedio de las razones entre Cpliq y Cpgas de
3 % los distintos compuestos
4 %INPUT:
5 %     BD: Base de datos donde estan los Cpliq y Cpgas
6 %OUTPUT:
7 %     r: Promedio de las razones Cpliq/Cpgas

```

```

9 Cpl = BD.Cpliq_kJ_kmolK;
Cpg = BD.Cpgas_kJ_kmolK;
11
13 [n,m] = size(BD.CAS);
15
17 sum = 0;
n_comp = 0;
for i=1:m
19     if str2num(Cpl{i})~=0 && str2num(Cpg{i})~=0
        n_comp = n_comp + 1;
        sum = sum + (str2num(Cpl{i})/str2num(Cpg{i}));
21     end
end
r = sum/n_comp;
23 end

```

## E.5. Cálculo de la entalpía de reacción

```

1 function BD = DHr(BD, Treforming)
% Funcion que calcula la entalpia de reaccion
3 % para cada compuesto a una cierta temperatura
% y la agrega a la base de datos.
5 % INPUT:
% BD: estructura de datos
7 % Treforming: temperatura a la que se reforma
% OUTPUT:
9 % BD: base de datos modificada

11 % Entalpia de formacion para el agua a 600C
DHfrec1 = DHfw(Treforming);
13
15 % Entalpia de formacion para el CO2 a 600C
DHfprod1 = DHfCO2(Treforming);

17 % Entalpia de formacion para el hidrogeno a 600C
DHfprod2 = DHfH2(Treforming);
19
21 % Entalpia de formacion para el compuesto a 600C
DHfrec2= DHf_600(BD, Treforming);

23 [n,m] = size(BD.CAS);

25 for i=1:m
    if isDHR(BD,i)
27         Est = BD.Stoichiometric{i};
        A = Est(1);
29         B = Est(2);
        D = Est(3);
31         E = Est(4);
        F = Est(5);
33         if i==2
            r1 = B*DHFrec1

```



```

35     r2 = A*DHFreact2(i)
36     p1 = D*DHFprod1
37     p2 = F*DHFprod2
38     p1+p2-r2
39     end
40     BD.DHrkJ_kmol{i} = num2str((D*DHFprod1 + F*DHFprod2)-(A*DHFreact2(i)+B*
41         DHfreact1));
42     end
43 end
44 end

```

### E.5.1. Cálculo de entalpía de formación del compuesto a la temperatura de reformado

```

function DHf = DHf_600(BD, Treforming)
2 % Funcion que calcula la entalpia de formacion del compuesto a la
3 % temperatura de reformado
4 %INPUT:
5 %BD: estructura de datos
6 % Treforming: temperatura a la que se reforma
7 %OUTPUT:
8 %BD: base de datos modificada
9
10 [n,m] = size(BD.CAS);
11 DHform = BD.DHfkJ_kmol;
12
13 for i=1:m
14     if isDHR(BD,i)
15         Tbp = str2num(BD.BoilingPoint_C{i});
16         Tmp = str2num(BD.MealtingPoint_C{i});
17         T = str2num(BD.Tdensity_C{i});
18
19         DHs = 0;
20         if T<Tmp
21             Cps = str2num(BD.Cpsol_kJ_kmolC{i});
22             DHs = Cps*(Tmp-T);
23             T = Tmp;
24         end
25
26         DHfus = 0;
27         if T==Tmp
28             DHfus = str2num(BD.DHfusion_kJ_kmol{i});
29         end
30
31         DHl = 0;
32         if T<Tbp
33             Cpl = str2num(BD.Cpliq_kJ_kmolK{i});
34             DHl = Cpl*(Tbp-T);
35             T = Tbp;
36         end
37
38         DHvap = 0;

```

```

40     if T==Tbp
        DHvap = str2num(BD.DHvap_kJ_kmol{i});
42     end

44     DHg = 0;
46     if T<Tforming
        Cpg = str2num(BD.Cpgas_kJ_kmolK{i});
        DHg = Cpg*(Tforming-T);
        T = Tforming;
48     end

50     DHf(i) = str2num(DHform{i})+DHs+DHfus+DHl+DHvap+DHg;
52     else
        DHf(i) = 0;
54     end
end
end

```

### E.5.2. Cálculo de entapía de formación del agua

```

1 function DHf = DHfw(Tforming)
2 % Funcion que calcula la entalpia de formacion del agua a la temperatura de
3 % reformado
4 %INPUT:
5 % Tforming: temperatura a la que se reforma
6 %OUIPUT:
7 %DHf: entalpia de formacion del agua a cierta T

9 DHform = -285840; %kJ/kmol
10 Cplw = 75.3; %kJ/kmolK
11 DHvw = 40650; %kJ/kmol
12 Cpgw = 37.5; %kJ/kmolK
13 Tbpw = 100; %C
14 T = 25; %C

15 DHf = DHform + Cplw*(Tbpw-T) + DHvw + Cpgw*(Tforming-Tbpw);
17 end

```

### E.5.3. Cálculo de entapía de formación del hidrógeno

```

1 function DHf = DHfH2(Tforming)
2 % Funcion que calcula la entalpia de formacion del hidrogeno a la
3 % temperatura de reformado
4 %INPUT:
5 % Tforming: temperatura a la que se reforma
6 %OUIPUT:
7 %DHf: entalpia de formacion del hidrogeno a certa T

9 CpgH2 = 27;
10 DHf = CpgH2*(Tforming-25);
11 end

```

#### E.5.4. Cálculo de entalpía de formación del $CO_2$

```
function DHf = DHfCO2(Treforming)
2 % Funcion que calcula la entalpia de formacion del CO2 a la temperatura de
  % reformado
4 % INPUT:
  % Treforming: temperatura a la que se reforma
6 % OUTPUT:
  % DHf: entalpia de formacion del CO2 a certa T
8 CpgCO2 = 29.27;
10 DHf = CpgCO2*(Treforming -25);
end
```

#### E.6. Cálculo de la matriz de propiedades macroscópicas

```
1 function A = PropMacros(BD,M)
  % Funcion que devuelve una matriz con los valores de las propiedades
3 % macroscopicas para cada compuesto que se tiene todos los datos
  % INPUT
5 %     BD: Base de datos de los compuestos
  %     M: Matriz de descriptores con los compuestos que se tienen todos
7 %     los datos
  % OUTPUT
9 %     A: Matriz con las propiedades macriscopicas
  %     Columnas:
11 %     CAS
  %     Eficiencia de hidrogeno
13 %     Emisiones
  %     Costos
15 %     Filas:
  %     Compuestos
17 [n,m] = size(BD.CAS);
19
  % Factor de emision
21 FE = 56100/(44*10^9); % kmol de CO2/kJ
23 i=1;
  for k=1:m
25
27     % Solo dejo los compuestos que tengo los datos de los descriptores
    if isDHR(BD,k)
29         EffH2 = BD.Stoichiometric{k}(5);
31         Emis1 = BD.Stoichiometric{k}(3); %kmolCO2/kmolOrg
33         Emis2 = abs(str2num(BD.DHrkJ_kmol{k}))*FE; %kmolCO2/kmolOrg
35         Cost = CostoNorm(BD,k);
37
39         A(i,1) = str2num(strrep(BD.CAS{k}, '- ', ''));
41         A(i,2) = EffH2;
43         A(i,3) = Emis1+Emis2;
45         A(i,4) = Cost;
47     end
49 end
```

```

37         i = i+1;
           end
39     end
41 end
end

```

### E.6.1. Normalización del costo del compuesto

```

function r = CostoNorm(BD,i)
2 % Funcion que normaliza el costo en CLP/ml
  %INPUT
4 %     BD: Base de datos
  %     i: indice del compuesto
6 %OUTPUT
  %     r: valor normaizado del costo en CLP/ml
8
P = str2num(BD.Price{i});
10
if strcmp(BD.UnitQ, 'g')
12     Q = str2num(BD.Quantity{i})/str2num(BD.Densityg_ml);
  else
14     Q = str2num(BD.Quantity{i});
  end
16
r = P/Q;
18
end

```

### E.6.2. Función auxiliar de entalpía de reacción

```

1 function r = isDHR(BD,i)
  % Funcion que devuelve 1 si estan todos los datos necesarios para calcular
  % el DHR de todos los compuestos de la BD.
3 %INPUT
  % BD: Base de datos
  %OUTPUT
7 % r: 1 si estan todos los datos necesarios
  % 0 caso contrario
9
T = str2num(BD.Tdensity_C{i});
11 Tbp = str2num(BD.BoilingPoint_C{i});
  Tmp = str2num(BD.MealtingPoint_C{i});
13 DHf = str2num(BD.DHfusion_kJ_kmol{i});
  DHv = str2num(BD.DHvap_kJ_kmol{i});
15 Cpg = str2num(BD.Cpgas_kJ_kmolK{i});
  Cpl = str2num(BD.Cpliq_kJ_kmolK{i});
17 Cps = str2num(BD.Cpsol_kJ_kmolC{i});
19
r = 0;

```

```

21 if Cpg==0
    return;
end
23
25 if T<=Tbp && DHv==0
    return;
end
27
29 if T<Tbp && Cpl==0
    return;
end
31
33 if T<=Tmp && DHf==0
    return;
end
35
37 if T<Tmp && Cps==0
    return;
end
39
41 r = 1;
end

```

### E.6.3. Función auxiliar de entalpía de reacción

```

function M = Descriptores(BD)
2 % Funcion que entrega una matriz con los valores de los descriptores. Es
% una matriz en cuyas columnas estan los descriptores y en las filas estan
4 % los compuestos.
%INPUT
6 %     BD: Base de datos.
%OUIPUT
8 %     M: Matriz con los descriptores
% columnas
10 % CAS-NumCar-NumHidr-NumOH
% filas
12 % Compuestos

14 [n,m] = size(BD.CAS);

16 M = zeros(1,11);
k = 1;
18 for i=1:m
    % solo deajo compuestos que se puede obtener el DHr
20     if isDHR(BD,i)
        cas = str2num(strrep(BD.CAS{i}, '-',' '));
22 %     DHf_std = str2num(BD.DHfkJ_kmol{i});
%     T_ref = str2num(BD.Tdensity_C{i});
24 %     Cps = str2num(BD.Cpsol_kJ_kmolC{i});
%     DHf = str2num(BD.DHfusion_kJ_kmol{i});
26 %     Cpl = str2num(BD.Cpliq_kJ_kmolK{i});
%     DHv = str2num(BD.DHvap_kJ_kmol{i});

```

```

28 %      Cpg = str2num(BD.Cpgas_kJ_kmolK{i});
%      DHr = str2num(BD.DHrkJ_kmol{i});
30 %      Tbp = str2num(BD.BoilingPoint_C{i});
%      Tmp = str2num(BD.MealtingPoint_C{i});
32 N_C = BD.CarbonNum{i};
N_H = BD.HydrogenNum{i};
34 N_OH = BD.HydroxideNum{i};
%      Est_CO2 = BD.Stoichiometric{i}(3);
36 %      Est_H2 = BD.Stoichiometric{i}(5);

38 M(k,1) = cas;
M(k,2) = N_C;
40 M(k,3) = N_H;
M(k,4) = N_OH;

42 %      M(k,1) = cas;
%      M(k,2) = DHf_std;
44 %      M(k,3) = DHr;
%      M(k,4) = Cps;
46 %      M(k,5) = DHf;
%      M(k,6) = Cpl;
48 %      M(k,7) = DHv;
%      M(k,8) = Cpg;
50 %      M(k,9) = Tmp;
%      M(k,10) = Tbp;
52 %      M(k,11) = N_C;
%      M(k,12) = N_H;
54 %      M(k,13) = N_OH;
%      M(k,14) = Est_CO2;
56 %      M(k,15) = Est_H2;
58 k = k+1;

    end
60 end
62 end

```

## E.7. Cálculo coeficiente de correlación de Spearman y *p-value*

```

function [rho, pvalue] = correlationSpearman(PropMacro, Descript)
2 % Funcion que entrega el valor de los coeficientes de correlacion
% entre descriptores y propiedades macroscopicas y ademas, el
4 % p-value asociado a cada uno de ellos.
% INPUT:
6 % PropMacro: Matriz con los valores de las propiedades
% macroscopicas definidas.
8 % Descript: Matriz de descriptores

10 [~,n] = size(PropMacro);
[~,m] = size(Descript);

12 for i=1:n
14     Y = PropMacro(:,i);
    for j=1:m

```

```

16     X = Descript(:,j);
17     [rho(j,i), pvalue(j,i)] = ...
18         corr(X,Y, 'type', ...
19             'Spearman'); % Almacenar coeficientes y pvalue
20                             % en filas para cada descriptor
21                             % y columnas las prop macro
22     end
23 end
24 end

```

## E.8. Rutina para análisis cualitativo

```

1 function name = GrafCuali(D,PM,nD,nPM)
2 % Funcion que realiza todos los graficos entre propiedades macroscopicas y
3 % descriptores para asi estudiar la relacion que existe entre ellas. Estos
4 % son guardados en archivos .pdf.
5 % INPUT:
6 %     D: Matriz de descriptores (sin CAS en la primera columna).
7 %     PM: Matriz de propiedades macroscopicas (sin CAS).
8 %     nD: Vector con el nombre de los descriptores.
9 %     nPM: Vector con el nombre de las propiedades macroscopicas
10 % OUTPUT:
11 %     name: Nombre de los archivos .pdf que se almacenan los graficos.
12
13 [~,n] = size(D);
14 [~,m] = size(PM);
15 k = 1;
16 for i=1:m
17     for j=1:n
18         h = figure('visible','off');
19         if i==3
20             semilogy(D(:,j),PM(:,i),'o')
21         else
22             plot(D(:,j),PM(:,i),'o')
23         end
24         xlabel(nD{j},'FontSize',12)
25         ylabel(nPM{i},'FontSize',12)
26         %title('Relacion entre Descriptores y Propiedades Macroscopicas')
27         name{k} = strcat(nD{j},nPM{i},'.pdf');
28         print(h,'-dpdf',name{k});
29         k = k+1;
30         close(h);
31     end
32 end
33 end

```

## E.9. Rutina para agrupar descriptores

```
function [PM1, PM2, PM3] = addMatrix(i,j,Descrip)
2 % Funcion que agrupa los descriptores por propiedad macroscopica con la que
  % estan mas relacionados (segun coef de correlacion)
4 %INPUT:
  %     i: vector con el numero del descriptor segun orden en matriz de
  %     descriptores
6 %     j: vector con el numero de la propiedad macroscopica segun orden en
  %     matriz de propiedades macroscopicas
8 %     Descrip: matriz de descriptores.
10 %OUTPUT:
  %     PM1: matriz con los descriptores para la propiedad macroscopica 1
  %     (eficiencia). Las columnas son los descriptores, mientras que las
  %     filas son los compuestos.
12 %     PM2: idem, pero para propiedad macroscopica 2 (emisiones).
14 %     PM3: idem, pero para propiedad macroscopica 3 (costos).
16 n = length(i);
  PM1 = [];
18 PM2 = [];
  PM3 = [];
20
  for k=1:n
22     if j(k)==1
        PM1 = [PM1 Descrip(:,i(k))];
24     end
26     if j(k)==2
        PM2 = [PM2 Descrip(:,i(k))];
28     end
30     if j(k)==3
        PM3 = [PM3 Descrip(:,i(k))];
32     end
  end
end
```