

ANÁLISIS DE SEMEJANZA EN LAS ASOCIACIONES MOLECULARES DE LA AMILOSA CON ALGUNOS COMPUESTOS ORGANOLÉPTICOS

Elsa Sánchez-García y Luis A. Montero*

Laboratorio de Química Computacional y Teórica, Facultad de Química, Universidad de La Habana, Cuba

RESUMEN

En el presente trabajo se realiza un análisis de semejanza de estructuras moleculares, con vistas a escoger el mejor valor límite del coeficiente de Tanimoto. Se desea obtener los mejores valores de propiedades termodinámicas basadas en consideraciones estadísticas correspondientes a asociaciones moleculares en sistemas formados por la amilosa y algunos compuestos organolépticos como el eugenol demetilado, el linalol, la vainillina y el benzoato de bencilo. El análisis de semejanza se usa con el fin de evitar estructuras redundantes en simulaciones que parten de estructuras aleatorias "a priori". Tales redundancias provocan desviaciones no deseadas en los valores de las funciones de partición y afectan la calidad de las simulaciones si no se eliminan adecuadamente. Para simular el comportamiento de la amilosa se utilizó la dextrina formada por tres unidades de α 1,4 -glucosa. Se demostró que para estos sistemas el valor idóneo de discriminación de Tanimoto es 0.95. Esto se debe a que, aunque los valores de las funciones termodinámicas de asociación no difieren si se trabaja con $T = 0.85$ o con $T = 0.95$, con el primero se consideran como iguales celdas que en realidad son diferentes, lo cual se debe a las múltiples posibilidades de ordenamientos espaciales diferentes en la molécula de dextrina estudiada.

ABSTRACT

In the present work an analysis of similarity of molecular structures is made, with views to choose the best value limit of the coefficient of Tanimoto. It is desired to obtain the best values of thermodynamic properties based on statistical considerations corresponding to molecular associations in systems formed by the amilosa and some organolépticos compounds like eugenol demetilado, linalol, the vainillina and the benzoate of bencilo. The similarity analysis is used with the purpose of avoiding redundantes structures in simulations that leave from random structures "a priori". Such redundancies cause deviations wished in the values of the partition functions and affect the quality of the simulations if they are not eliminated suitably. In order to simulate the behavior of the amilosa the dextrin formed by three units was used of α 1,4 - glucose, obtaining that for these systems the suitable value of coefficient of Tanimoto is 0.95. This must to that, although the values of the thermodynamic functions of association do not differ if works with $T = 0.85$ or $T = 0.95$, with first are considered like equal cells that in fact are different, which must to the multiple possibilities of space orderings different in the studied dextrin molecule.

INTRODUCCIÓN

La amilosa es un polisacárido formado por varias unidades de α 1,4 -glucosa que constituye un importante soporte cromatográfico y ha sido objeto de diferentes estudios debido, entre otras características, a su capacidad para formar complejos de inclusión^{1,2}. La amilosa es utilizada para la retención de diferentes tipos de sustancias, como por ejemplo los compuestos organolépticos. Estos compuestos presentan múltiples aplicaciones entre las que podemos citar la actividad escabicida y pediculicida del benzoato de bencilo, el poder saborizante y la actividad antioxidante de la vainillina, así como las propiedades analgésicas y antioxidantes del eugenol^{3,5}.

El análisis de semejanza^{6,8} se ha convertido en una importante herramienta en áreas como el diseño molecular y la química aplicada a la medicina donde una premisa central validada por la experiencia es que moléculas que son estructuralmente similares tienen actividades biológicas similares⁷. Este análisis comprende la especificación de una estructura determinada caracterizada por un conjunto de uno o más descriptores que son a su vez comparados con el correspondiente conjunto de descriptores de las otras estructuras a comparar. Los coeficientes de semejanza proporcionan entonces una medida cuantitativa del grado de relación estructural entre cada par de representaciones⁹. Algunos coeficientes son medidas de distancias o de falta de semejanza entre dos objetos mientras otros miden la semejanza

directamente. En la mayoría de los casos un coeficiente de semejanza puede tomar valores entre 0 y 1 o puede ser normalizado en dicho rango⁹.

La idea de determinar una medida numérica de semejanza entre dos objetos, cada uno caracterizado por un conjunto común de atributos, es común en varias disciplinas. Por esta razón algunos coeficientes de semejanza tienen características en común como es el caso del coeficiente de asociación de Tanimoto (de Jaccard) cuyo complemento es el coeficiente de distancia de Soergel y es monótono con el coeficiente de Dice⁹. Para variables dicotómicas, el coeficiente $S_{A,B}$ de Tanimoto para medir la semejanza entre un sistema A y otro B se calcula mediante la expresión:^{9,10}

$$S_{A,B} = \frac{c}{a+b-c} \quad (1)$$

y toma valores entre 0 y 1. En esta expresión a y b son propiedades comparables correspondientes a cada uno de los objetos A y B, respectivamente, mientras que c es una variable que se obtiene de asociar las propiedades a y b en las mismas condiciones que ellas individualmente. A mayor semejanza entre los sistemas, entonces c será más similar al valor de a o de b, éstas tenderán a ser idénticas entre si, y la expresión tiende al valor unitario.

En el presente trabajo se realiza un análisis de semejanza, con vistas a escoger el mejor valor límite del coeficiente de Tanimoto para el estudio de los valores termodinámico-estadísticos de las interacciones moleculares en sistemas formados por la amilosa y los compuestos organolépticos eugenol demetilado, linalol, vainillina y benzoato de bencilo.

METODOLOGÍA DE TRABAJO

Con vistas a hacer menos complejo el sistema objeto de estudio, se seleccionó la dextrina formada por tres unidades de α 1,4-glucosa (Figura 1), con el objetivo de reproducir el comportamiento de la amilosa, que como ya habíamos mencionado, es un polisacárido lineal de alto peso molecular formado por varias unidades de α 1,4 -glucosa.

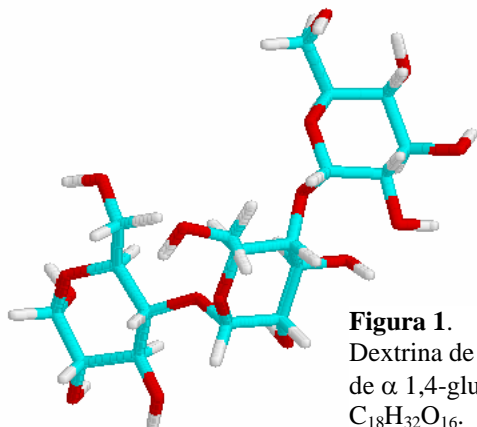


Figura 1.
Dextrina de tres unidades
de α 1,4-glucosa
 $C_{18}H_{32}O_{16}$.

Para nuestro trabajo utilizamos la aproximación **MMH** (Hipersuperficies de Múltiples Mínimos) cuyos fundamentos aparecen descritos en otros artículos ya publicados^{11,12} **MMH** combina procedimientos químico-cuánticos para el cálculo de las energías de los mínimos locales con la estadística para la determinación de las funciones termodinámicas de asociación.

En términos estadísticos¹² y tomando como estado de referencia a las moléculas no asociadas, la energía interna, la entropía y la energía libre de Helmholtz de asociación de las moléculas pueden ser expresadas respectivamente como:

$$\Delta E_{\text{ass}} = E - \varepsilon^{\text{ref}} = RT^2 \frac{q^{**}}{q^*} \quad (2)$$

$$S_{\text{ass}} = R \ln q^* + \frac{\Delta E_{\text{ass}}}{T} \quad (3)$$

$$\Delta F_{\text{ass}} = F - \varepsilon^{\text{ref}} = -RT \ln q^* \quad (4)$$

donde T es la temperatura en Kelvin, R la constante de los gases expresada en J/mol·K y q^* es la función de partición molecular o sea, la suma de los pesos estadísticos de los estados correspondientes al conjunto dado, que se expresa como:

$$q^* = \sum_i g_i e^{\frac{-\Delta \varepsilon_i}{RT}} \quad (5)$$

Se toma como estado de referencia convencional una celda con el mismo número de moléculas aunque no interactuantes y en el mismo estado rotovibracional y traslacional que las moléculas asociadas. Por tanto, la energía de la celda respecto a la escala de referencia es:

$$\Delta \varepsilon_i = \varepsilon_i - \varepsilon^{\text{ref}} = \varepsilon_i - (\varepsilon_{\text{dextrina}} + \varepsilon_{\text{organoléptico}}) \quad (6)$$

siendo $\Delta \varepsilon_i$ el término que aparece en la ecuación (3).

De este modo, en cada caso se construyen agregados moleculares, o celdas posibles del espacio de configuraciones en ese conjunto estadístico, formados por la molécula de dextrina y el compuesto organoléptico eugenol demetilado, vainillina, linalol o benzoato de bencilo (Figura 2). Para el cálculo de geometrías y energías de estos sistemas se utiliza el programa MOPAC con el hamiltoniano semiempírico PM3.^{13,14}

El estudio de estos agregados moleculares implica un gran esfuerzo computacional, partiendo de la generación de un mínimo de alrededor de 25 celdas para cada sistema molecular dextrina-compuesto organoléptico donde estos aparecen en posiciones seleccionadas aleatoriamente y, obviamente, no optimizadas¹⁵. Luego se lleva a cabo la optimización de la geometría (búsqueda en el “camino” de los

gradientes de cada mínimo en la correspondiente superficie de energía potencial con respecto a las coordenadas nucleares del sistema) y la consiguiente definición de las energías de cada arreglo molecular, hasta encontrar el conjunto de mínimos locales con el cual sea posible la descripción estadística del sistema. Una vez encontrado dicho conjunto se lleva a cabo el procesamiento estadístico¹⁵ de estos resultados para encontrar las funciones termodinámicas de asociación para dicho sistema. Luego, se demuestra la convergencia de las energías de asociación y es posible buscar aquellas celdas más estables en cada caso para así describir las interacciones en estos sistemas moleculares. Es en este proceso donde puede ocurrir una redundancia de estructuras degeneradas que no tienen significación en la descripción del sistema. Si desde más de una estructura aleatoria inicial, el camino de los gradientes llega a un mismo mínimo (o aproximadamente el mismo) las estructuras repetidas representan degeneraciones que no tienen nada que ver con la descripción del sistema, sino con la simple casualidad de que la estructura de partida condujera a situaciones intermoleculares similares.

ANÁLISIS DE SEMEJANZA

Se llevó a cabo el análisis de semejanza con vistas a escoger el mejor coeficiente de Tanimoto para este tipo de sistemas. Es necesario ante todo tener en cuenta que en estas superficies de energía potencial con respecto a las coordenadas nucleares, o hipersuperficies, la degeneración puede ser de dos tipos:

Aquellas celdas que tienen la misma energía y el mismo ordenamiento molecular, o sea, las que se pueden considerar iguales tanto desde el punto de vista energético como geométrico. Este tipo de degeneración es generalmente redundante ya que no aporta información nueva a la función de partición. Debe destacarse que en el caso de sistemas con un cierto nivel de simetría pueden existir degeneraciones de este tipo que NO son redundantes si los átomos son distinguibles, como es el caso. Estas redundancias SI son significativas y no deben ser descartadas.

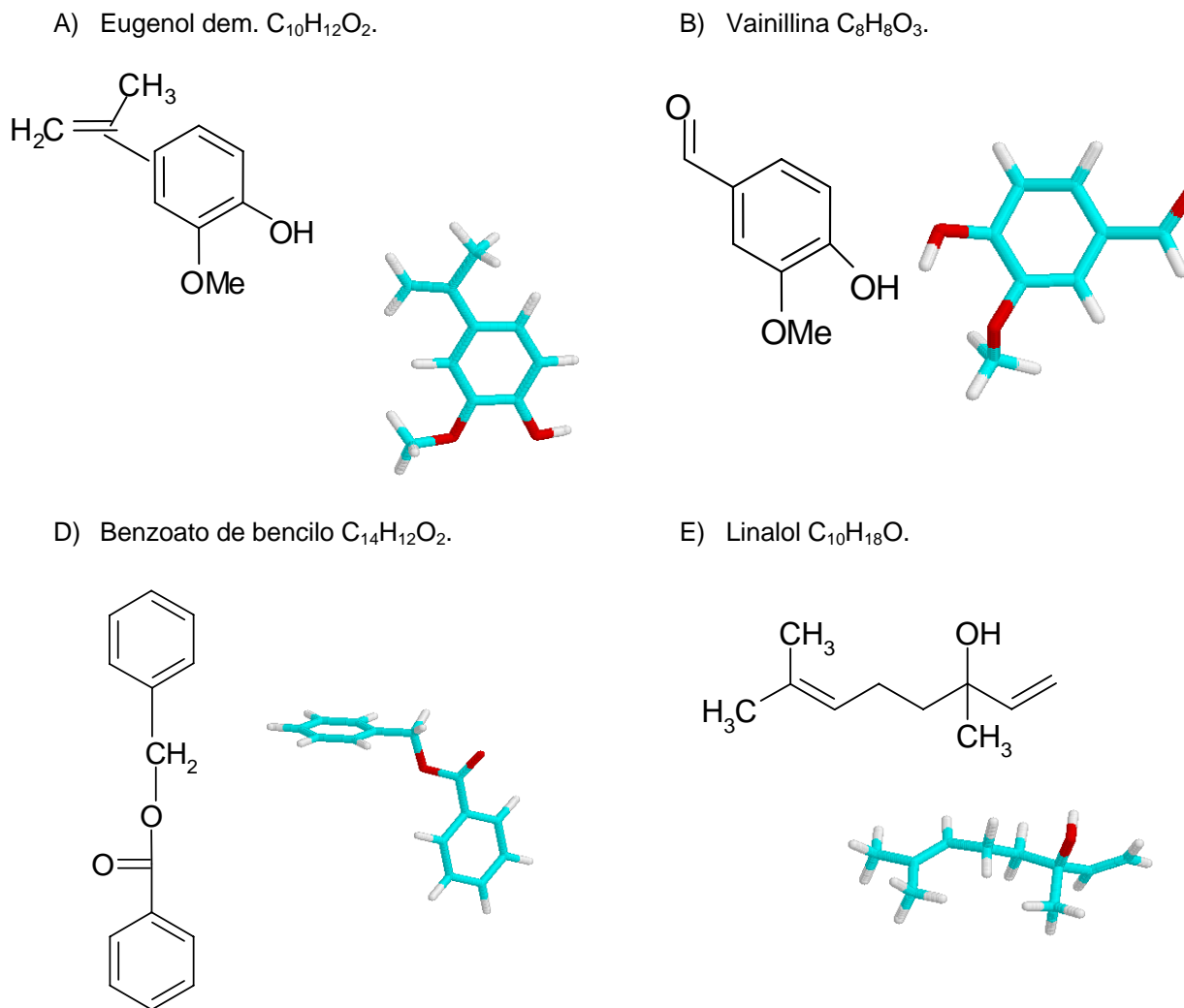


Figura 2. Compuestos organolépticos estudiados.

Aquellas que tienen diferentes ordenamientos moleculares pero la misma energía, lo que sí representa una degeneración significativa. Entonces, si las celdas son realmente diferentes deben ser consideradas en la función de partición, aunque tengan valores de energías muy similares. Este es el caso de conformaciones similares en sistemas simétricos cuando los átomos son distinguibles.

Ambos tipos de degeneración aparecen frecuentemente, sobre todo en clusters pequeños, que como es de esperar tienen menos posibilidades de presentar ordenamientos moleculares diferentes. Como consecuencia de este análisis el programa Q2¹⁵ de procesamiento estadístico contiene una subrutina llamada Tanimoto especialmente diseñada con este propósito. Esta subrutina analiza la semejanza entre arreglos moleculares para desechar la degeneración del tipo i). El procedimiento de Tanimoto calcula la semejanza entre las estructuras par a par mediante la comparación de todos los clusters con una diferencia de energía menor de 0.096 kJ/mol con respecto a una anterior (este valor se toma teniendo en cuenta la confiabilidad estimada para un hamiltoniano semiempírico).

Luego se convierten las coordenadas de cada átomo de internas a cartesianas. De este modo, se construye la matriz [D] de los módulos de los vectores de posición con respecto al origen de coordenadas fijado en el átomo 1.

$$[D] \equiv [r_1, r_2, r_3, \dots, r_N] \quad (7)$$

donde N es el número total de átomos..

El índice de Tanimoto correspondiente a la comparación de las celdas A y B se determina mediante la expresión:

$$T = \frac{M}{(A + B - M)} \quad (8)$$

donde A, B, M se calculan mediante las siguientes expresiones:

$$M = \sum_{i=1}^N r_i^A r_i^B \quad (9)$$

$$A = \sum_{i=1}^N r_i^A r_i^B \quad (10)$$

$$B = \sum_{i=1}^N r_i^B r_i^B \quad (11)$$

r_i^A es un elemento de $[D_A]$ y r_i^B de $[D_B]$ correspondientes a las estructuras A y B, respectivamente.

Entonces, fijamos un valor límite de discriminación para considerar las celdas iguales o no desde el punto de vista geométrico. Por ejemplo si el valor calculado de T (índice de Tanimoto) es mayor que 0.85 (valor límite fijado empíricamente) se puede considerar que esos arreglos moleculares son equivalentes y estamos en presencia de clusters geoméricamente degenerados. Así, si T es menor de 0.85 los clusters son diferentes y estamos en presencia de degeneración del tipo ii), que sí contribuye a la función de partición.

RESULTADOS

En este caso probamos con un valor inicial de $T = 0.85$, siguiendo experiencias previas con otros sistemas estudiados¹⁶, sin embargo se demostró que debido a la posibilidad de la dextrina de adoptar diferentes ordenamientos espaciales, con este valor de T se eliminaban como iguales ordenamientos moleculares que en realidad eran diferentes. Por ello se llevó a cabo el siguiente análisis:

Primeramente comparamos los valores calculados de E_{ass}, F_{ass} y S_{ass} con $T = 0.85$ y con $T = 0.95$. Como se observa en las Figuras 9, 10 y 11 (Tablas I-III), los resultados obtenidos en cada caso no difieren prácticamente; por lo cual, desde el punto de vista del resultado final, podríamos utilizar, en principio, cualquiera de estos valores límites de discriminación. Llama la atención que al refinar el índice de Tanimoto a valores superiores, la entropía de asociación aumenta como producto de la consideración de más estados degenerados, lo que es una consecuencia esperada y confirmatoria de las consideraciones estadísticas en las que se fundamenta el análisis. En la Tabla IV vemos la cantidad de celdas degeneradas para cada sistema a $T = 0.85$ y $T = 0.95$ y es posible comprobar que no difieren mucho. La mayor diferencia se observa para el caso del linalol, sin embargo, como se puede observar claramente en las Figuras 9-11, no afecta significativamente los valores calculados.

Tabla I. Valores calculados de energía de asociación para cada sistema dextrina-compuesto organoléptico.

	E _{ass} (kJ/mol)	
	T = 0.85	T = 0.95
eugenol demetilado	-29.52	-29.51
vainillina	-30.17	-29.95
benzoato de bencilo	-31.66	-31.60
linalol	-31.44	-31.34

Sin embargo, se hace necesario hacer un análisis más profundo considerando si aquellas celdas eliminadas como iguales geométrica y energéticamente, en realidad lo son.

Tabla II. Valores calculados de energía libre asociación para cada sistema dextrina-compuesto organoléptico.

	F_{ass} (kJ/mol)	
	T = 0.85	T = 0.95
eugenol demetilado	-33.38	-33.38
vainillina	-32.82	-32.92
benzoato de bencilo	-35.09	-35.10
linalol	-35.12	-35.15

Tabla III. Valores calculados de entropía de asociación para cada sistema dextrina-compuesto organoléptico.

	S_{ass} J/mol K	
	T = 0.85	T = 0.95
eugenol demetilado	12.95	12.98
vainillina	8.88	9.95
benzoato de bencilo	11.52	11.74
linalol	12.36	12.80

Tabla IV. Cantidad de celdas degeneradas para cada sistema dextrina-compuesto organoléptico.

	# de celdas degeneradas	
	T = 0.85	T = 0.95
eugenol demetilado	2	1
vainillina	4	1
benzoato de bencilo	6	0
linalol	7	2

Por ejemplo, las celdas dextrina-eugenol 9 y 18 son eliminadas como degeneradas tanto energética como geométrica, al utilizar el valor límite del coeficiente de Tanimoto 0.85. Sin embargo, al comparar las Figuras 3 y 4 vemos a simple vista como estas celdas realmente son diferentes desde el punto de vista geométrico, por lo cual es necesario trabajar con un coeficiente de semejanza mayor. De este modo escogemos un valor límite de $T = 0.95$. Se selecciona debido a que, lógicamente, debe ser mayor que 0.85 y en realidad la mayor parte de las celdas eliminadas en cada caso tenían valor de T' que bien podía ser mayor que 0.95 o mantenerse muy próximo a 0.85. En las Figuras 5 y 6 se muestran las dos celdas que resultan degeneradas con este nuevo valor del coeficiente de Tanimoto. Así confirmamos, apoyados por algunas de las distancias entre átomos más importantes, que las celdas dextrina-eugenol 9 y 20 son realmente degeneradas.

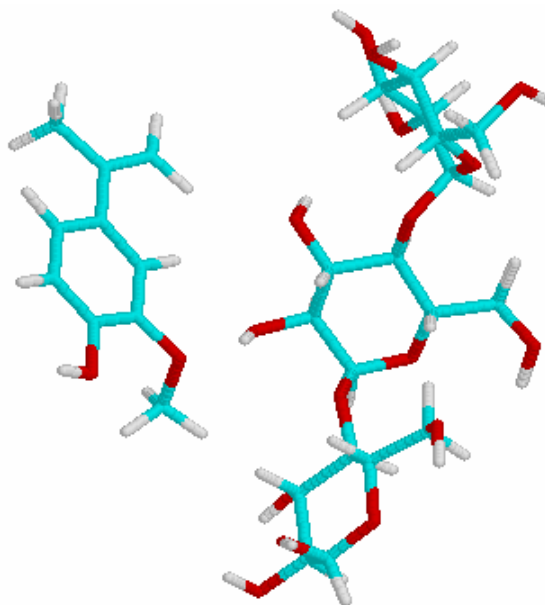


Figura 3. Celda 9 dextrina-eugenol.

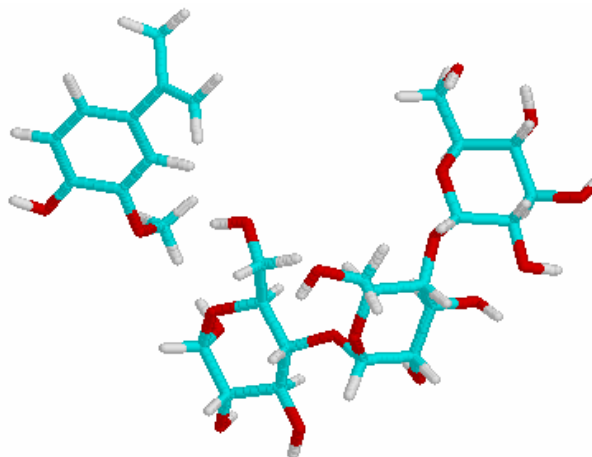


Figura 4. Celda 18 dextrina-eugenol.

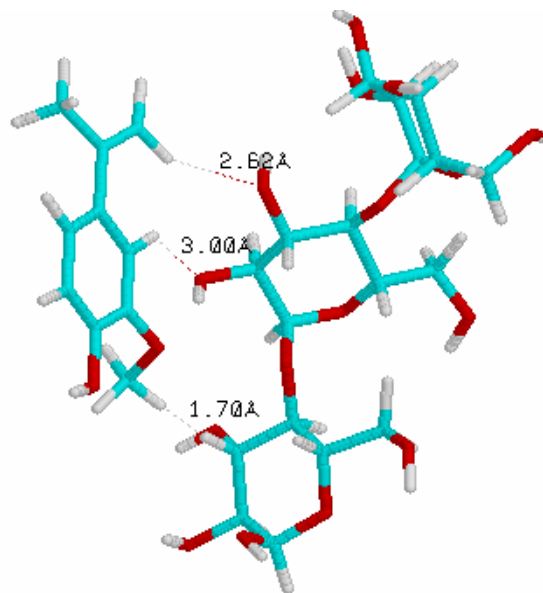


Figura 5. Celda 9 dextrina-eugenol.

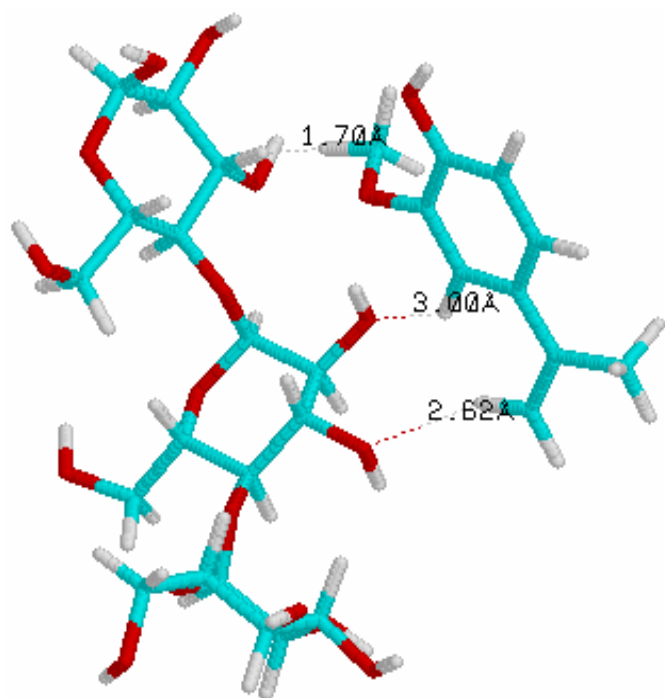


Figura 6. Celda 20 dextrina-eugenol.

Otro ejemplo se muestra en las Figuras 7 y 8 donde aparecen dos celdas dextrina-vainillina que se consideran degeneradas utilizando un valor límite de $T = 0.95$. Hemos señalado en dichas figuras algunas de las distancias más significativas con el objetivo de corroborar la similitud entre dichas celdas. De esta forma es posible comprobar que son también degeneradas geoméricamente.

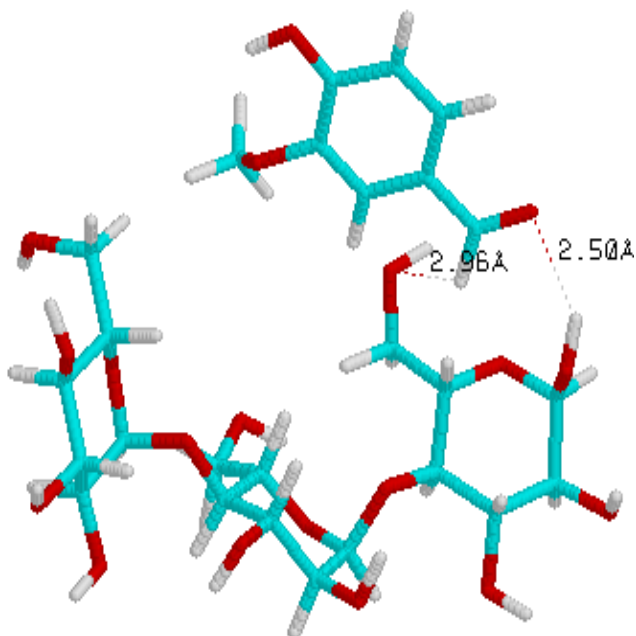


Figura 7. Celda 1 dextrina-vainillina.

CONCLUSIONES

Al realizar el análisis de semejanza en los sistemas moleculares dextrina-compuesto organoléptico, fue posible demostrar que en este caso el valor límite de discriminación de Tanimoto idóneo fue de 0.95. Se observó que, aunque los valores de las funciones termodinámicas de asociación, no diferían si se trabajaba tanto con $T = 0.85$ como con $T = 0.95$, con el primer valor de T se consideraban como iguales celdas diferentes en realidad. Esto es de esperar debido a que, desde el punto de vista estructural, es necesario considerar las múltiples posibilidades de ordenamientos espaciales diferentes, presentes en la dextrina estudiada, debido a su relativa mayor complejidad estructural con respecto a otros sistemas moleculares más sencillos.

No obstante lo anteriormente expuesto es preciso señalar que este trabajo demuestra que en este análisis debe partirse de consideraciones racionales en el momento de establecer números límite y cualquier tipo de criterio de convergencia. Por ejemplo, carece de sentido llevar la convergencia de la suma estadística de las energías en las simulaciones a límites inferiores a la propia expectativa de predicción energética del hamiltoniano seleccionado para la hipersuperficie. Igualmente, una semejanza mayor que este límite es seguramente una semejanza a los efectos del presente método. Todo ello también debe condicionarse al tamaño (número de átomos y moléculas) del sistema y a la simetría del mismo. El índice propuesto de 0.95 es un valor de referencia con buenas expectativas de éxito.

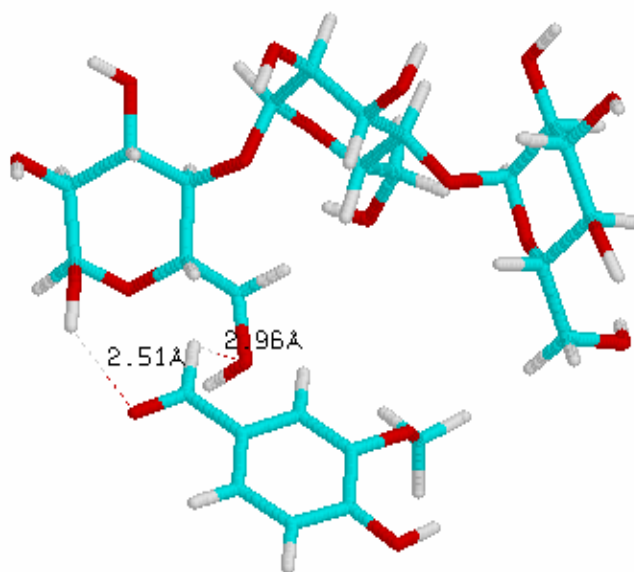


Figura 8. Celda 5 dextrina- vainillina.

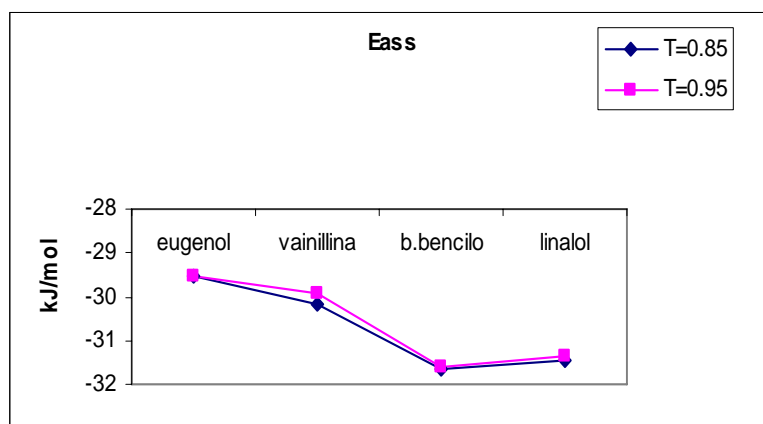


Figura 9. Valores calculados de energía de asociación para cada sistema dextrina-compuesto organoléptico.

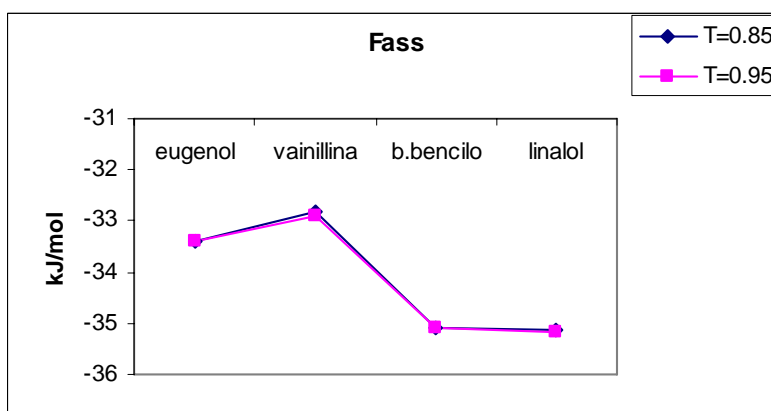


Figura 10. Valores calculados de energía libre de asociación para cada sistema dextrina-compuesto organoléptico.

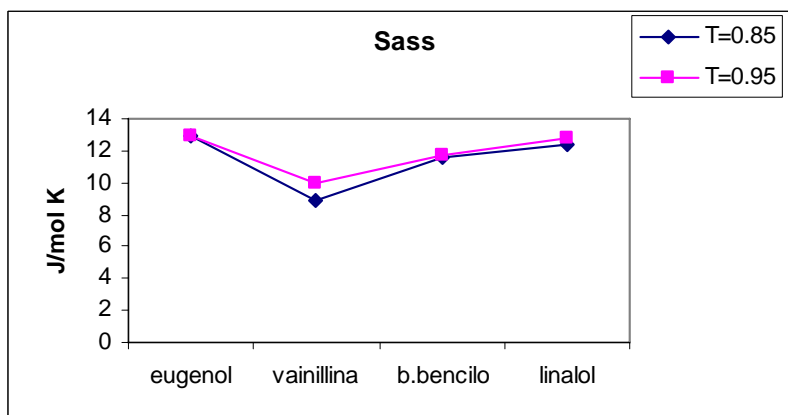


Figura 11. Valores calculados de entropía de asociación para cada sistema dextrina-compuesto organoléptico.

AGRADECIMIENTOS

Los autores desean agradecer a los revisores de la Revista Cubana de Física por sus acertados comentarios. Este trabajo se ha realizado bajo

los auspicios de la Universidad de La Habana y del Servicio Alemán de Intercambio Académico (DAAD). La asesoría del Dr. Roberto Cruz ha sido decisiva para la utilización de técnicas de semejanza.

REFERENCIAS

- (1) WULFF A., G.S. and O. HOELLER (1998): **Carbohydr. Res.** 307, 19-32.
- (2) SHIMADA H., J.K.; T. TAKADA; S. KITAMURA and K.J. KAJIWARA (2000): **Phys. Chem. B** 104, 2136 - 2147.
- (3) KIKUZAKI S., H.H.; Y. KAWAI and N. NAKATANI (1999): **Phytochemistry**, 52, 1307-1312.
- (4) OGATA M., M.H.; K. SHIMOTOHNO; S. URANO and T.J. ENDO (1997): **Am. Oil Chem Soc.**, 74, 557-562.
- (5) AHMED S., M.A. **et al.** (2000): **Pharmazie**, 55, 314 - 316.
- (6) VARMUZA, K.; M. KARLOVITS and W. DEMUTH (2003): **Analytica Chimica Acta**, 490, 313-324.
- (7) MARTIN, Y.C.; J.L. KOFRON and L.M. TRAPHAGEN (2002): **Journal of Medicinal Chemistry** 45, 4350-4358.
- (8) HOLLIDAY, J. D.; N. SALIM and P. WILLETT (2005): **ACS Symposium Series**, 894, 77-95.
- (9) WILLET, P.J. (1998): **Chem. Inf. Comput. Sci.**, 38, 983-996.
- (10) FISANICK, W.L. and A.H.; RUSINKO III (1994): **J. Chem. Inf. Comput Sci.**, 34, 130-134.
- (11) MONTERO E., L.A. **et al.** (1998): **A J. Am. Chem. Soc.**, 120, 12023-12033.
- (12) MONTERO M., L. A. and J. FABIAN (2000): **Int. J. Quant. Chem.**, 79, 8-16.
- (13) STEWART P., J.J. (1993): MOPAC v. 6 release for PC computers in the Laboratory of Computational and Theoretical Chemistry Universidad de La Habana and v. 7 implemented for IBM R6000/AIX and Linux systems.
- (14) STEWART P., J.J. (1989): **Journal of Computational Chemistry**, 10, 209-264.
- (15) MONTERO, L. A., GRANADA and Q Programs for PC computers Available on request 1996-2000.
- (16) SÁNCHEZ, E. M., L.A.; J.M. HERMIDA; R. CRUZ y G. GONZÁLEZ (2000): **Revista Cubana de Física**, 17, 41-45.