

# EN BUSCA DE UNA METODOLOGÍA PARA LA INCLUSIÓN DE SOLVENTE EN CÁLCULOS DE ESTRUCTURA ELECTRÓNICA PARA COMPUESTOS CON PROPIEDADES ÁCIDO-BASE

## LOOKING FOR A METHODOLOGY FOR THE INCLUSION OF SOLVENT IN ELECTRONIC STRUCTURE CALCULATIONS FOR COMPOUNDS WITH ACID-BASE PROPERTIES

Yenifer Cubillos<sup>1</sup>, Ismael Ortiz Verano<sup>2</sup>

<sup>1</sup> Estudiante U.D.C.A, <sup>2</sup> Químico, Msc Química, Universidad de Ciencias Aplicadas y Ambientales UDCA, Facultad de Ciencias, Programa de Química, Calle 222 No. 55 – 37. Bogotá, Colombia. yencubillos@udca.edu.co

### RESUMEN

Se estudió teóricamente la inclusión de solvente “agua” en el cálculo de propiedades termodinámicas mediante el modelado de indicadores ácido base. Se decidió seguir el método de solvente implícito, para después calcular las propiedades termodinámicas de algunos indicadores ácido-base ampliamente usados en laboratorios de química general y de esta forma determinar la constante de acidez de manera teórica. Este estudio se realizó en los paquetes computacionales Mopac y Gamess, a través de la interfaz Gabedit. Los resultados obtenidos fueron confrontados por los reportados por Patterson [1].

**Palabras clave:** Solvente implícito, constante de acidez, indicadores ácido-base, Mopac, Gamess.

### SUMMARY

The inclusion of water as a solvent on the calculation of thermodynamic properties was studied theoretically by modeling some acid-base indicators. Implicit solvent model was selected for calculating thermodynamical properties of widely used acid-base indicators in general chemistry labs, and then determine the acidity constant from theoretical data. In this study the computational packages Mopac and Gamess were used, using the graphical interface Gabedit. Obtained results were confronted with the experimental ones published previously by Patterson [1].

**Keywords:** Implicit solvent, acidity constant, acid-base indicators, Mopac, Games.

### INTRODUCCIÓN

El rápido desarrollo de la tecnología computacional ha permitido al campo de la química computacional convertirse en una rama firmemente establecida en investigación a nivel académico e industrial [2]. La comunidad académica se beneficia de esta tecnología gracias a la disponibilidad de software de modelado y la disminución de costos en los equipos de cómputo que puedan correr estos programas [3].

Esta tecnología está usándose para aumentar el aprendizaje de estudiantes a través de la visualización de estructuras y fenómenos químicos, sin embargo los estudios computacionales generalmente no se llevan a cabo en condiciones reales, de tal forma que no se describe totalmente el comportamiento químico de una muestra a menos que el sistema real se encuentre en fase gaseosa. Para superar este obstáculo, se pueden incluir en el sistema de estudio las moléculas del solvente, lo cual se denomina “solvente explícito”, con la consecuencia de tratar un sistema con un número considerablemente mayor de grados de libertad y por lo tanto más complejo, para el cual se requeriría mucho más tiempo de cálculo, espacio de almacenamiento en disco, memoria RAM, etc. Es decir, mayores recursos computacionales. Otro tipo de metodología de inclusión del solvente se denomina “solvente implícito”, en la cual se representa al solvente como un medio continuo y se considera el efecto promedio de muchas moléculas de solvente por medio de diferentes funciones analíticas.

Actualmente se dispone de diferentes metodologías para el tratamiento de solvente implícito en cálculos de estructura electrónica, aunque no todas las metodologías están dispo-

nibles en un solo paquete computacional, por lo que generalmente se hace necesario el uso de más de uno [4]. Las simulaciones computacionales de sistemas químicos con inclusión de agua de manera explícita representan los acercamientos teóricos más detallados para estudiar la influencia del solvente en la conformación de moléculas biológicas complejas [5]. Por otro lado los modelos de solvente implícito brindan una relevante eficiencia computacional mediante la reducción de interacciones soluto explícito solvente a las características de su campo, lo cual se expresa como una función de la configuración del soluto de manera individual [6]. Los métodos de solvente implícito o continuo reducen los grados de libertad del sistema usando una aproximación “promedio” de las interacciones del solvente e iones móviles con la biomolécula [7]. El componente polar de los modelos continuos usualmente describe la respuesta dieléctrica del solvente a la distribución de carga de un soluto. Los modelos dieléctricos continuos modernos pueden libremente agruparse en dos categorías básicas: modelo Generalizado de Born (GB) y la aproximación de Poisson–Boltzmann (PB) [8].

La base de la teoría PB es la aproximación del solvente como un continuo dieléctrico, que es el supuesto de una polarización lineal y local del solvente debido a un campo aplicado. Bajo esta aproximación, es posible derivar la ecuación de Poisson desde la teoría de mecánica estática y aplicarla a líquidos simples [9].

La ecuación de Poisson asume una distribución de carga (usualmente la carga del soluto) en un continuo dieléctrico; la ecuación PB generaliza esta aproximación del continuo eléctrico para incluir una distribución de carga móvil (p. ej., contra iones y co-iones) que responden al potencial electrostático. La teoría de PB asume que las cargas móviles siguen una distribución Boltzmann en la aproximación de campo medio [10] para dar una ecuación diferencial parcialmente no lineal para el potencial electrostático adimensional  $\phi(x)$ :

$$-\nabla \cdot \epsilon(x) \nabla \phi(x) - \sum_{i=1}^m c_i q_i e^{-q_i \phi(x) - v_i(x)} = \frac{4\pi e^2}{kT} \rho(x) \quad (1)$$

Donde los coeficientes son función de una posición  $x$ ;  $\epsilon(x)$  es el coeficiente dieléctrico describiendo la variación de la respuesta entre la relativamente baja polarizabilidad del soluto y la alta polarizabilidad del solvente,  $c_i$  es la concentración de especies iónicas móviles  $i$ ,  $q_i$  es la carga de especies iónicas móviles  $i$ ,  $v_i(x)$  describe las interacciones estéricas entre el soluto y las especies iónicas  $i$ ,  $e_c$  es la carga del electrón,  $k$  es la constante de Boltzmann,  $T$  es la temperatura absoluta y  $\rho(x)$  es la distribución de carga del solvente en unidades de electrones. La ecuación (1) representa la forma “completa” o no lineal de la ecuación PB [8].

El método analítico generalizado de Born (GB) supone otra alternativa para el cálculo del término electrostático de la energía libre de solvatación. Debido a su menor costo computacional, comparado con PB, esta metodología se ha convertido en un método bastante popular en dinámica molecular, para el reemplazo del solvente explícito [11].

A cada átomo de la molécula le corresponde una esfera de radio  $a_i$  con carga  $q_i$  centrada en el núcleo. En el interior del átomo, se asume un material dieléctrico de constante 1. La molécula está envuelta de un solvente de alta permisividad dieléctrica. La energía libre electrostática se determina a partir de la solvatación individual de Born para cada átomo, corregida por la perturbación del resto de átomos, según la ecuación (2).

$$\Delta G_{elec} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)} \quad (2)$$

Donde  $r_{ij}$  es la distancia entre cargas, y  $R_i$  y  $R_j$  son los denominados radios de Born (*effective Born radii*). Uno de los algoritmos más comunes para  $f_{GB}$  es la función desarrollada en la ecuación (3):

$$f_{GB} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)]^{\frac{1}{2}} \quad (3)$$

Uno de los parámetros más importantes es el valor de los radios de Born, ya que no son propiedades atómicas intrínsecas, sino que dependen de la conformación del soluto, por lo que se han de calcular tras cada cambio conformacional. Reflejan el grado de enterramiento de un átomo en el interior del soluto: para un átomo cercano a la superficie, los radios de Born son más pequeños, pudiendo igualarse al radio de van der Waals para aquellos átomos de cadena laterales totalmente expuestas al solvente. El cálculo de estos radios se deriva de los radios de van der Waals implementados en el *campo de fuerza* o de valores experimentales. A partir de este modelo, se han generado diversas modificaciones que afectan a la forma de la función  $f_{GB}$  y/o al modo en que se calculan los radios efectivos. En este sentido, cada vez se tiende a obtener funciones analíticas más rápidas, transferibles y que funcionen bien en los sistemas de estudio [12].

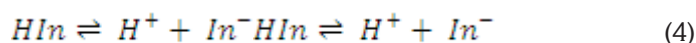
Los ejemplos más importantes del uso de métodos PB son tratamientos de proteínas en solución [13-14]. Desafortunadamente, los métodos de PB a menudo son lentos. Sin embargo, los métodos de GB también tienen sus debilidades. Por ejemplo, que la precisión del modelo GB depende críticamente de la precisión de los llamados radios Born que caracterizan las posiciones de cada carga atómica parcial respecto a la superficie molecular de la estructura [12].

El uso de modelos de solvente implícito es particularmente atractivo para las aplicaciones destinadas a las predicciones de estructura basada en la energía, para las que son críticas las evaluaciones de energía conformacional rápidas y precisas. Las descripciones con solvente implícito permiten la incorporación de términos que representan la energía libre de solvatación sin un promedio de más grados de libertad que en solventes explícitos en todos los átomos con funciones puntuales [6].

Una de las propiedades químicas que se ven muy afectadas por la naturaleza del solvente son las constantes de acidez, y es el objetivo de este estudio. El  $pK_a$  de una molécula o grupo funcional es una importante propiedad química y biológica, que ayuda en la determinación de la protonación a un pH dado así como el sitio de protonación de la molécula [15]. El  $pK_a$  es el logaritmo negativo de  $K_a$ .

El conocimiento de los valores de  $pK_a$  es importante para el tratamiento cuantitativo de sistemas que implican equilibrios ácido-base en solución, un ejemplo de estos sistemas son los indicadores ácido-base. Los indicadores son, generalmente, sustancias orgánicas de carácter ácido o básico débil, que tienen la propiedad de que su molécula y el ion correspondiente presentan coloraciones diferentes, es decir, el color del compuesto disociado es diferente del no disociado [16].

Una forma de calcular el  $pK_a$  para el sistema en equilibrio de un indicador ácido base como el naranja de metilo puede explicarse según la siguiente demostración, siendo HIn la representación de una molécula del indicador sin disociar,  $H^+$  su ácido conjugado e  $In^-$  su base conjugada:



La expresión para la constante de equilibrio de la reacción es:

$$K_a = \frac{[H^+][In^-]}{[HIn]} \quad (5)$$

Y si sabemos que hay una función de estado termodinámica fundamental que gobierna todos los procesos que se realizan a temperatura y a presión constante, conocida como la energía libre de Gibbs descrita por la siguiente ecuación:

$$\Delta G = -RT \ln K \quad (6)$$

$$K = e^{\frac{\Delta G}{RT}} \quad (7)$$

Podemos asumir que el cálculo de  $K_a$  puede efectuarse considerando la energía libre de Gibbs para la reacción de disociación.

$$\Delta G_{Reacción} = \sum \Delta G_{productos} - \sum \Delta G_{Reactivos} \quad (8)$$

$$\Delta G_{Reacción} = \sum \Delta G_{[H^+][In^-]} - \sum \Delta G_{[HIn]} \quad (9)$$

$$\Delta G = \Delta H - T\Delta S \quad (10)$$

Y convertirse la constante calculada por medio de la ecuación 7 a  $pK_a$  aplicando el logaritmo negativo.

$$pK_a = -\text{Log } K_a \quad (11)$$

## MATERIALES Y METODOS

Se propusieron los siguientes indicadores para calcular el  $pK_a$ : Naranja de metilo, rojo de metilo, rojo fenol, azul bromofenol, azul bromotimol, verde bromocresol, púrpura bromocresol y fenolftaleína; que son ampliamente usados en laboratorios de química general.

Para poder obtener los datos termodinámicos a usar en la ecuación (9) fue necesario inicialmente construir las estructuras tridimensionales de los indicadores neutros y en su forma básica o desprotonada. Estas estructuras fueron construidas en el programa Avogadro [17], y guardadas en el formato xyz.

Se llevó a cabo una preoptimización de la geometría de cada una de las estructuras construidas, usando el programa MOPAC 2012 [18] al nivel de teoría PM7. Posteriormente la geometría de cada una de los sistemas de estudio fue optimizada con el paquete computacional GAMESS [19], al nivel de teoría HF/6-31G. A cada una de las estructuras optimizadas se le llevó a cabo un análisis termodinámico mediante un cálculo de frecuencias a los mismos niveles de teoría a los que se hicieron las optimizaciones de geometría. A partir de los cálculos descritos anteriormente fue posible tener los datos termodinámicos necesarios para calcular las constantes de equilibrio.

Con el objetivo de evaluar el efecto del solvente sobre las constantes de equilibrio, los cálculos descritos en el párrafo anterior también se llevaron a cabo con la inclusión de solvente de manera implícita, usando la metodología COSMO en MOPAC 2012, y PCM en GAMESS.

## RESULTADOS Y DISCUSIÓN

La condición de equilibrio termodinámico en un sistema cerrado a temperatura constante viene definida en términos de la variación de la función de Gibbs (G), que es la función de estado termodinámica fundamental que gobierna los

procesos que se realizan a temperatura y a presión constantes, según lo expresado en la ecuación (10). El signo de  $\Delta G$  depende de los signos de los cambios en entalpía ( $\Delta H$ ) y entropía ( $\Delta S$ ), así como de la temperatura absoluta ( $T$ ) y de esta forma advierte la espontaneidad del sistema, si  $\Delta G$  es negativo significa que la reacción ocurre espontáneamente, por otra parte si el valor de  $\Delta G$  de la reacción es positivo, esta solo ocurrirá si se aplica energía al sistema para promover la reacción.

En las tablas 1 y 2 se muestran las energías libres de disociación calculadas usando la ecuación 9, y los valores calculados de  $pK_a$  para los 8 indicadores propuestos en este trabajo. Los resultados obtenidos fueron comparados con los reportados por Patterson usando un método espectrofotométrico [1]. Los valores experimentales de información termodinámica no estuvieron siempre disponibles, debido a que no se lograba la convergencia o el cálculo se detenía por falta de memoria del equipo.

## CONCLUSIONES

La dificultad en la simulación de estos sistemas recae en su gran dimensionalidad ya que al trabajar con un método semi-empírico, puede que, en el caso de los métodos semiempíricos, estos no se hayan parametrizado para los elementos más pesados y esta sea una causa de la no convergencia de algunas de las estructuras.

La metodología empleada no nos permite emitir un juicio sobre las ventajas de una metodología sobre otra debido a que las simulaciones no arrojaron resultados suficientes para ello.

Por otro lado, incorporar métodos computacionales al protocolo puede disminuir la problemática en la enseñanza de decidir que experimentos se pueden modelar y de esta manera disminuir los riesgos generados por manipulación de reactivos, de tal forma que se busca impulsar este campo de

Indicador	Sin Solvente		Con Solvente		pKa Reportada
	$\Delta G$ Reacción (J/mol)	pKa	$\Delta G$ Reacción (J/mol)	pKa	
Naranja de Metilo	-29305,697	5,14	-31552,963	5,53	3,43
Rojo de Metilo	26451,959	-4,64	71533,16	-12,54	4,9
Rojo Fenol	-25574,987	4,48	-26830,011	4,70	7,66
Azul Bromofenol	ND		ND		3,95
Azul Bromotimol	ND		ND		7
Verde Bromocresol	ND		ND		4,62
Purpura Bromocresol	ND		ND		6,19
Fenolftaleina	25686,634	-4,50	29608,733	-5,19	9,37

**Tabla 1.** Energía libre de Gibbs de reacción calculada con MOPAC 2012 sin solvente y con la inclusión de agua de forma implícita

Indicador	Sin Solvente		Con Solvente		pKa Reportada
	$\Delta G$ Reacción (J/mol)	pKa	$\Delta G$ Reacción (J/mol)	pKa	
Naranja de Metilo	-24148,294	4,23	-24097,623	4,22	3,43
Rojo de Metilo	ND		ND		4,9
Rojo Fenol	-19889,016	3,49	ND		7,66
Azul Bromofenol	ND		ND		3,95
Azul Bromotimol	ND		ND		7
Verde Bromocresol	ND		ND		4,62
Purpura Bromocresol	ND		ND		6,19
Fenolftaleina	ND		29240,467	-5,13	9,37

**Tabla 2.** Energía libre de Gibbs de reacción calculada con MOPAC 2012 sin solvente y con la inclusión de agua de forma implícita

la ciencia que actualmente no es tan explotado como podría serlo. Sin embargo, los sistemas con los que se pretende llevar a cabo la mejora en la enseñanza de la química deben ser seleccionados con mayor cuidado, y se deben llevar a cabo mayor número de pruebas a dichos sistemas.

**AGRADECIMIENTOS.** Este trabajo fue parte del Programa Jóvenes Investigadores de la Universidad de Ciencias Aplicadas y ambientales U.D.C.A, bajo el amparo del Semillero de Investigación en Química Teórica y Computacional, de la facultad de Ciencias.

## BIBLIOGRAFIA

- [1] Paterson, G.S. *J.Chem.Ed.* 1999 Vol 76 No.3. 395–398.
- [2] Krieger, J. H. *Chem. Eng. News* **1997**, 75(May 12), 30
- [3] Poon, T. An integrated molecular modeling and melting point experiment for the organic chemistry laboratory. *J.Chem.Ed.* 1999, Vol 76 No.7, p. 983.
- [4] Zheng, F., & Zhan, C. G. Computational modeling of solvent effects on protein-ligand interactions using fully polarizable continuum model and rational drug design. *Commun. Comput Phys.* 2013. 13, 31-60.
- [5] Brooks, C. L., Pettitt, B. M., & Karplus, M. *Advances in Chemical Physics, Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics.* John Wiley & Sons. 1990
- [6] Feig, M., & Brooks, C. L. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current opinion in structural biology*, 2004. 14(2), 217-224.
- [7] Roux B, Simonson T: Implicit solvent models. *Biophys Chem* 1999, 78:1-20
- [8] Baker, N. A. Improving implicit solvent simulations: a Poisson-centric view. *Current opinion in structural biology*. 2005. 15(2), 137-143.
- [9]. Beglov D, Roux B: An integral equation to describe the solvation of polar molecules in liquid water. *J Phys Chem B* 1997, 101:7821-7826.
- [10]. Holm C, Kekicheff P, Podgornik R (Eds): *Electrostatic Effects in Soft Matter and Biophysics.* Boston: Kluwer Academic Publishers; 2001.
- [11] Onufriev, A. (2013). The generalized Born model: its foundation, applications, and limitations.
- [12] Gracia, O. R. Herramientas de cribado virtual aplicadas a inhibidores de tirosina quinasas. Contribución al desarrollo del programa PRALINS para el diseño de quimiotecas combinatorias (2007). (Doctoral dissertation, Universitat Ramon Llull)
- [13] Honig, B., & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* (1995). 268(5214), 1144-1149.
- [14] Rogers, N. K. The modelling of electrostatic interactions in the function of globular proteins. *Progress in biophysics and molecular biology.* (1986). 48(1), 37-66.
- [15] Riojas, A. G., & Wilson, A. K. Solv-ccCA: Implicit Solvation and the Correlation Consistent Composite Approach for the Determination of pKa. *Journal of Chemical Theory and Computation*, (2014). 10(4), 1500-1510.
- [16] Teijón, J. M., García, J. A. Reacciones de Transferencia de Protones. En *Química: Teoría y problemas* (131-158). Madrid, España: Tebar. (1996)
- [17] Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics.* 2012. 4(1), 17.
- [18] Stewart, J. MOPAC2012, Stewart Computational Chemistry, Version 13.240L web: <http://OpenMOPAC.net>
- [19] M.W.Schmidt, K.K.Baldrige, J.A.Boatz, S.T.Elbert, M.S.Gordon, J.H.Jensen, S.Koseki, N.Matsunaga, K.A.Nguyen, S.J.Su, T.L.Windus, M.Dupuis, J.A.Montgomery. General Atomic and Molecular Electronic Structure System” *J.Comput.Chem.* 1993