



UNIVERSIDAD AUTÓNOMA METROPOLITANA – IZTAPALAPA

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

**“ESTUDIO COMPUTACIONAL DE
EQUILIBRIOS QUÍMICOS EN SOLUCIÓN:
ACIDEZ Y FORMACIÓN DE COMPLEJOS”**

Tesis que presenta

Q. AIDA MARIANA REBOLLAR ZEPEDA

Para obtener el grado de

DOCTOR EN CIENCIAS (QUÍMICA)

ASESORA:

DRA. ANNIA GALANO JIMÉNEZ

México, D.F., 5 de Septiembre de 2016

AGRADECIMIENTOS

A veces uno quisiera decir tantas cosas en la conclusión de esta etapa y no olvidar a nadie que fue importante para llevar a cabo este gran logro. Sin embargo, en este momento no tengo palabras, solo quiero decirles que es lo que pasa por mi mente y poder transmitirles todo lo que esta etapa significó para mí: desde el comienzo fue una gran emoción, un gran reto; en el desarrollo fue una batalla, existieron lagrimas y muchas alegrías; al final existió la incredulidad y una gran nostalgia por dejar este periodo de mi vida.

Es por eso, y muchas otras cosas que quiero agradecer a todos los que de una u otra forma han estado ahí para mí en los momentos en los que los he necesitado. Ya sea para apoyarme en el ámbito académico, personal o humano. Todo lo valoro y sepan que son parte de este gran logro.

Un agradecimiento especial a la profesora: Dra. Annia Galano Jiménez que, como directora de tesis, me ha orientado, apoyado y corregido en mi labor científica siempre con un interés y entrega que han sobrepasado, con mucho, todas mis expectativas como alumna. Para mí, ella es mi madre académica; siempre mostrando tanta pasión en todo lo que realiza. Y aunque probablemente le di varios dolores de cabeza, agradezco de todo corazón los años que la vida me ha dado la oportunidad de estar con ella.

Aída Mariana Rebollar Zepeda

TABLA DE CONTENIDOS

RESUMEN	3
ABSTRACT	4
ABREVIATURAS	5
1.- INTRODUCCIÓN	6
2.- ANTECEDENTES	11
3.- HIPÓTESIS	13
4- OBJETIVOS	13
4.1 Objetivo General	13
4.2 Objetivos Particulares	13
5- METODOLOGÍA	14
5.1. CONSTANTES DE ACIDEZ	15
5.1.1. Moléculas a estudiar	15
5.1.2. Esquemas de reacción	18
5.1.3. Métodos de cálculo	26
5.1.4. Método de correlaciones	31
5.1.5. Estimación de errores	32
5.2 CONSTANTES DE ACOMPLEJAMIENTO	32
5.2.1. Moléculas a estudiar	32
5.2.2. Esquemas de reacción	33
5.2.3. Métodos de cálculo	36
5.2.4. Estimación de errores	37
6- RESULTADOS Y DISCUSIÓN	38
6.1 Constantes de Acidez	38
6.1.1. Fenoles	38
6.1.1.1. Mono-funcionales PCM	38
6.1.1.2 Mono-funcionales SMD	41
6.1.1.3. Poli-funcionales PCM	43
6.1.1.4. Poli-funcionales SMD	45

6.1.1.5. Conjunto completo de fenoles con PCM	48
6.1.1.6. Conjunto completo de fenoles con SMD	50
6.1.1.7. Método de correlaciones para fenoles	52
6.1.1.8. Generalidades para fenoles	53
6.1.2. Ácidos Carboxílicos	54
6.1.2.1 Mono-funcionales PCM	54
6.1.2.2 Mono-funcionales SMD	56
6.1.2.3 Poli-funcionales PCM	58
6.1.2.4. Poli-funcionales SMD	60
6.1.2.5. Conjunto completo de ácidos carboxílicos con PCM	63
6.1.2.6. Conjunto de ácidos carboxílicos con SMD	65
6.1.2.7. Método de correlaciones para ácidos carboxílicos	67
6.1.2.8 Generalidades para ácidos carboxílicos	68
6.1.3 Aminas	70
6.1.3.1 Mono-funcionales PCM	73
6.1.3.2. Mono-funcionales SMD	76
6.1.3.3 Poli-funcionales PCM	79
6.1.3.4 Poli-funcionales SMD	82
6.1.3.5 Conjunto completo de aminas con PCM	84
6.1.3.6 Conjunto completo de aminas con SMD	85
6.1.3.7. Método de correlaciones para aminas	87
6.1.3.8 Generalidades para aminas	89
6.1.4. Efectos del conjunto de funciones de base	90
6.1.5. Generalidades pKas.	92
6.2 Constantes de Acomplejamiento	95
6.2.1 Geometría del Cu(II) en solución acuosa	95
6.2.2 Validación de estrategia de puntos simples	98
6.2.3 Complejos de Cu (II) con amoniac	100
6.2.4 Complejos de Cu (II) con acetato (CH ₃ COO ⁻)	105
6.2.5. Generalidades de ambos ligantes mono-funcionales	108
6.2.6. Complejos de Cu(II) con dopamina	108
7- CONCLUSIONES	113
8 - PERSPECTIVAS	114
9- REFERENCIAS	115
10- ANEXOS	119
11- PUBLICACIONES	141

Resumen

En esta tesis se estudiaron equilibrios químicos en solución acuosa con métodos teóricos. En particular se investigaron equilibrios de acidez y de formación de complejos. Se exploraron diferentes aspectos de las metodologías de cálculo con el propósito de identificar aquellas combinaciones que dan resultados en mejor acuerdo con los datos experimentales. Se utilizaron diferentes especies químicas: fenoles, ácidos carboxílicos y aminas para obtener las constantes de acidez; y amoníaco, acetatos y dopamina para conocer los valores de constantes de complejamiento con Cu(II). En ambos casos se incluyeron tanto especies mono-funcionales como poli-funcionales. Se encontró que el aspecto más importante de la metodología de cálculo es sistemáticamente el esquema de reacción utilizado. En el caso de equilibrios de acidez la característica clave es que estos tengan idéntico número y tipo de iones a ambos lados del equilibrio. Por otra parte, para los equilibrios de formación de complejos el aspecto que más influye en la calidad de los resultados es incluir un número suficiente de moléculas de agua explícitas en los modelos. De manera general se recomienda para equilibrios de acidez utilizar el funcional PBE0 con el esquema de intercambio protónico y SMD para modelar el solvente. Además en este proyecto queda demostrado que algunas metodologías son confiables para cualquier tipo de molécula, sin importar ser mono-funcional o poli-funcional. Otra alternativa es emplear el método de las correlaciones con este u otros funcionales. Para constantes de complejamiento se recomienda incluir 8 moléculas de agua explícitas para los sistemas de Cu(II), 4 de coordinación y 4 de solvatación, además del modelo de solvente SMD, y el funcional de la densidad BMK. Se espera que las recomendaciones derivadas de esta tesis puedan ser de ayuda a investigaciones futuras en el tema.

Abstract

Different equilibria in aqueous solution were investigated in this thesis using theoretical chemistry related methods, in particular acid-base equilibria and chelation equilibria. Diverse methodological aspects were explored, to identify which combinations provide the best agreement with the available experimental data. A variety of chemical species were included in this study: phenols, carboxylic acids and amines for the acidity constants; and ammonia, acetate and dopamine for the Cu(II) chelation constants. For both kind of equilibria mono-functional and poly-functional species were included. It was found that the most important methodological aspect is, systematically, the used reaction scheme. For acidity the key feature is that the reaction scheme has identical number and kind of ions at both sides of the equilibria. On the other hand, for chelation reactions the factor that influences the most the quality of the calculated results is including a large enough number of explicit water molecules in the model. In general, for acid-base equilibria it is recommended using the PBE0 functional, with the proton exchange scheme, and SMD to model the solvent. Another alternative is to use the correlation method with that, or other density functional methods. For complexation reactions it is recommended to include 8 explicit water molecules, 4 in the coordination sphere and 4 in the solvation sphere, in addition to the SMD solvent; and the BMK functional. It is expected that the recommendations derived from this study can be helpful to future investigations on the subject.

ABREVIATURAS

COSMO:	Conductor-like Screening Model
ΔG_s :	Energía libre de Gibbs en solución.
ΔG_g :	Energía libre de Gibbs en fase gas.
$\Delta \Delta G_{\text{solv}}$:	Energía libre de Gibbs de solvatación.
EAP:	Error absoluto promedio.
EAPG:	Error absoluto promedio global.
GGA:	Aproximación de gradiente generalizado.
H-F:	Hartree Fock.
Hyper-GGA:	Hiper-aproximación de gradiente generalizado.
HPI:	Aminas Primarias
HS1:	Aminas Secundarias
HT1:	Aminas Terciarias
HNI:	Neurotransmisores
HA:	Especie Neutra
LDA:	Aproximación de densidad local.
Meta-GGA:	Meta-aproximación de gradiente generalizado.
PCM:	Modelo Continuo Polarizable
RPA:	de sus siglas en inglés, Random Phase Approximation
SMD:	Modelo de Densidad Solvatada
TFD:	Teoría de funcionales de la densidad.

1.- INTRODUCCIÓN

La gran mayoría de los procesos químicos que ocurren en los organismos vivos y en la vida cotidiana involucran especies que tienen un comportamiento ácido-base y/o posibilidad de formar complejos con metales. Esto hace que sea de suma importancia conocer tanto sus constantes de equilibrio como los detalles de las diferentes reacciones químicas involucradas (orden de desprotonación y sitios de complejamiento) para estudiar adecuadamente el comportamiento de estas especies cuando participan en otros procesos. Por ejemplo, para entender el mecanismo de acción de un antioxidante a un determinado pH es necesario conocer su especie ácido-base predominante en estas condiciones.¹⁻³ También sería necesario conocer si es susceptible de formar complejos con metales (actividad antioxidante secundaria).^{4,5}

Con respecto al comportamiento ácido-base, en la literatura existen varias teorías que describen este comportamiento. Arrhenius definió a los ácidos como sustancias que donan iones H^+ y a las bases como sustancias que donan iones OH^- . Brønsted definió un ácido como un donador de protón y la base como un aceptor de protón. Lewis definió a un ácido como una sustancia que acepta electrones y a una base una sustancia donadora de un par de electrones no enlazante para compartirlo en un enlace. Aunque cualquiera de estas definiciones es válida de manera general, en este trabajo serán más relevantes las definiciones de Brønsted para las constantes de acidez y las de Lewis para las constantes de complejamiento, debido a la naturaleza de los equilibrios en cada caso. Las constantes de acidez (K_a) son comúnmente reportadas como pK_a s, que se definen como $-\log(K_a)$. Es de suma importancia conocer estas constantes ya que no solo caracterizan la acidez de los compuestos sino que también determinan la fracción de ellos que existen en cada forma ácido-base a un determinado pH , lo que influye en la reactividad y también afecta propiedades tales como la solubilidad.

En el caso particular de los compuestos farmacéuticos estos equilibrios influyen además en la forma de dosificación y el régimen de administración. Esto se debe a que los medicamentos tienen que atravesar una o más membranas celulares para llegar al torrente sanguíneo y alcanzar su sitio de acción. Cuando el transporte a través de las membranas es pasivo, son

esencialmente las especies neutras las que pueden transportarse, por lo que la especiación ácido-base, según los pKas, es de suma importancia.

Otro tipo de sustancias que también presentan este tipo de equilibrio ácido-base son los neurotransmisores. Estas moléculas y la mayoría de los fármacos son especies poli-donadoras de protones, o sea que tienen varios sitios posibles de desprotonación. Para este tipo de moléculas es importante además conocer la acidez relativa de los diferentes sitios para poder conocer con exactitud la estructura de la especie predominante a distintos valores de *pH*. Los grupos funcionales con comportamiento ácido-base más frecuentemente encontrados en fármacos y neurotransmisores son fenoles, ácidos carboxílicos y aminas. Es por ello que estos serán los estudiados en el presente trabajo.

Por otro lado, la formación de enlaces metal-ligando puede alterar notablemente las propiedades tanto del ion metálico como de los ligantes. De hecho los complejos metálicos presentan propiedades físicas y químicas diferentes a las de los fragmentos que lo conforman. Los valores de las constantes de complejamiento son sumamente importantes porque reflejan en qué proporción se encuentran los complejos con respecto a los fragmentos separados. Además es importante conocer el sitio de complejamiento más probable cuando los ligantes son poli-funcionales.

Los elementos metálicos ingresan a nuestro organismo a través de alimentos, el agua y el aire.⁶ Nuestro cuerpo almacena estos metales en tejidos orgánicos como el riñón, hígado, páncreas, siendo el más importante el cerebro. Para que el cerebro funcione adecuadamente se requiere de vitaminas, minerales, aminoácidos, ácidos grasos y neurotransmisores. Estos últimos son los responsables de diferentes funciones cerebrales específicas, por ejemplo: la epinefrina se ha relacionado con la motivación y la energía; la norepinefrina con la estabilidad emocional; la serotonina con el humor, el sueño y el apetito; y la dopamina con el placer y la satisfacción.⁷ Además cuando existe algún desequilibrio en las concentraciones de los neurotransmisores se ha sugerido que puede provocar varias enfermedades tales como: Parkinson, esquizofrenia, depresión, ansiedad, sobrepeso, anorexia, bulimia, migrañas, entre otras.⁸ En caso de exposición continua a metales, estos pueden acumularse hasta alcanzar niveles tóxicos. Estas concentraciones relativamente altas pueden provocar la formación de

complejos con algún neurotransmisor actuando como ligante. Esto disminuye la concentración local del neurotransmisor necesaria y puede provocar enfermedades.

Al mismo tiempo los metales son importantes para el buen funcionamiento del cuerpo humano. En particular se ha mostrado que el Cu(II) en animales y en personas es necesario para que varias enzimas realicen sus funciones.⁹ También se ha mostrado que el Cu(II) es necesario para el crecimiento de los niños, está involucrado en mecanismos de defensa, ayuda a tener huesos fuertes, a transportar el hierro, interviene en el metabolismo del colesterol y la glucosa y en el desarrollo del cerebro.⁶ Las fuentes fundamentales de Cu(II) en la dieta humana son mariscos, leche, pan integral, nueces, almendras, papas, huevo y agua.⁶ Sin embargo, cuando se encuentra en exceso puede llegar a ser tóxico. Una de las causas de su toxicidad es la capacidad oxidante que tiene, por su papel en la formación de radicales libres tóxicos como el $\bullet\text{OH}$.¹⁰ El exceso de Cu se ha relacionado con la enfermedad de Wilson,¹¹ las cataratas y problemas hepáticos. Otra de las enfermedades que se asocia a altas concentraciones de cobre es el mal de Alzheimer.¹²

Acomplejar al cobre es una posible vía química para prevenir los efectos negativos asociados a altas concentraciones de este metal. De modo que las terapias con agentes quelantes son una alternativa en estos casos.^{13,14} Por lo tanto, conocer los valores de sus constantes de complejamiento se vuelve un tema de interés ya que permiten conocer cuáles son los agentes quelantes más eficaces. Otra de las características importantes de los complejos de Cu(II), que es necesario conocer, es el sitio de complejamiento en el organismo. Esta información permite caracterizar a las especies químicas que se forman. Es por ello que este aspecto también es estudiado en el presente trabajo.

Si se tiene en cuenta todo lo anterior, las constantes de equilibrio en solución son de suma importancia, lo que hace necesario encontrar estrategias que proporcionen valores confiables. Las estrategias experimentales más utilizadas se basan en la ley de acción de masas y un método con el cual se puedan medir las concentraciones de las especies involucradas en el equilibrio, por ejemplo la espectroscopia ultravioleta- visible. Sin embargo, encontrar las constantes de equilibrio experimentalmente a veces se convierte en un reto, por ejemplo, cuando se trabaja con especies tóxicas, cuando se tienen intermediarios de vida muy corta, o

cuando se tiene una sustancia con poca solubilidad.¹⁵ En el caso de las constantes de acidez se han dedicado una gran cantidad de trabajos a obtener valores de pK_a s utilizando métodos teóricos.¹⁶⁻²⁰ Sin embargo esto aún constituye un reto para la química computacional y errores de hasta 2 unidades de pK_a se consideran aceptables.²¹ Por otra parte, el cálculo de constantes de complejamiento confiables está aún casi sin explorar. Hasta donde sabemos, existe solamente un trabajo anterior en el tema de complejos,²² y ni siquiera se ha establecido un umbral de calidad en este caso.

Las dificultades de utilizar métodos teóricos para calcular constantes de acidez y/o complejamiento en solución acuosa son varias, pero las dos más importantes se describen a continuación. La primera es que en el proceso de desprotonación o de complejamiento no siempre se conserva el número y tipo de especies cargadas a ambos lados del equilibrio. Por lo tanto, no hay una cancelación sistemática de errores cuando se calculan las energías de reacción. Como resultado, la calidad de estas energías relativas suele ser peor que para los procesos en el que el número y tipo de especies cargadas es el mismo a ambos lados del equilibrio.²¹ La segunda fuente de error se debe a los modelos usados para describir el solvente. Estos no son siempre suficientemente buenos para obtener energías de solvatación, lo que introduce un error en el cálculo de las energías libres de Gibbs y por lo tanto en las constantes de equilibrio.

Los métodos de solvente más utilizados en la literatura son los que se conocen como métodos continuos. El mayor problema en este caso es que las interacciones específicas soluto-solvente de corto alcance no se describen adecuadamente. Esto puede solucionarse incluyendo simultáneamente micro-solvatación, lo que se conoce como modelos discretos-continuo,²³ de campo de reacción supramolecular,²⁴ o modelo cúmulo-continuo.²⁵ Cuando se usa este tipo de estrategia se incluye un número pequeño de moléculas de solvente explícitas. De este modo las interacciones de largo alcance se incluyen por medio del modelo continuo, y las de corto alcance mediante la presencia de las moléculas de solvente explícitas. También puede modelarse el solvente usando modelos discretos, incluyendo una gran cantidad de moléculas de disolvente. El inconveniente en este caso es que con métodos *ab-initio* o de Teoría de Funcionales de la Densidad (TFD), esto sería tan costoso computacionalmente que no es una opción viable actualmente.

Para el cálculo de pKas la estrategia que ha demostrado frecuentemente ser más exitosa, se basa en obtener una máxima cancelación de errores. Dentro de esta estrategia el método más usado hasta el momento se conoce como método de intercambio protónico, método isodésmico, o método relativo para calcular pKas. Como su nombre lo indica se basa en calcular energías de intercambio protónico en lugar de energías de desprotonación. Para ello se incluye un ácido de referencia que debe ser muy parecido al de interés y del que es necesario conocer el pKa experimental con exactitud. El principal problema con este enfoque es que la calidad de los resultados depende en gran medida de la elección del ácido de referencia (HRef).

Otra estrategia que ha sido recientemente propuesta para disminuir los errores al calcular pKas es un método conocido como de correlaciones.²⁶⁻³² En esta estrategia se usa un conjunto de moléculas de referencia para las cuales se conocen los pKa experimentales y se obtiene una correlación lineal de estos valores con la energía calculada para el proceso ácido-base de interés. De esta manera se obtienen dos parámetros, la pendiente y la ordenada al origen, que luego son utilizadas para calcular los pKas de interés. Además la comprobación de los parámetros obtenidos puede validarse utilizando un conjunto de moléculas de prueba para las que también se conocen los valores experimentales de pKas. Estas correlaciones se hacen de manera separada para cada familia de compuestos. De este modo se disminuyen significativamente las discrepancias entre valores calculados y experimentales. Sin embargo, su mayor inconveniente es justamente que depende de valores experimentales.

En el presente trabajo se explorarán diferentes metodologías de cálculo, validando la calidad de los resultados obtenidos mediante comparación con datos experimentales de un conjunto de moléculas de referencia. Con base en esta comparación se identificarán las metodologías óptimas, y estas serán utilizadas luego para predecir las constantes de equilibrio tanto de acidez como de complejamiento de diferentes compuestos de interés biológico. Una característica adicional que se buscará en las metodologías es que no dependan de datos experimentales. Además nuestra meta es encontrar metodologías que disminuyan significativamente los errores comúnmente aceptados, de ser posible con errores en logK menores a una unidad. Esto permitiría obtener valores de constantes de equilibrio en solución

acuosa confiables, para cuando se tengan moléculas de interés y en la literatura aunque su valor no esté reportado.

2.- ANTECEDENTES

La búsqueda para determinar constantes de acidez con métodos teóricos, confiables y con un costo computacional razonable, teniendo como base métodos teóricos lleva más de una década³³. Existe abundante literatura sobre este tema. Sin embargo, en esta parte del proyecto se pretende enfatizar ciertas contribuciones que ayudaron a la motivación del presente proyecto.

Una de ellas es la investigación realizada por Liptak y colaboradores.³³ En este trabajo se emplearon métodos compuestos, CBS-QB3 y CBS-APNO, para calcular pKas de moléculas fenólicas con un solo sitio de desprotonación. Esta metodología se realizó con el modelo continuo CPCM y ciclos termodinámicos en los que se calculan de manera separada las energías de solvatación y las energías en fase gas, para luego obtener las energías en solución como suma de ambas. Las desviaciones que se obtuvieron fueron de alrededor de 0.5 unidades de pKas. Dissanayake y Senthilnithy³⁴ también usaron ciclos termodinámicos y energías CBS-QB3 para calcular los pKas de 12 ácidos hidroxicinámicos (con dos posibles sitios de desprotonación). La desviación con respecto a valores experimentales fue de 0.32 unidades de pKa. En ambos casos los cálculos se realizaron para moléculas pequeñas. La generalización de esta estrategia no es viable, debido al alto costo computacional que implicaría para moléculas de tamaño mediano y grande.

Pliego y colaboradores,³⁵ emplearon 17 especies orgánicas incluyendo ácidos carboxílicos, alcoholes y fenoles, con un solo sitio de desprotonación. Estos autores utilizaron ciclos termodinámicos con el método MP2 y los modelos de solvente continuos SM5.42R y PCM. Los valores que obtuvieron tienen una desviación de 2.2 unidades de pKa. Esta desviación es mucho mayor que la obtenida por Liptak y colaboradores para fenoles, y también presenta el inconveniente del costo computacional.

Frau y colaboradores,³⁶ usaron el método isodésmico junto con ciclos termodinámicos para alcoholes alifáticos, ácidos carboxílicos, aminas, fenoles, ácidos benzoicos y piridinas, (con un solo sitio de desprotonación). Las energías en fase gas se calcularon usando dos funcionales B3LYP y M05-2X, además del método compuesto CBS-4B3 y para obtener las energías de solvatación usaron los modelos continuos SMD y CPCM. La desviación que se obtuvo en este caso fue de 0.5-1 unidades de pKas.

En todos los trabajos mencionados hasta el momento se utilizó agua como disolvente. Sin embargo, experimentalmente hay pKas reportados en otros solventes, especialmente alcoholes. Pliego y colaboradores³⁷ realizaron otra contribución en el tema, esta vez con dos tipos de solvente metanol y agua y utilizando dos métodos: el directo y el isodésmico. Estos autores emplearon el funcional X3LYP junto con los modelos continuos SM8 y SMD. El estudio se realizó para 26 ácidos carboxílicos, 24 fenoles y 23 aminas. La desviación obtenida fue de 3.6 unidades de pKa.

Estos son algunos de los trabajos de investigación publicados hasta el momento en los que se ha intentado utilizar métodos teóricos para cálculos de pKa. Los trabajos antes mencionados han sido los más exitosos, en el sentido de que son los que han obtenido desviaciones más pequeñas con respecto a los valores experimentales. Sin embargo es importante mencionar que existen una gama amplia de artículos y fuentes sobre este tipo de investigación.³⁸⁻⁴⁷ El estado del arte en este tema fue revisado en detalle en 2009 por Ho y Coote.²¹

Como se mencionó anteriormente, hasta el momento solo hay un trabajo en la literatura en el que se ha usado usar métodos teóricos para el cálculo de constantes de complejamiento del cobre que se acerquen a los valores experimentales. En este se estima la constante de complejamiento de Cu (II) con hidroxilos utilizando el funcional de la densidad B3LYP junto con el modelo continuo COSMO. La estrategia para mejorar los valores de energías se basó en ir aumentando el número de moléculas de agua explícitas que se incluyen en el modelo. Los mejores resultados se obtuvieron para 18 moléculas de agua y el error obtenido en este caso fue de 0.7 Kcal/mol, mientras que utilizando solo 4 moléculas de agua el error se incrementa a 6 Kcal/mol.²²

Los trabajos antes discutidos demuestran que aún queda mucho por hacer en esta temática, por lo que los esfuerzos continúan para encontrar estrategias de cálculo con errores pequeños, cercanos a los que se obtienen en estudios experimentales. Además, la mayor parte de las investigaciones realizadas tratan solamente con especies mono-funcionales; esto es, con compuestos ácido-base con un solo sitio posible de desprotonación y ligantes con un solo posible sitio de complejamiento. El problema se hace aún más complicado para especies poli-funcionales.

3.- HIPÓTESIS

En esta tesis se asume que si se encuentra una metodología confiable para obtener constantes de equilibrio en solución acuosa de moléculas simples (mono-funcionales), ésta podrá aplicarse de manera igualmente confiable a moléculas más complejas (poli-funcionales).

4- OBJETIVOS

4.1 Objetivo General

El objetivo principal de este trabajo es encontrar una metodología computacional que permita obtener valores confiables de constante de equilibrio en solución acuosa, en particular para equilibrios ácido-base y de acomplejamiento, con aplicación a sistemas o moléculas de interés biológico.

4.2 Objetivos Particulares

Para poder cumplir con el objetivo general de esta tesis, se siguieron los siguientes objetivos particulares:

1) Calcular con diferentes metodologías los pKas de una amplia serie de ácidos carboxílicos, fenoles y aminas (mono-funcionales) para los que se conoce sus valores experimentales.

- 2) Comparar los valores calculados de pKa con los valores experimentales e identificar las metodologías con menores errores en ambos casos.
- 3) Identificar los sitios de deprotonación más probables en las moléculas poli-próticas.
- 4) Calcular los pKas de varios fármacos y neurotransmisores (poli-funcionales) con las metodologías identificadas como más confiables para el conjunto de moléculas mono-funcionales. Estimar las desviaciones con respecto a los valores experimentales.
- 5) Usar diferentes estrategias para el cálculo de constantes de acomplejamiento de Cu(II) con ligantes sencillos (amoníaco y acetatos).
- 6) Comparar los valores calculados de las constantes de acomplejamiento con los valores experimentales e identificar las metodologías con menores errores.
- 7) Identificar el sitio de acomplejamiento más probable en la dopamina.
- 8) Calcular las constantes de acomplejamiento de Cu(II) con dopamina utilizando las metodologías identificadas como más confiables para NH₃ y acetato. Utilizar los datos experimentales disponibles para estimar las desviaciones.
- 7) Comprobar si las metodologías con buenos resultados para sistemas sencillos (especies mono-funcionales) son también eficientes para sistemas complejos (especies poli-funcionales)

5- METODOLOGÍA

Dado que el objetivo fundamental de este trabajo es identificar metodologías viables para el cálculo de magnitudes termodinámicas, se utilizaron diferentes protocolos computacionales. En general tanto para constantes de acidez como para constantes de complejamiento se cambiaron diferentes aspectos de la modelación incluyendo esquemas de reacción y número de moléculas de agua explícitas, método de estructura electrónica y modelo de solvente. Los resultados obtenidos con las diferentes variantes fueron comparados con los datos

experimentales disponibles para así reconocer los protocolos que conducen a errores más bajos. Además se usaron dos conjuntos de moléculas en cada caso. Uno de ellos incluyendo solamente moléculas relativamente simples (mono-funcionales) y el otro incluyendo moléculas más complejas (poli-funcionales).

5.1. CONSTANTES DE ACIDEZ

5.1.1. Moléculas a estudiar

Se estudiaron 3 familias de compuestos: fenoles, ácidos carboxílicos y aminas. Esto con la idea de si no existiese una estrategia única que demostrara ser exitosa para el cálculo de cualquier pKa, poder al menos hacer recomendaciones específicas según el tipo de sitio ácido-base involucrado en el equilibrio de acidez de interés. Para este estudio se escogieron como moléculas mono-funcionales 6 fenoles, 10 ácidos carboxílicos y 10 aminas (Tabla 1). De estas últimas 3 son aminas terciarias, 3 son aminas secundarias, 3 son aminas primarias y además se incluyó el NH₃. Como moléculas poli-funcionales se escogieron 4 en cada caso (fenoles, ácidos carboxílicos y aminas). Sus estructuras se muestran en la Tabla 2 y sus valores de pKa experimentales en el Anexo 1. En todos los casos fueron moléculas de interés biológico, fármacos para el caso de ácidos y fenoles y neurotransmisores para el caso de aminas. Es importante hacer notar que aunque estas moléculas poli-funcionales, como su nombre lo indica, pueden presentar más de un tipo de sitio ácido-base, en cada caso se estudió solamente el pKa correspondiente a la familia de interés.

En todos los casos, y para poder poner a prueba nuestra hipótesis, se escogieron sistemas poli-funcionales para los que existieran valores experimentales de pKa reportados en la literatura (Anexo 1). Estos valores se incluyen en las Tablas 1 y 2. Para estos sistemas se hizo primeramente, en los casos necesarios, un estudio previo para identificar el sitio de desprotonación más probable. Para ello se calcularon las energías de desprotonación para todos los sitios posibles, y se seleccionó como más probable el que requiere menor energía para perder el protón.

Tabla 1. Moléculas mono-funcionales a estudiar en el presente trabajo

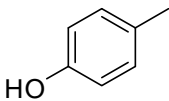
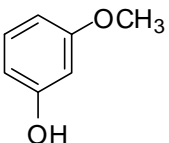
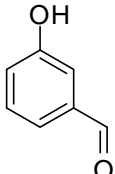
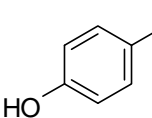
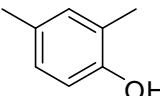
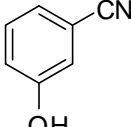


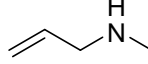
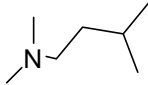

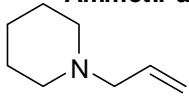
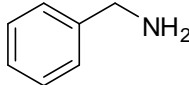
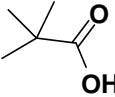

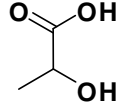
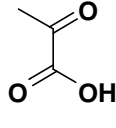
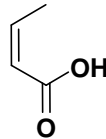
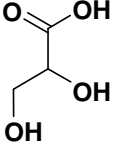
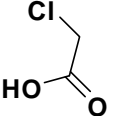
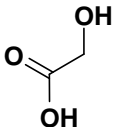
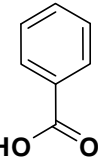
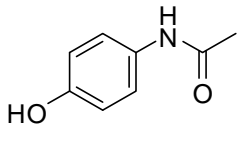
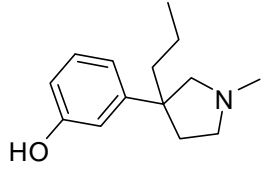
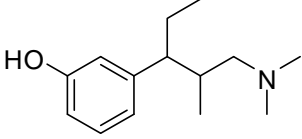
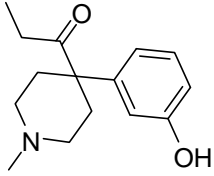
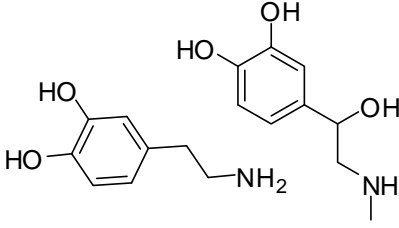
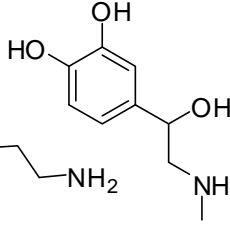
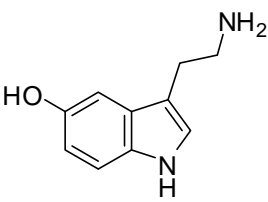
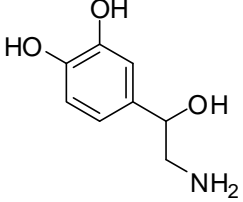
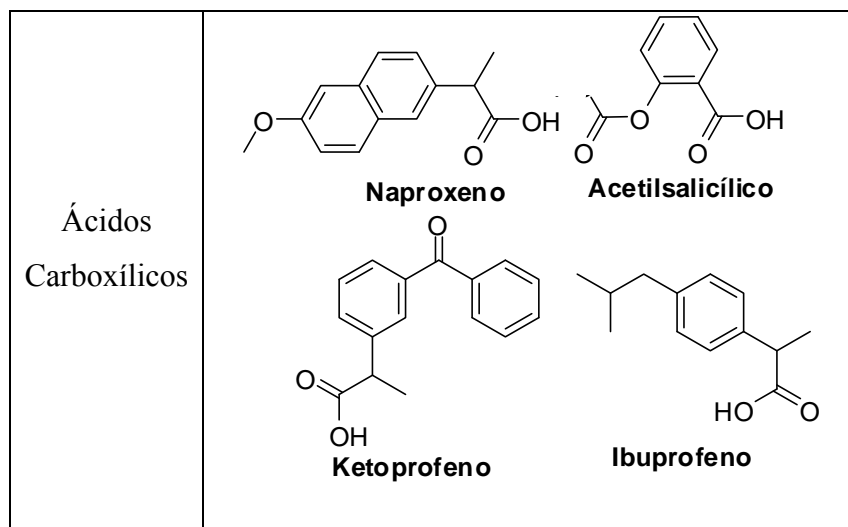
<p>Fenoles</p>	<div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="text-align: center;">  4-Metilfenol </div> <div style="text-align: center;">  3-Metoxifenol </div> <div style="text-align: center;">  3-Hidroxibenzaldehído </div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start; margin-top: 10px;"> <div style="text-align: center;">  4-Metiltiofenol </div> <div style="text-align: center;">  2,4-Dimetilfenol </div> <div style="text-align: center;">  3-Cianofenol </div> </div>
<p>Aminas (escritas en su forma neutra, o sea base conjugada)</p>	<div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="text-align: center;"> NH_3 Amoniaco </div> <div style="text-align: center;"> —NH_2 Metilamina </div> <div style="text-align: center;"> $\text{—CH}_2\text{—NH}_2$ Etilamina </div> <div style="text-align: center;">  Dimetilamina </div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start; margin-top: 10px;"> <div style="text-align: center;">  Trimetilamina </div> <div style="text-align: center;">  Alilmetil-amina </div> <div style="text-align: center;">  Dimetil-sec-butilamina </div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start; margin-top: 10px;"> <div style="text-align: center;">  Piperidina </div> <div style="text-align: center;">  Nalilpiperidina </div> <div style="text-align: center;">  Benzilamina </div> </div>
<p>Ácidos Carboxílicos</p>	<div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="text-align: center;">  Acético </div> <div style="text-align: center;">  Fórmico </div> <div style="text-align: center;">  Láctico </div> <div style="text-align: center;">  Pirúvico </div> <div style="text-align: center;">  Crotónico </div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start; margin-top: 10px;"> <div style="text-align: center;">  Glicérico </div> <div style="text-align: center;">  Cloroacético </div> <div style="text-align: center;">  Glicólico </div> <div style="text-align: center;">  Benzoico </div> </div>

Tabla 2. Moléculas poli-funcionales a estudiar en el presente trabajo.

<p>Fenoles</p>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Paracetamol</p> </div> <div style="text-align: center;">  <p>Profadol</p> </div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 20px;"> <div style="text-align: center;">  <p>Tapentadol</p> </div> <div style="text-align: center;">  <p>Ketobemidona</p> </div> </div>
<p>Aminas (escritas en su forma neutra, o sea base conjugada)</p>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Dopamina</p> </div> <div style="text-align: center;">  <p>Epinefrina</p> </div> </div> <div style="display: flex; justify-content: space-around; align-items: center; margin-top: 20px;"> <div style="text-align: center;">  <p>Serotonina</p> </div> <div style="text-align: center;">  <p>Norepinefrina</p> </div> </div>



5.1.2. Esquemas de reacción

Para las familias de los fenoles y ácidos carboxílicos, se probaron veintidós esquemas de reacción que se reportan en la Tabla 3, junto con las expresiones correspondientes al cálculo directo de pK_a s a partir de los valores de ΔG_s . Para estas familias de moléculas su primera desprotonación corresponde al proceso de formación de la especie aniónica a partir de la neutra, a diferencia de la familia de las aminas que su primera desprotonación es de la especie protonada a la neutra. Además la fuerza la sal de amonio que son los ácidos conjugados varía dependiendo de que estas sean primarias, secundarias o terciarias. Por lo tanto en el caso de las aminas los esquemas de reacción, no son idénticos a los de ácidos y fenoles, aunque sí equivalentes (Tabla 4).

Tabla 3. Esquemas de reacción para las familias de los ácidos carboxílicos y fenoles, y expresiones para calcular directamente los valores de pK_a a partir de ΔG_s .

Esquema	Equilibrio	$pK_a =$
A	$HA \leftrightarrow H^+ + A^-$	$\frac{\Delta G_s}{RT \ln(10)}$

B	$HA + H_2O \leftrightarrow H_3O^+ + A^-$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$
C1	$HA + 2H_2O \leftrightarrow H_3O^+ + A^-(H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 2\log[H_2O]$
C2	$HA + 3H_2O \leftrightarrow H_3O^+ + A^-(2H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 3\log[H_2O]$
C3	$HA + 4H_2O \leftrightarrow H_3O^+ + A^-(3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 4\log[H_2O]$
D	$HA + Ref^- \leftrightarrow A^- + HRef$	$\frac{\Delta G_s}{RT \ln(10)} + pKa(HRef)$
E1	$HA + OH^-(3H_2O) \leftrightarrow A^-(H_2O) + 3H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 3\log[H_2O]$
E2	$HA + OH^-(3H_2O) \leftrightarrow A^-(2H_2O) + 2H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 2\log[H_2O]$
E3	$HA + OH^-(3H_2O) \leftrightarrow A^-(3H_2O) + H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + \log[H_2O]$
F1	$HA + 4H_2O \leftrightarrow H_3O^+(3H_2O) + A^-$	$\frac{\Delta G_s}{RT \ln(10)} - 4\log[H_2O]$
F2	$HA + 5H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 5\log[H_2O]$
F3	$HA + 6H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(2H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 6\log[H_2O]$
F4	$HA + 7H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 7\log[H_2O]$

CN1	$HA(H_2O) + H_2O \leftrightarrow H_3O^+ + A^-(H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$
CN2	$HA(H_2O) + 2H_2O \leftrightarrow H_3O^+ + A^-(2H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 2\log[H_2O]$
CN3	$HA(H_2O) + 3H_2O \leftrightarrow H_3O^+ + A^-(3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 3\log[H_2O]$
FN1	$HA(H_2O) + 3H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 3\log[H_2O]$
FN2	$HA(H_2O) + 4H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(2H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 4\log[H_2O]$
FN3	$HA(H_2O) + 5H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 5\log[H_2O]$
EN1	$HA(H_2O) + OH^-(3H_2O) \leftrightarrow A^-(H_2O) + 4H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 4\log[H_2O]$
EN2	$HA(H_2O) + OH^-(3H_2O) \leftrightarrow A^-(2H_2O) + 3H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 3\log[H_2O]$
EN3	$HA(H_2O) + OH^-(3H_2O) \leftrightarrow A^-(3H_2O) + 2H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 2\log[H_2O]$

Tabla 4. Esquemas de reacción para las aminas, mostrando las expresiones correspondientes para calcular directamente los valores de pK_a a partir de ΔG_s .

Esquema	Esquema de reacción	$pK_a =$	Usado para*
A	$NH_4^+ \rightleftharpoons H^+ + NH_3$	$\frac{\Delta G_s}{RT \ln(10)}$	NH_4^+ , HP^+ , HS^+ , HT^+
B	$NH_4^+ + H_2O \rightleftharpoons H_3O^+ + NH_3$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$	NH_4^+ , HP^+ , HS^+ , HT^+
C1	$NH_4^+(H_2O) \rightleftharpoons H_3O^+ + NH_3$	$\frac{\Delta G_s}{RT \ln(10)}$	NH_4^+ , HP^+ , HS^+ , HT^+
C2	$NH_4^+(2H_2O) \rightleftharpoons H_3O^+ + NH_3 + H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + \log[H_2O]$	NH_4^+ , HP^+ , HS^+
C3	$NH_4^+(3H_2O) \rightleftharpoons H_3O^+ + NH_3 + 2H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 2\log[H_2O]$	NH_4^+ , HP^+
CN1	$NH_4^+(H_2O) + H_2O \rightleftharpoons H_3O^+ + NH_3(H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$	NH_4^+ , HP^+ , HS^+ , HT^+
CN2	$NH_4^+(2H_2O) \rightleftharpoons H_3O^+ + NH_3(H_2O)$	$\frac{\Delta G_s}{RT \ln(10)}$	NH_4^+ , HP^+ , HS^+

CN3	$NH_4^+(3H_2O) \rightleftharpoons H_3O^+ + NH_3(H_2O) + H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + \log[H_2O]$	NH ₄ ⁺ , HP ⁺
D	$NH_4^+ + Ref \rightleftharpoons NH_3 + HRef$	$\frac{\Delta G_s}{RT \ln(10)} + pKa(HRef)$	NH ₄ ⁺ , HP ⁺ , HS ⁺ , HT ⁺
E1	$NH_4^+(H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3 + 5H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 5 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺ , HT ⁺
E2	$NH_4^+(2H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3 + 6H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 6 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺
E3	$NH_4^+(3H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3 + 7H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 7 \log[H_2O]$	NH ₄ ⁺ , HP ⁺
EN1	$NH_4^+(H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3(H_2O) + 4H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 4 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺ , HT ⁺
EN2	$NH_4^+(2H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3(H_2O) + 5H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 5 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺
EN3	$NH_4^+(3H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3(H_2O) + 6H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 6 \log[H_2O]$	NH ₄ ⁺ , HP ⁺

F1	$NH_4^+ + 4H_2O \rightleftharpoons NH_3 + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 4 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺ , HT ⁺
F2	$NH_4^+ (H_2O) + 3H_2O \rightleftharpoons NH_3 + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 3 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺ , HT ⁺
F3	$NH_4^+ (2H_2O) + 2H_2O \rightleftharpoons NH_3 + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 2 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺
F4	$NH_4^+ (3H_2O) + H_2O \rightleftharpoons NH_3 + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$	NH ₄ ⁺ , HP ⁺
FN1	$NH_4^+ + 5H_2O \rightleftharpoons NH_3 (H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 5 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺ , HT ⁺
FN2	$NH_4^+ (H_2O) + 4H_2O \rightleftharpoons NH_3 (H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 4 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺ , HT ⁺
FN3	$NH_4^+ (2H_2O) + 3H_2O \rightleftharpoons NH_3 (H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 3 \log[H_2O]$	NH ₄ ⁺ , HP ⁺ , HS ⁺
FN4	$NH_4^+ (3H_2O) + 2H_2O \rightleftharpoons NH_3 (H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - 2 \log[H_2O]$	NH ₄ ⁺ , HP ⁺

* HP⁺= aminas protonada primaria, HS⁺= aminas protonada secundaria, HT⁺= aminas protonada terciaria

El esquema **A** de la Tabla 3 es el más usado en la literatura, probablemente debido a su simplicidad. Como podemos observar este esquema corresponde a una desprotonación simple, por lo cual, incluye explícitamente al protón. Esto es una desventaja desde el punto de vista teórico ya que se sabe que los métodos de cálculo reproducen mal las energías de solvatación de esta especie en particular⁴⁸. Por lo tanto la energía del protón se obtuvo a partir de los valores experimentales. Estos valores en sí mismos representan una fuente de error adicional ya que varían desde -259 hasta -264 Kcal/mol.²¹ Solo esta variación representa alrededor de 3 unidades de *pKa*. En este trabajo se usaron los valores recomendados por Camaioni y Schwerdtfeger,⁴⁹ que son $\Delta G_g(\text{H}^+) = -4.39$ kcal/mol y $\Delta G_{\text{solv}}(\text{H}^+) = -265.89$ kcal/mol, lo que da un valor de $\Delta G_s(\text{H}^+) = -270.28$ kcal/mol.

Una posible estrategia para evitar este problema es utilizar una molécula de agua como co-reactivo (Esquema **B**). Esta estrategia se puede ampliar incluyendo más de una molécula de agua y modelando tanto al protón como a las otras especies iónicas que aparezcan en los equilibrios explícitamente solvatados (esquemas de reacción **C**). Las variaciones dentro de los esquemas **C** consisten en el número de aguas explícitas que se usan para micro-solvatar a la base conjugada de ácidos y fenoles (o a la especie ácida catiónica en el caso de aminas). El esquema **D**, corresponde a un intercambio protónico entre la especie ácida de interés y un ácido de referencia, que debe ser estructuralmente similar y para el cual es necesario conocer su *pKa* experimental. Para las moléculas que se estudian en este proyecto los ácidos de referencia que se usaron son el fenol para los fenoles, el ácido acético para los ácidos carboxílicos y para las aminas se usó la metil-amina para aminas primarias, la dimetil-amina para aminas secundarias y la trimetil-amina para aminas terciarias. En el caso particular de este esquema, se suelen obtener valores más confiables de *pKas* (que con los esquemas **A**, **B** y **C**), ya que el número y tipo de especies cargadas se conserva a ambos lados de la ecuación química, sin importar la familia de moléculas que se analice. El principal problema con este enfoque es que la calidad de los resultados depende en gran medida de la elección del ácido de referencia (HRef), en particular de que tan similares sean estructuralmente HRef y HA, así como de la exactitud del valor experimental del *pKa* de HRef. Desde el punto de vista teórico, los esquemas **D** y **A** tienen el mismo inconveniente, que dependen directamente de datos experimentales. Esto debilita el carácter predictivo de estas metodologías.

Las diferentes variantes del esquema *E* se formularon con la ayuda de la ley de Hess y el equilibrio de autoprotólisis del agua. Por lo tanto estos esquemas no incluyen explícitamente al protón, sino al hidroxilo solvatado. En estos esquemas se solvata explícitamente a todas las especies iónicas (no solo el OH⁻ sino también a las involucradas en los equilibrios ácido-base de interés). En el caso de los ácidos carboxílicos y fenoles se espera que este tipo de esquema de reacción proporcione valores más cercanos al experimental que el método directo, ya que se igualan el número y tipo de especies iónicas a ambos lados del equilibrio. Por el contrario, para el caso de las aminas la compensación de cargas se lleva a cabo en los esquemas *F*. En estos esquemas, al igual que en los *E*, se solvata explícitamente a todas las especies iónicas presentes, solo que en este caso el ión ajeno al par de interés no es el OH⁻ sino el ion hidronio.

Los esquemas *CN*, *FN* y *EN* son derivados de los esquemas *C*, *F* y *E*. La diferencia radica en que se introduce una molécula de agua extra que solvata a la especie neutra del par ácido base. Los números en los acrónimos de los esquemas de reacción se refieren al número de moléculas de agua incluidas explícitamente.

Además, en el caso de las aminas el número de aguas explícitas incluidas en las modelaciones depende de que las aminas sean primarias, secundarias o terciarias. Esto ya que en cada subgrupo varía el número de H ácidos y por tanto el número de sitios que pueden involucrarse en la micro-solvatación. Por lo tanto el número máximo de moléculas de aguas explícitas que se usó fue 3 para aminas primarias, 2 para aminas secundarias y 1 para aminas terciarias.

La mayoría de los esquemas de reacción para las moléculas estudiadas en este proyecto permiten conservar el número y tipo de especies cargadas a ambos lados del equilibrio. Además para todos los esquemas de reacción que incluyen moléculas de agua explícitamente, es necesario hacer la corrección para el estado estándar del agua líquida, es decir 55.55 mol/L. De hecho se ha demostrado previamente que no tener en cuenta esta corrección lleva a errores sistemáticos.^{50,51}

5.1.3. Métodos de cálculo

Teoría de Funcionales de la Densidad (TFD)

La base de la TFD son los teoremas de Hohenberg y Kohn,⁵² los cuales establecen que la energía electrónica del estado basal está determinada completamente por la densidad electrónica. Estos teoremas son:

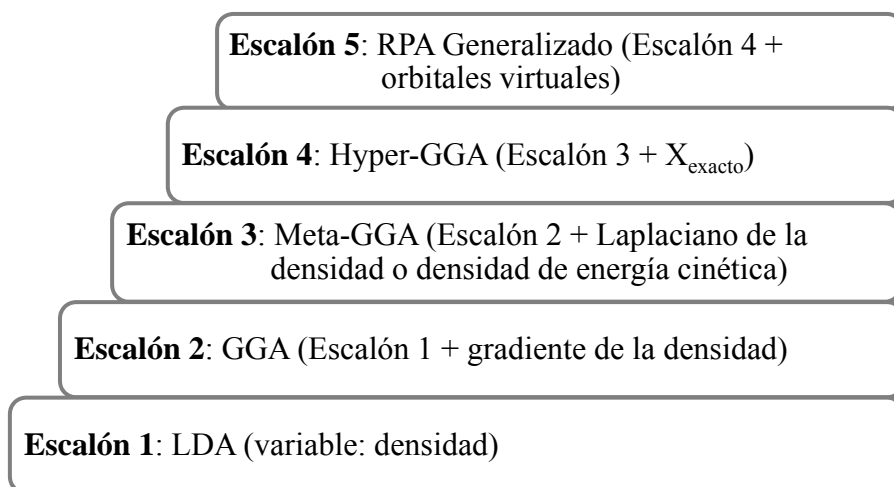
- 1) Para un sistema multielectrónico en su estado basal, la densidad electrónica (ρ) exacta del estado base tiene una relación unívoca con el potencial externo (v). Así que como la función de onda (Ψ) es un funcional de v , entonces se puede establecer que Ψ es un funcional de ρ . En consecuencia cualquier observable relacionado con el estado basal, como lo es la energía, es un funcional de la densidad.
- 2) Usando la densidad electrónica como una variable básica, existe un principio variacional análogo al de la mecánica cuántica. Dada una densidad electrónica aproximada, asumida como una cantidad real positiva que integra al número de electrones, la energía dada por esta densidad es superior a la energía exacta del estado basal.

Sin embargo, Hohenberg y Kohn no establecieron cómo calcular la energía a partir de la densidad, ya que se desconoce la expresión del funcional exacto. Fueron Kohn y Sham⁵³ quienes por primera vez propusieron una forma de aproximar el funcional $F[\rho]$. Para ello consideraron un sistema de referencia ficticio formado por un sistema de N electrones no interactuantes que experimentan el mismo potencial externo (inducido por los núcleos atómicos). Este potencial es tal que la densidad de probabilidad electrónica del estado fundamental del sistema de referencia es igual a la del estado fundamental del sistema real. Como el sistema de referencia consta de partículas no interactuantes, la función de onda de su estado fundamental es un determinante de Slater de los espín-orbitales ortonormales de Kohn-Sham de más baja energía para el sistema de referencia.

El término clave en la aplicación práctica de la formulación de Kohn-Sham es el funcional de la energía de intercambio y correlación (E_{xc}). Aunque este representa una fracción pequeña de la energía total, es el único término desconocido en las ecuaciones de Kohn-

Sham. Actualmente existen diversas aproximaciones para calcular E_{XC} . La más simple se conoce como aproximación de densidad local (LDA, del inglés Local Density Aproximation) y consiste en suponer que en cada punto la energía de intercambio y correlación depende sólo de la densidad en ese punto, este valor se considera como el que tendría un gas de electrones libres. Las aproximaciones de gradiente generalizado (GGA, del inglés Generalized Gradient Aproximation) son semilocales, ya que consideran en cada punto el valor de la densidad y de sus gradientes. Por otra parte, la introducción de funcionales GGA y funcionales híbridos (con porcentaje de intercambio exacto, X_{exacto}) produjo una importante mejora en términos de precisión para aplicaciones químicas.

La clasificación de las diferentes aproximaciones puede hacerse si se tienen en cuenta las variables fundamentales que se utilizan para definir el funcional de intercambio y correlación. J. P. Perdew propuso “la escalera de Jacob”⁵⁴⁻⁵⁶ para clasificar sistemáticamente los funcionales dependiendo de estas variables que permiten incrementar su precisión o al menos esperar una mejora en cada paso de la escalera:



Las aproximaciones a la TFD usadas en este trabajo fueron: BLYP, B3LYP, BHandHLYP, PBE, PBE0, PW91, BMK, TPSS y M05-2X. Se escogieron estos funcionales de la densidad porque presentan diferentes características como son: número de parámetros ajustados empíricamente, la inclusión o no del intercambio exacto y su cantidad, así como su ubicación en los diferentes escalones de la escalera de Jacob (Tabla 5). La idea de esta variación fue investigar si alguna de estas características juega un papel significativo en la obtención de valores más confiables de pK_a .

Tabla 5. Características de los funcionales de la densidad usados en la metodología de constantes de acidez

FDA	% Ex	NPAE	Clasificación	Año	Escalón	Ref.
BHandHLYP	50		<i>Hyper-GGA</i>	1992	4	57
BLYP		5	<i>GGA</i>	1988	2	58
B3LYP	20	8	<i>Hyper-GGA</i>	1993	4	59
PW91			<i>GGA</i>	1992	2	60
PBE			<i>GGA</i>	1997	2	61
PBE0	25		<i>Hyper-GGA</i>	1999	4	62
TPSS			<i>Meta-GGA</i>	2003	3	63
BMK	42	16	<i>Hyper-meta-GGA</i>	2004	4	64
M05-2X	56	20	<i>Hyper-meta-GGA</i>	2006	4	65

NPAE= número de parámetros ajustados empíricamente

Además, se usó el método de función de onda Moller-Plesset a segundo orden (MP2) para los sistemas mono funcionales. Todos los funcionales de la densidad y MP2, se usaron en combinación con el conjunto de funciones de base 6-311++G (d, p).

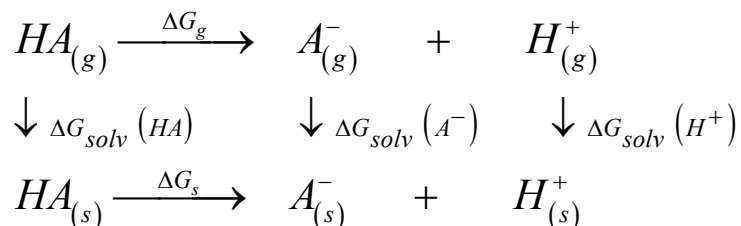
Energía de solvatación

En el presente estudio se utilizaron dos estrategias diferentes para obtener las energías de solvatación:

-Estrategia Mixta: la modelación se realiza en dos etapas, en la primera (conocida como de ciclos termodinámicos) se hacen las optimizaciones de geometrías y los cálculos de frecuencia al vacío, para obtener las energías libres en fase gas (ΔG_g) y luego se realizan cálculos de puntos simple en solución para obtener las energías libres de Gibbs de solvatación (ΔG_{sol}). Es importante mencionar que cuando nos referimos a la diferencia de energía libre es solo para ejemplificar la diferencia de energía libre de productos menos reactivos. Para ello se utilizó el modelo continuo polarizable PCM, específicamente IEF-PCM.⁶⁶ En este tipo de cálculos la elección de la cavidad del soluto es importante porque las energías calculadas

y las propiedades obtenidas dependen del tamaño de la cavidad. En el presente estudio la cavidad ha sido construida utilizando el modelo UAHF,⁶⁷ es el modelo topológico para obtener el radio atómico usando el nivel de teoría HF/6-31+g(d). Entonces utilizando la ley de Hess se calculan las energías libres en solución (ΔG_s) como $\Delta G_g + \Delta G_{solv}$.

Por ejemplo para el esquema de la reacción A, el ciclo termodinámico es el siguiente:



Por lo tanto la energía libre de Gibbs en solución (ΔG_s) se obtiene como la suma de energía libre de Gibbs de la reacción en el vacío (ΔG_g) y la diferencia en energías de solvatación ($\Delta\Delta G_{solv}$):

$$\Delta G_s = \Delta G_g + \Delta\Delta G_{solv} \quad (1)$$

donde $\Delta\Delta G_{solv}$ y ΔG_g son calculadas de la siguiente forma:

$$\Delta\Delta G_{solv} = \Delta G_{solv}(A^-) + \Delta G_{solv}(H^+) - \Delta G_{solv}(HA) \quad (2)$$

$$\Delta G_g = G_g(A^-) + G_g(H^+) - \Delta G_g(HA) \quad (3)$$

Aquí ΔG_{solv} representa las energías de solvatación de las diferentes especies. En la aplicación de estos ciclos es importante recordar que los cálculos se realizan para estado estándar de 1atm y por tanto es necesario adicionar la corrección $\Delta nRT \ln(\dot{R}T)$ a los valores calculados en fase gas para convertirlos a estado estándar 1M. En esta expresión los símbolos R y \dot{R} representan la constante de los gases en unidades de J/mol K y L atm/mol K, respectivamente.

Es importante mencionar que la estrategia mixta antes descrita se utilizó porque era la recomendada en el año 2010 cuando se empezó este proyecto y se trabajó con el programa Gaussian 03⁶⁸ porque era la versión disponible en ese momento.

-Estrategia Todo en Solución: todos los cálculos (energías, geometrías y correcciones termodinámicas) se realizan en solución y, a partir de estos cálculos, se obtienen las energías libres de Gibbs en solución (ΔG_s), para lo cual en este proyecto se utilizó el modelo continuo SMD⁶⁹ con el programa Gaussian 09.⁷⁰ Se escogió el modelo de solvatación SMD, porque utiliza la densidad electrónica total del soluto en lugar de definir cargas atómicas puntuales, lo que se considera un mejor modelo. Además el disolvente se representa por un medio dieléctrico con tensión superficial en la frontera soluto-solvente.⁶⁹ El SMD es considerado un modelo de solvatación universal, debido a su aplicabilidad a cualquier soluto cargado o no cargado, en cualquier solvente o medio líquido para el cual pocos descriptores clave son conocidos. Este método corresponde a la estrategia conocida como campo de reacción autoconsistente (SCRF, del inglés Self Consistent Reaction Field), donde el término “campo de reacción” se refiere al campo eléctrico que surge como consecuencia de los momentos dipolares que inducen las moléculas de solvente en el soluto y viceversa. Este modelo de solvatación, también toma en cuenta los valores de conectividad de los enlaces entre átomos, cuya definición requiere del uso de campos de fuerza; adicionalmente toma en cuenta una gran cantidad de propiedades como la densidad de carga, índice de refracción, energía libre en solución con base en parámetros de acidez, entre otras.

Las energías libres de Gibbs en solución (ΔG_s) se calcularon con ambas estrategias (*Mixta y Todo en solución*), como se explica anteriormente para cada una de las familias evaluadas en este proyecto. Se incluyen las correcciones termodinámicas a 298.15 K. En todos los casos el estado de referencia utilizado es de 1M. La energía libre de Gibbs en solución para las reacciones de desprotonación a su vez se utiliza para calcular las constantes de acidez (K_a), de acuerdo con:

$$K_a = e^{-\Delta G_s / RT} \quad (4)$$

Y los de pKas se obtienen, según su definición, como:

$$pK_a = -\log(K_a) \quad (5)$$

Las moléculas de interés biológico estudiadas en este proyecto (Tabla 2), fármacos y neurotransmisores, tienen varios sitios de desprotonación. Por lo tanto pueden tener más de

un pK_a . Esto implica que antes de calcular los valores de ΔG_s es necesario evaluar la acidez relativa de estos sitios e identificar el sitio de desprotonación más probable. Para tal fin se modelaron las especies protonadas y se estimaron las energías de desprotonación de los diferentes sitios ácidos. Se seleccionó como más probable en cada caso el sitio con energía de desprotonación más baja. Se incluyeron estas moléculas (fármacos y neurotransmisores) para investigar si la presencia de otros grupos funcionales afecta las predicciones realizadas con las metodologías empleadas, en comparación con aquellas que solo presentan un grupo funcional (el involucrado en el proceso de desprotonación de interés).

Las metodologías que se utilizaron para el cálculo de pK_a s, para este tipo de moléculas (polifuncionales) se escogieron con base a los resultados obtenidos para las moléculas de prueba. Para las moléculas de interés biológico se utilizaron solamente las variantes que proporcionan valores más confiables de pK_a para los conjuntos de moléculas de prueba. Por lo tanto, para cada familia estudiada se seleccionaron los mejores esquemas de reacción y métodos de estructura electrónica (aquellos con errores menores a 2 unidades de pK_a s).

5.1.4. Método de correlaciones

En esta metodología se utilizó el conjunto de moléculas mono-funcionales (MF), para obtener los parámetros m y C_0 de cada familia de compuestos separadamente. Para ellos se hace una correlación lineal de valores calculados de ΔG_{BA} vs. valores experimentales de pK_a :

$$pK_a^{\text{exp}} = m \Delta G_{BA}^{\text{MF}} + C_0 \quad (6)$$

donde $\Delta G_{BA}^{\text{MF}}$ representa la diferencia en energía libre de Gibbs entre cada ácido mono-funcional A y su base conjugada B ($\Delta G_{BA}^{\text{MF}} = G_B^{\text{MF}} - G_A^{\text{MF}}$). Como se ve en esta ecuación la pendiente es igual a m y la ordenada al origen es igual a C_0 . Luego estos dos parámetros se utilizan para calcular el pK_a de un conjunto de moléculas independientes, en nuestro caso las poli-funcionales (PF), según:

$$pK_a^{\text{calc}} = m \Delta G_{BA}^{\text{PF}} + C_0 \quad (7)$$

donde ΔG_{BA}^{PF} representa la diferencia en energía libre de Gibbs entre cada ácido poli-funcional A y su base conjugada B ($\Delta G_{BA}^{PF} = G_B^{PF} - G_A^{PF}$).

Esta metodología se realizó solamente usando SMD y todos los cálculos con solvente implícito y sin moléculas de agua explícitas.

5.1.5. Estimación de errores

Se calcularon los valores de pKa de cada una de las moléculas mostradas en las tablas 1 y 2 para cada par esquema de reacción / método de cálculo. Estos valores se compararon con la data experimental y se estimaron los errores absolutos en cada caso:

$$EA = |\text{pKa}^{\text{calc}} - \text{pKa}^{\text{exp}}| \quad (8)$$

Con estos valores se obtuvo el error absoluto promedio (EAP) para cada par esquema de reacción / método de cálculo como el promedio simple de los valores de EA correspondientes. Se usa el umbral propuesto por Ho y Coote (2 unidades de pKa)²¹ como criterio para identificar las metodologías confiables.

Además se definió un error absoluto promedio global (EAPG) incluyendo simultáneamente los resultados obtenidos con todos los métodos de cálculo para cada esquema de reacción. Esto se hizo, luego de encontrar que los errores cambian poco con el método pero mucho con el esquema, para facilitar las comparaciones entre los numerosos resultados obtenidos.

5.2 CONSTANTES DE ACOMPLEJAMIENTO

5.2.1. Moléculas a estudiar

Para el cálculo de las constantes de equilibrio de complejamiento con el Cu(II) se modeló primeramente un conjunto de ligantes de prueba, que tienen estructuras relativamente simples y además presentan un solo sitio de complejamiento. Estos ligantes relativamente pequeños son amoníaco (NH₃) y acetato (CH₃OO⁻), y se usaron para identificar la mejor metodología mediante la comparación con los valores experimentales, que se conocen para estos casos.

Hay que destacar que estos ligantes han sido seleccionados debido a que sus sitios de complejamiento involucran los mismos átomos (N y O) que la molécula de interés biológico a investigar en el presente proyecto (dopamina). Otro ligante que se tenía contemplado en este proyecto era el fenol, por consistencia con las moléculas que se usaron en la parte de las constantes de acidez, ya que la dopamina también incluye un grupo fenol, sin embargo este ligante se tuvo que descartar, debido a que no existen datos experimentales de constantes de complejamiento reportados en la literatura.

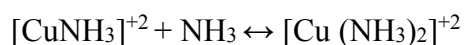
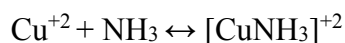
Como se ha mencionado anteriormente, la dopamina puede formar complejos con Cu(II), los cuales son de interés biológico, porque se ha llegado a relacionar este tipo de complejo con enfermedades tales como Alzheimer, Parkinson y Wilson, por consecuencia es de suma importancia para el presente proyecto⁷¹. En la literatura podemos encontrar que para el Cu(II) existe controversia en cómo se encuentra este metal en solución, por esa razón se realizó un estudio conformacional del Cu(II), tanto en forma octaédrica como tetra coordinada plana, en solución acuosa.

5.2.2. Esquemas de reacción

Posteriormente utilizando la geometría más probable de la primera esfera de solvatación para Cu(II) en solución acuosa, se propusieron los posibles esquemas de reacción para los equilibrios de formación sucesiva con los ligantes NH₃ y acetato. En el caso de la dopamina se consideraron tanto complejos mono-dentados como bi-dentados, mientras que NH₃ actúa siempre como ligante mono-dentado y el acetato como ligante bi-dentado. Los diferentes esquemas de reacción se diferencian entre sí en el número de moléculas de agua explícitas (0, 4, 6, 8 y 12) con que se modela al Cu(II). Las primeras 4 moléculas de aguas se corresponden a la esfera de coordinación y el resto a la esfera de solvatación. A continuación se muestran los esquemas de reacción que se propusieron para cada uno de los ligantes.

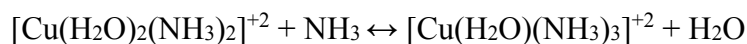
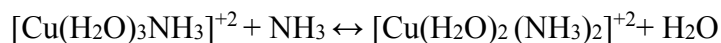
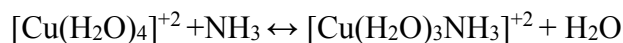
Cu(II) con NH₃:

Equilibrios de Formación Sucesivas Sin Aguas Explícitas (EFS0):

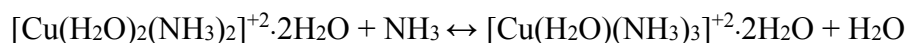
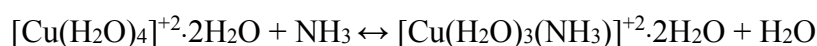




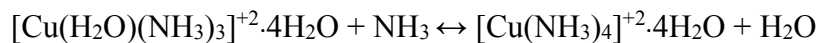
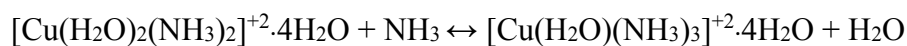
Equilibrios de Formación Sucesivas con 4 Aguas Explícitas (EFS4)



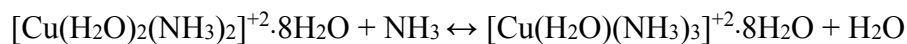
Equilibrios de Formación Sucesivas con 6 Aguas Explícitas (EFS6)



Equilibrios de Formación Sucesivas con 8 Aguas Explícitas (EFS8)

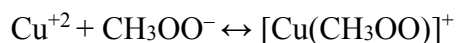


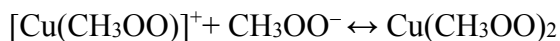
Equilibrios de Formación Sucesivas con 12 Aguas Explícitas (EFS12)



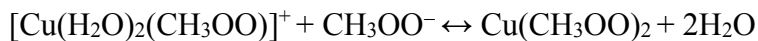
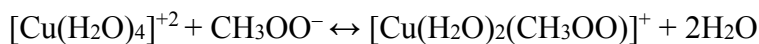
Cu(II) con acetato:

Equilibrios de Formación Sucesivas sin Aguas Explícitas (EFS0)





Equilibrios de Formación Sucesivas con 4 Aguas Explícitas (EFS4)



Equilibrios de Formación Sucesivas con 6 Aguas Explícitas (EFS6)



Equilibrios de Formación Sucesivas con 8 Aguas Explícitas (EFS8)



Equilibrios de Formación Sucesivas con 12 Aguas Explícitas (EFS12)



Para cada esquema de reacción se calculó el valor de ΔG_s correspondiente y con estos valores se obtuvieron las constantes de equilibrio para las reacciones de complejamiento usando la ecuación:

$$K = e^{-\Delta G_s / RT} \quad (9)$$

Al igual que para los equilibrio ácido-base, en este caso incluyen las correcciones termodinámicas a 298.15 K en el cálculo de ΔG_s , y el estado de referencia utilizado es de 1M. También se incluyó la corrección para el estado estándar del agua, es decir 55.55 mol/L para todos los esquemas diferentes a EFS0.

La comparación de los resultados obtenidos al usar cada esquema de reacción con los valores experimentales disponibles, se realizó utilizando el logK. Esta comparación permitió seleccionar los esquemas de reacción con menores errores para cada ligante. Estos fueron los que se utilizaron luego para la molécula de interés biológico contemplada en este proyecto, que es el neurotransmisor dopamina.

Como se puede observar en la Figura 1, la dopamina presenta 3 posibles sitios de complejamiento, los dos grupos OH y el grupo amino. Además, dada la ubicación de los OH en el fenol, la dopamina podría actuar como ligante mono-dentado o bi-dentado. Por lo tanto fue necesario realizar primeramente una evaluación de cuál sería la ruta de complejamiento más probable. O sea identificar el sitio, o los sitios que enlazan más fuertemente con Cu(II) y también si el complejamiento involucra uno o dos sitios. Una vez obtenida la ruta de complejamiento más probable, se utilizaron los funcionales de la densidad identificados como los que mejor reproducen los valores experimentales de ligantes simples.

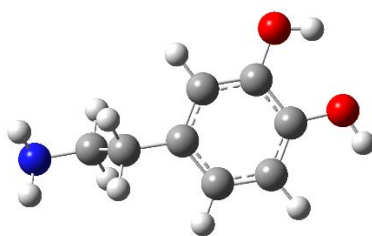


Figura 1. Estructura de la dopamina en forma neutra.

5.2.3. Métodos de cálculo

Las energías necesarias, para todas las metodologías que se probaron en el cálculo de las constantes de complejamiento, se obtuvieron con el paquete computacional Gaussian 09.⁷⁰ Además se usaron diversos funcionales de la densidad. Por consistencia con las constantes de acidez se decidió utilizarlos siguientes funcionales: BLYP, B3LYP, BHandHLYP, PBE, PBE0, PW91, BMK, TPSS; sin embargo se decidió intercambiar el funcional M05-2X por M06 y M05 (Tabla 6), debido a que estos fueron parametrizados para metales de transición, por lo que se espera que estos proporcionen resultados más confiables en el caso de las constantes de complejamiento. En todos los casos los funcionales antes mencionados se utilizaron para realizar las optimizaciones de geometría y los cálculos de frecuencia en combinación con el conjunto de funciones de bases 6-311+G (d) y el modelo continuo de solvatación SMD.

Tabla 6. Características de los funcionales de la densidad usados en la metodología de constantes de acidez.

FDA	% Ex ^{HF}	Parametros Optimizados	Clasificación	Año	Escalón	Referencia
M05	28	20	<i>Hybrid-meta-GGA</i>	2006	4	72
M06	27	38	<i>Hybrid-meta-GGA</i>	2008	4	73

Debido al gran número de cálculos que se pretendían realizar en este proyecto se buscó una estrategia para minimizar el tiempo de cómputo. Esta estrategia se basó en verificar si los parámetros geométricos para cada molécula varían significativamente, o no, al cambiar de funcional a utilizar. Para ello se seleccionó la molécula $[\text{Cu}(\text{H}_2\text{O})_2(\text{NH}_3)_2]^{2+}$ ya que es sencilla y presenta dos ligantes diferentes (Figura 2). De comprobarse que los parámetros geométricos son muy similares, con solo un método de cálculo se realizarían las optimizaciones de geometría y los cálculos de frecuencia, y para el resto de los funcionales propuestos se harían solamente cálculos de punto simple. Sin embargo, de no ser así habría que hacer las optimizaciones de geometría y los cálculos de frecuencia con todos los métodos de cálculo.

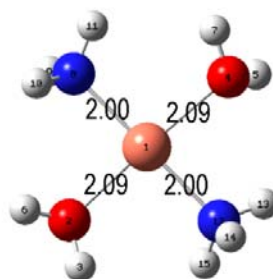


Figura 2. Estructura de $[\text{Cu}(\text{H}_2\text{O})_2(\text{NH}_3)_2]^{2+}$.

5.2.4. Estimación de errores

Se utilizaron las mismas estrategias que para los equilibrios de acidez, solo que en este caso los valores de EA se obtuvieron como:

$$EA = |\log K^{\text{calc}} - \log K^{\text{exp}}| \quad (8)$$

Donde K representa las constantes de equilibrios de formación sucesivos. Los valores de $\log K$ experimentales se reportan en el Anexo 2. Con estos valores se obtuvo el EAP para cada par esquema de reacción / método de cálculo como el promedio simple de los valores de EA correspondientes, y con los valores de EAP se calcularon los EAPG para cada esquema.

6- RESULTADOS Y DISCUSIÓN

6.1 Constantes de Acidez

6.1.1. Fenoles

Para este tipo de moléculas la especie ácida es neutra y la base conjugada es aniónica, por lo que se calcularon los valores de pKas utilizando los esquemas de reacción y las ecuaciones mostrados en la Tabla 3. Los errores absolutos se analizaran primero separadamente para las moléculas mono-funcionales y poli-funcionales, y luego de manera conjunta. Los resultados se presentan separadamente para las dos estrategias usadas en el cálculo de energías de solvatación: (i) usando ciclos termodinámicos junto con el modelo continuo PCM y (ii) todo en solución con SMD.

6.1.1.1. Mono-funcionales PCM

En la Figura 4 se muestra una gráfica con los EAP obtenidos para este tipo de moléculas con todos los métodos de cálculo y esquemas de reacción. Todos los valores de EAP obtenidos se reportan en el Anexo 3, mientras que en la tabla 7 se muestran todos los esquemas de reacción con los que se encontró al menos un valor de $EAP < 2$. En esta tabla se muestran en color azul todos los valores menores a 2 unidades de pKa y en color azul y letras en negrita los valores menores a 1 unidades de pKa. Este mismo modo de resaltar se emplea en las otras tablas de la tesis que reportan valores de EAP.

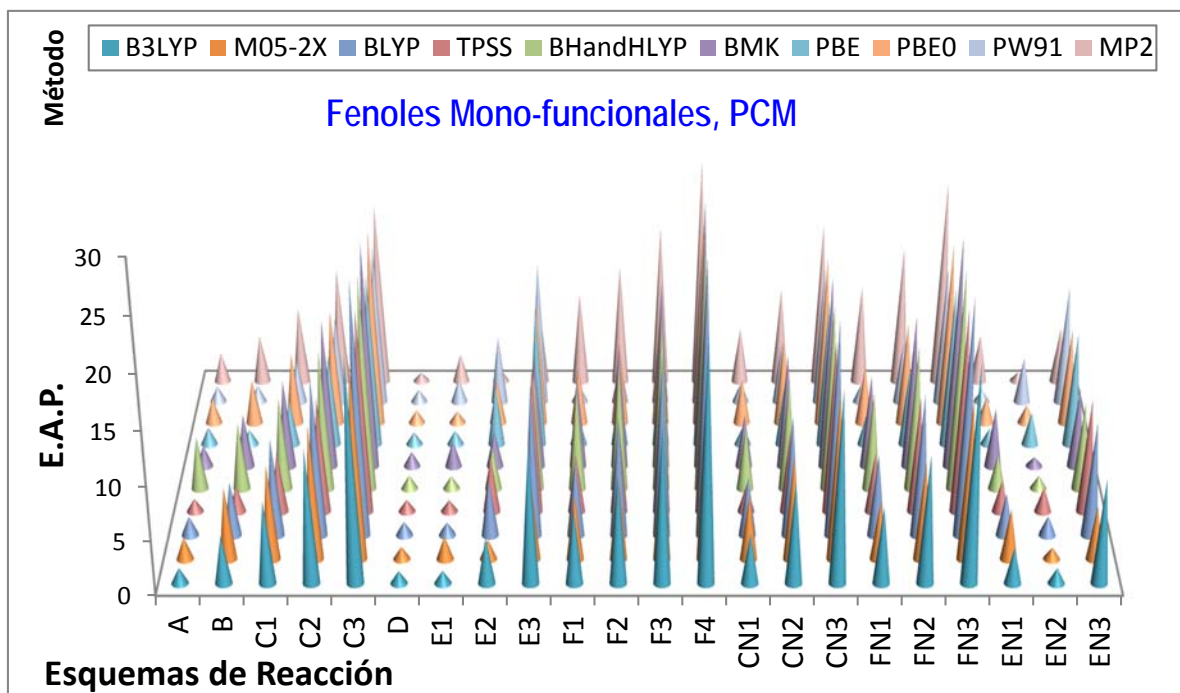


Figura 4. Errores absolutos de los valores de pKa calculados con la estrategia de PCM y ciclos termodinámicos para todos los fenoles mono-funcionales.

Tabla 7. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para fenoles mono-funcionales, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	M05-2X	BLYP	TPSS	BH&HLYP	BMK	PBE	PBE0	PW91	MP2	EAPG
A	1.46	2.04	1.72	1.22	4.92	1.96	1.63	2.35	1.49	2.74	2.15
B	4.45	6.63	4.96	2.41	6.21	4.94	1.38	4.15	1.59	4.58	4.13
D	1.17	1.16	1.24	1.14	1.25	1.36	1.14	1.21	1.17	0.80	1.16
E1	1.20	2.14	1.15	1.08	1.20	2.46	1.31	1.08	2.16	2.62	1.64
E2	4.30	1.84	5.22	5.24	3.91	3.38	5.61	4.09	6.46	1.13	4.12
F1	6.85	5.97	6.96	5.38	9.41	8.59	2.68	5.33	1.54	8.60	6.13
CN1	4.66	6.02	5.49	2.58	5.90	4.93	1.67	4.28	1.81	5.25	4.26
FN1	7.06	5.35	7.49	5.55	9.10	8.59	2.89	5.45	1.78	9.27	6.25
EN1	3.38	4.63	3.88	2.88	3.25	5.50	1.52	2.49	1.12	4.62	3.33
EN2	1.45	1.10	1.77	2.29	1.22	0.86	3.05	1.71	4.26	0.85	1.86

Como muestra la Figura 4, el método de cálculo tiene una moderada influencia sobre la calidad de los resultados mientras que la de los esquemas de reacción tiene una influencia marcada. Los esquemas de reacción que permitieron obtener los errores más bajos son **A**, **D**, **E1** y **EN2**. Dos de ellos (**E1** y **EN2**) son propuestos en este proyecto. Los valores de EAPG que se obtuvieron en estos casos son: **A** (2.15), **D** (1.16), **E1** (1.64) y **EN2** (1.85). Analizaremos en detalle cada uno de ellos.

Para el esquema **A** se encontraron seis funcionales de la densidad con valores de EAP < 2: B3LYP, BLYP, TPSS, BMK, PBE y PW91; obteniéndose el menor EAP (1.22) con TPSS. Para el esquema **E1** también se encontraron 6 funcionales con EAP < 2: B3LYP, BLYP, TPSS, BHandHLYP, PBE y PBE0, siendo TPSS y PBE0 los que producen el error más bajo (1.08). En el caso del esquema **EN2**, los métodos de cálculo con EAP < 2 son nuevamente 6, MP2 y 5 funcionales de la densidad (B3LYP, BLYP, M05-2X, BHandHLYP y PBE0). En este caso los mejores valores se logran con MP2 (0.85) y BMK (0.86). De hecho en estos dos casos se obtienen valores de EAP menores a 1 unidad de pKa. El esquema **D** es el único caso en el que se obtienen valores de EAP < 2 para todos los métodos de cálculo utilizados. El mejor resultado lo produce MP2 (EAP=0.80), mientras que de los métodos DFT los valores más bajos de EAP se obtienen con TPSS y PBE (EAP=1.14 en ambos casos). El hecho de que para el esquema **D** el número y tipo de especies cargadas a ambos lados de equilibrio sea el mismo parece ser el factor determinante en la buena calidad de los resultados obtenidos. Esta característica también la cumplen los esquemas **E1** y **EN2**. Además es interesante que en el caso de los esquemas **E**, se obtienen mejores resultados con una sola molécula de agua explícita solvatando al anión (**E1**) que con un mayor número de moléculas de agua (**E2** y **E3**). Por el contrario, en el caso de los esquemas **EN** el número óptimo de moléculas de agua es 2.

De los 36 pares esquema de reacción / método de cálculo con valores de EAP < 2 los 10 mejores resultados individuales son:

- el esquema **D** en combinación con MP2, TPSS y PBE;
- el esquema **EN2** con los métodos MP2, BMK y M05-2X;
- el esquema **E1** con PBE0 y TPSS;
- el esquema **EN1** con el funcional PW91; y

- el esquema **E2** con el método MP2.

6.1.1.2 Mono-funcionales SMD

Los EAP obtenidos en este caso se muestran en la Figura 5. Los valores se reportan en el Anexo 4 y en la tabla 8 se muestran todos los esquemas de reacción con los que se encontró al menos un valor de EAP < 2. Al igual que ocurría al modelar el solvente con PCM y usar ciclos termodinámicos para el cálculo de la energía de solvatación, la influencia del esquema de reacción en los valores de EA es mucho mayor que la del método de estructura electrónica. Sin embargo al utilizar SMD los esquemas que dan mejores resultados son **D**, **E1**, **E2**, **E3** y **EN3**. Todos ellos tienen idéntico número y tipo de especies cargadas a ambos lados del equilibrio y todos menos el esquema **D** son propuestos en este proyecto. Los valores de EAPG que se obtuvieron en estos casos son: **D** (0.61), **E1** (1.81), **E2** (1.15), **E3** (1.88) y **EN3** (1.55).

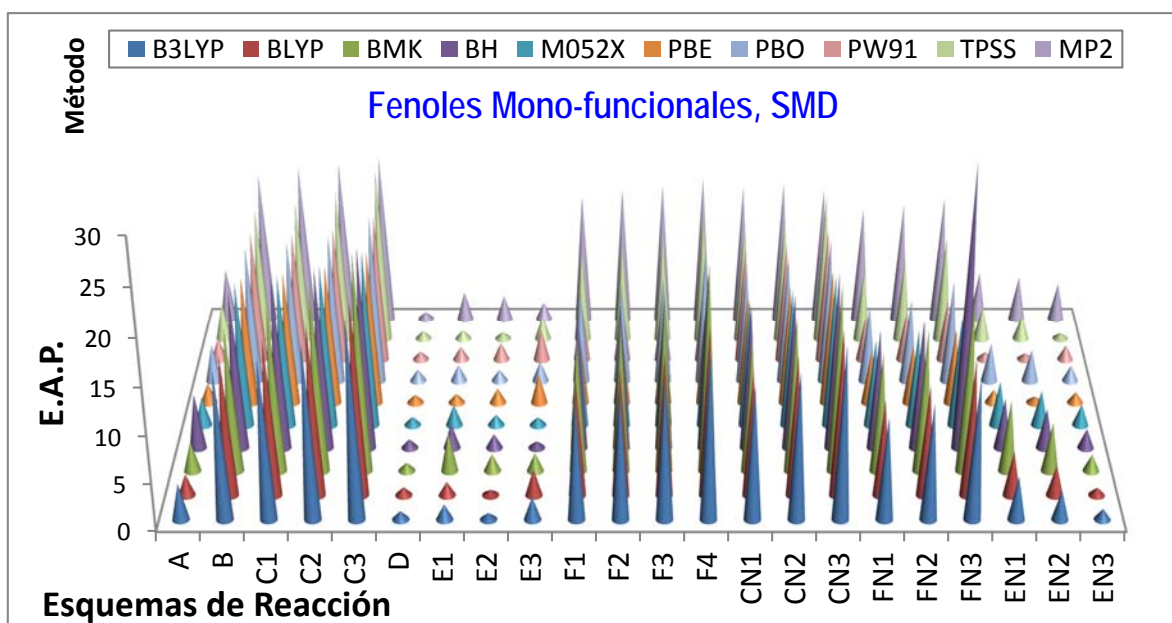


Figura 5. Errores absolutos de los valores de pKa calculados con la estrategia de todo en solución y usando SMD para todos los fenoles mono-funcionales.

Tabla 8. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para fenoles mono-funcionales, usando SMD (todo en solución) en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BMK	BH&HLYP	M05-2X	PBE	PBE0	PW91	TPSS	MP2	EAPG
D	0.63	0.62	0.68	0.63	0.55	0.59	0.56	0.63	0.56	0.67	0.61
E1	1.62	1.36	3.70	2.23	1.79	0.85	1.14	1.41	0.98	3.05	1.81
E2	0.73	0.48	1.39	0.96	1.09	1.49	0.53	1.85	0.43	2.50	1.15
E3	1.98	2.62	1.51	0.81	0.57	2.70	1.57	2.76	2.26	1.98	1.88
EN1	4.17	4.18	6.78	4.75	4.29	0.92	3.43	0.67	3.32	4.88	3.74
EN2	2.63	2.76	4.48	3.47	3.59	0.46	2.52	0.64	2.04	4.33	2.69
EN3	0.74	0.93	2.10	1.72	2.23	0.97	0.99	1.32	0.72	3.81	1.55

Para el esquema **D** todos los valores de EAP obtenidos son no solo menores a 2 unidades de pKa, sino también menores a 1 unidad de pKa. Los menores valores de EAP se obtuvieron con M05-2X (0.55), PBE0 (0.56) y TPSS (0.56), mientras que el máximo EAP se obtuvo con BMK y es solo de 0.68 unidades de pKa. En el caso del esquema **E2**, solo dos métodos (PW91 y MP2) tiene valores de EAP > 2 y cinco métodos producen valores de EAP < 1 (B3LYP, BLYP, BHandHLYP, PBE0 y TPSS). Los valores más bajos de EAP se obtienen para este esquema con TPSS (0.43), BLYP (0.48) y PBE0 (0.53). Para el esquema **EN3**, se obtienen valores de EAP < 2 con 7 de los 10 métodos de estructura electrónica probados y de ellos 5 con EAP < 1 (B3LYP, BLYP, PBE, PBE0 y TPSS). Los menores valores de EAP se corresponden a TPSS (0.72) y B3LYP (0.74). En el caso del esquema **E1**, 7 funcionales de la densidad producen EAP < 2, y de ellos 2 (PBE y TPSS) tienen EAP < 1, los valores son 0.85 y 0.98 respectivamente. Para el esquema **E3** se obtienen valores de EAP < 2 con 5 métodos de cálculo y los mejores valores con M05-2X (0.57) y BHandHLYP (0.81), que son además los únicos con EAP < 1.

Además es importante destacar que, en el caso de fenoles mono-funcionales, los EAP obtenidos usando SMD son significativamente más bajos que los que se obtienen con PCM y ciclos termodinámicos. Con SMD se encontraron 41 pares esquema de reacción / método de cálculo con valores de EAP < 2 (36 con PCM) y de ellos 28 con los EAP < 1 (solo 3 con

PCM). Además 20 de los valores de EAP obtenidos con SMD son menores que el mejor valor con PCM (0.80). Los 10 mejores resultados individuales con SMD son:

- el esquema **D** en combinación con M05-2X, BLYP, PBE, PBE0 y TPSS;
- el esquema **E2** en combinación con BLYP, PBE0 y TPSS;
- el esquema **E3** con el funcional M05-2X; y
- el esquema **EN2** con el funcional PBE.

6.1.1.3. Poli-funcionales PCM

Los resultados obtenidos se muestran de forma resumida en la Figura 6, los valores se reportan en el Anexo 5 y se muestran todos los esquemas de reacción con los que se encontró al menos un valor de EAP < 2 en la tabla 9. La figura 6 muestra que los resultados obtenidos son similares a los que ya se reportaron para fenoles mono-funcionales. Cabe notar, que como ya se demostró anteriormente que MP2 no es mejor que los diferentes funcionales de la densidad utilizados y los cálculos de las moléculas poli-funcionales son computacionalmente más costosos, MP2 no se incluyó en esta parte.

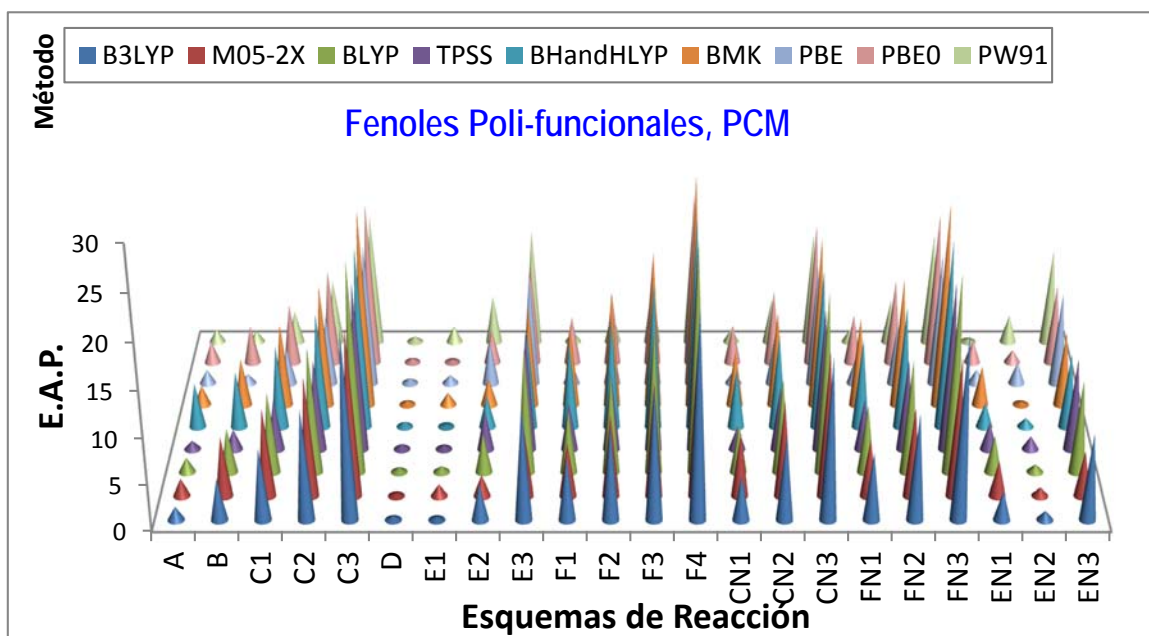


Figura 6. Errores absolutos promedio de los pKa calculados con la metodología de ciclos termodinámicos y el modelo continuo PCM para fenoles poli-funcionales.

Tabla 9. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para fenoles poli-funcionales, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	M05-2X	BLYP	TPSS	BH&HLYP	BMK	PBE	PBE0	PW91	EAPG
A	1.46	1.85	1.64	0.94	4.81	2.14	1.51	2.34	1.44	2.01
B	4.46	6.44	4.80	2.19	6.10	5.12	1.04	4.14	1.20	3.94
D	0.45	0.32	0.49	0.55	0.52	0.45	0.57	0.43	0.59	0.49
E1	0.34	1.21	0.64	0.43	0.23	1.35	0.91	0.29	1.81	0.80
F1	6.86	5.77	6.80	5.16	9.30	8.78	2.36	5.31	1.14	5.72
CN1	4.84	6.85	5.32	2.50	6.44	6.09	1.42	4.31	1.46	4.36
FN1	7.24	6.19	7.32	5.47	9.64	9.75	2.74	5.49	1.41	6.14
EN1	3.20	3.79	4.04	2.95	2.71	4.34	1.54	2.45	0.32	2.82
EN2	1.01	0.89	0.85	1.43	0.97	0.34	2.35	1.40	3.17	1.38

Los esquemas de reacción que permitieron obtener los errores más bajos son **A**, **D**, **E1** y **EN2** pero en este caso los valores de EAPG que se obtuvieron son menores con respecto a lo obtenido con PCM y ciclos termodinámicos: **A** (2.01), **D** (0.49), **E1** (0.80) y **EN2** (1.38). Dado que las moléculas poli-funcionales estudiadas son de mayor tamaño que las mono-funcionales, esto sugiere que mientras mayor sea la concentración de cargas en las especies iónicas mayor será el error que se comete en los cálculos de pKa.

En el caso del esquema **D** los EAP son todos menor a la unidad de pKa con valores desde 0.32 (M05-2X) hasta 0.59 (PW91). De modo que para este esquema y fenoles poli-funcionales puede utilizarse de manera confiable cualquiera de los funcionales probados. Para el esquema **A** se encontraron seis funcionales de la densidad con valores de EAP < 2: B3LYP, M05-2X, BLYP, TPSS, PBE y PW91. El menor valor de EAP (0.94) corresponde a los cálculos TPSS y es el único caso con este esquema para el que EAP es menor a la unidad de pKa. Al igual que para el esquema **D**, EAP < 2 para el esquema **E1** con todos los funcionales, siendo BHandHLYP y PBE0 los que producen el error más bajo (0.23 y 0.39, respectivamente). En el caso del esquema **EN2**, se obtienen valores EAP < 2 con la mayoría los métodos de cálculo. Las únicas excepciones son PBE y PW91. En este caso los mejores

valores se logran con BMK (0.34) y BLYP (0.85). Además de estos, también con los funcionales M05-2X y BHandHLYP se obtienen valores de EAP < 1.

Para este conjunto de fenoles se encontraron 39 pares esquema de reacción / método de cálculo con valores de EAP < 2, de ellos 21 con valores de EAP < 1. Los 10 mejores resultados individuales son:

- el esquema **D** en combinación con B3LYP, M05-2X, BMK y PBE0;
- el esquema **E1** con B3LYP, BHandHLYP, PBE0 y TPSS;
- el esquema **EN1** con el funcional PW91; y
- el esquema **EN2** con el método BMK.

Comparando los resultados obtenidos con la estrategia de ciclos termodinámicos y modelo de solvente PCM para fenoles mono-funcionales y poli-funcionales, existen varias combinaciones de esquema de reacción / método de cálculo que en ambos casos dan resultados con valores de EAP razonablemente buenos. Estas son: esquema **D** con B3LYP, PBE, PBE0 y TPSS; esquema **E1** con B3LYP, PBE0 y TPSS; esquema **EN2** con BMK y esquema **A** con TPSS. Esto soporta la hipótesis planteada.

6.1.1.4. Poli-funcionales SMD

Los valores de EAP se muestran de manera resumida en la Figura 7, los valores se reportan en el Anexo 6 y en la tabla 10 se muestran los datos para todos los esquemas de reacción con al menos un valor de EAP < 2. Los esquemas que dan mejores resultados en este caso son **D**, **E1**, **E2** y **EN3**. Los valores de EAPG que se obtuvieron en estos casos son 0.38, 1.50, 0.98 y 1.33, respectivamente.

Tanto para el esquema **D** como para el esquema **EN3** todos los valores de EAP obtenidos son menores a 2 unidades de pKa. Para el esquema **D** además todos son menores a 1 unidad de pKa. Los mejores resultados para este esquema se obtuvieron con los funcionales BLYP (0.18), PBE0 (0.23), B3LYP (0.27) y BMK (0.29). Sin embargo, al igual que ocurre para fenoles mono-funcionales, en el caso de los poli-funcionales el esquema **D** puede utilizarse con cualquiera de los métodos de cálculo obteniéndose buen acuerdo con los datos

experimentales de pKas. De modo que este esquema parece ser la mejor opción para el cálculo de compuestos fenólicos en general. En el caso del esquema *EN3* los mejores resultados se obtienen con B3LYP (EAP=0.89) y TPSS (EAP=0.60).

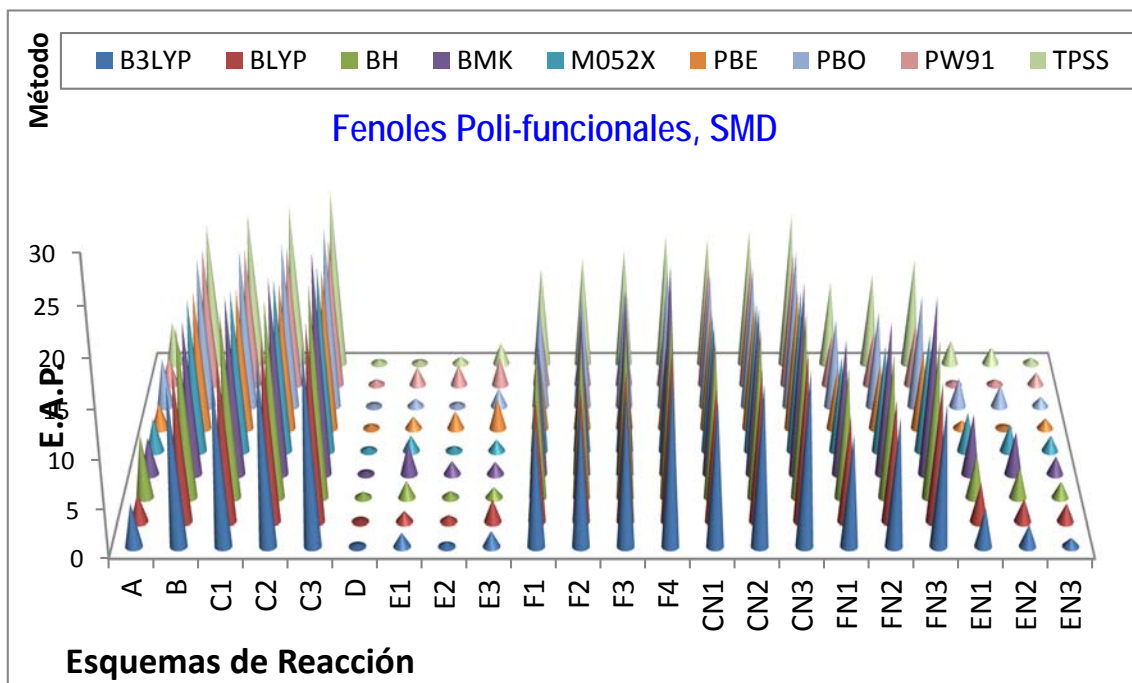


Figura 7. Errores absolutos promedio de pKa calculados con SMD para fenoles poli-funcionales.

Tabla 10. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para fenoles poli-funcionales, usando SMD en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BH&HLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	EAPG
D	0.27	0.18	0.59	0.29	0.42	0.44	0.23	0.58	0.44	0.38
E1	1.45	1.10	1.78	3.24	1.66	1.25	0.80	1.88	0.38	1.50
E2	0.47	0.67	0.68	1.39	0.53	1.84	0.37	2.11	0.74	0.98
E3	1.64	2.34	1.10	1.30	1.21	3.15	1.91	2.87	2.34	1.98
EN1	3.97	4.05	4.57	6.27	4.07	0.83	3.04	0.40	2.63	3.31
EN2	2.28	2.37	3.22	4.42	2.81	0.24	2.26	0.56	1.81	2.22
EN3	0.89	1.91	1.69	1.86	1.62	1.07	1.00	1.32	0.60	1.33

Para los esquemas **E1** y **E2** los valores de EAP obtenidos con todos los funcionales de la densidad menos uno (BMK y PW91, respectivamente) son menores a 2 unidades de pKa. Los mejores resultados para el esquema **E1** se obtuvieron con TPSS (0.38) y PBE0 (0.80) y son los únicos métodos que dan valores de EAP <1 con este esquema. Por su parte, con el esquema **E2** se obtuvieron valores de EAP <1 con 6 funcionales de la densidad (B3LYP, BLYP, BHandHLYP, M05-2X, PBE0 y TPSS). Los mejores resultados en este caso corresponden a PBE0 (EAP = 0.37) y B3LYP (EAP = 0.47).

Comparando los resultados obtenidos cuando las energías de solvatación se obtienen con SMD para fenoles mono-funcionales y poli-funcionales, se encontraron varios pares esquema de reacción / método de cálculo que en ambos casos dan resultados en buen acuerdo con los datos experimentales. Estos son: el esquema **D** con cualquier funcional de la densidad, y en particular con BLYP, PBE, PBE0 y TPSS; el esquema **E1** con PBE0 y TPSS; el esquema **E2** con B3LYP, PBE0 y TPSS; el esquema **EN1** con PW91; el esquema **EN2** con PBE y PW91 y el esquema **EN3** con B3LYP, PBE0 y TPSS. Esto soporta la hipótesis planteada.

Por otra parte, comparando los resultados obtenidos para fenoles poli-funcionales, con los modelos de solvente SMD y PCM, se observa que hay tendencias que se mantienen para ambas estrategias. En ambos casos los esquemas de reacción **D** y **E1** se encuentran dentro de los que mejor reproducen los datos experimentales. Además se obtienen valores de EAP <1 para el esquema **D** con todos los funcionales de la densidad probados, para el esquema **E1** con los funcionales PBE0 y TPSS y para el esquema **EN1** con el funcional PW91.

Sin embargo también se encontraron algunas diferencias como consecuencia de cambiar la forma de modelar el solvente. Por ejemplo, mientras que el esquema **A** da resultados aceptables cuando se usa con PCM, al usarlo con SMD el error es mayor al límite aceptado en la literatura. Además, de manera general (y tal como ocurre para fenoles mono-funcionales) las desviaciones con respecto al experimento para los esquemas **D** y **E** (que son los que mejor funcionan para fenoles) son menores con SMD que con PCM.

Al usar SMD en el cálculo de fenoles poli-funcionales se encontraron 44 pares esquema de reacción / método de cálculo con valores de EAP < 2 y de ellos 23 con los EAP < 1. Los 10 mejores resultados individuales en este caso son:

- el esquema **D** en combinación con B3LYP, BLYP, BMK, M05-2X PBE, PBE0 y TPSS;
- el esquema **E1** en combinación con TPSS;
- el esquema **E2** con el funcional PBE0;
- el esquema **EN1** con el funcional PW91; y
- el esquema **EN2** con el funcional PBE.

6.1.1.5. Conjunto completo de fenoles con PCM

El haber encontrado que existen estrategias para el cálculo de pKas que conducen a buenos resultados tanto para las moléculas mono-funcionales como para las poli-funcionales permite fundamentar la hipótesis planteada, al menos para el caso de compuestos fenólicos. Por lo tanto en esta sección se investigan todos los fenoles estudiados como un conjunto único para poder hacer recomendaciones en cuanto a las mejores estrategias de cálculo para fenoles en general. Los valores de EAP obtenidos para este caso se muestran de manera resumida en la Figura 8, los valores se reportan en el Anexo 7 y la tabla 11 contiene los datos para todos los esquemas de reacción con al menos un valor de EAP < 2.

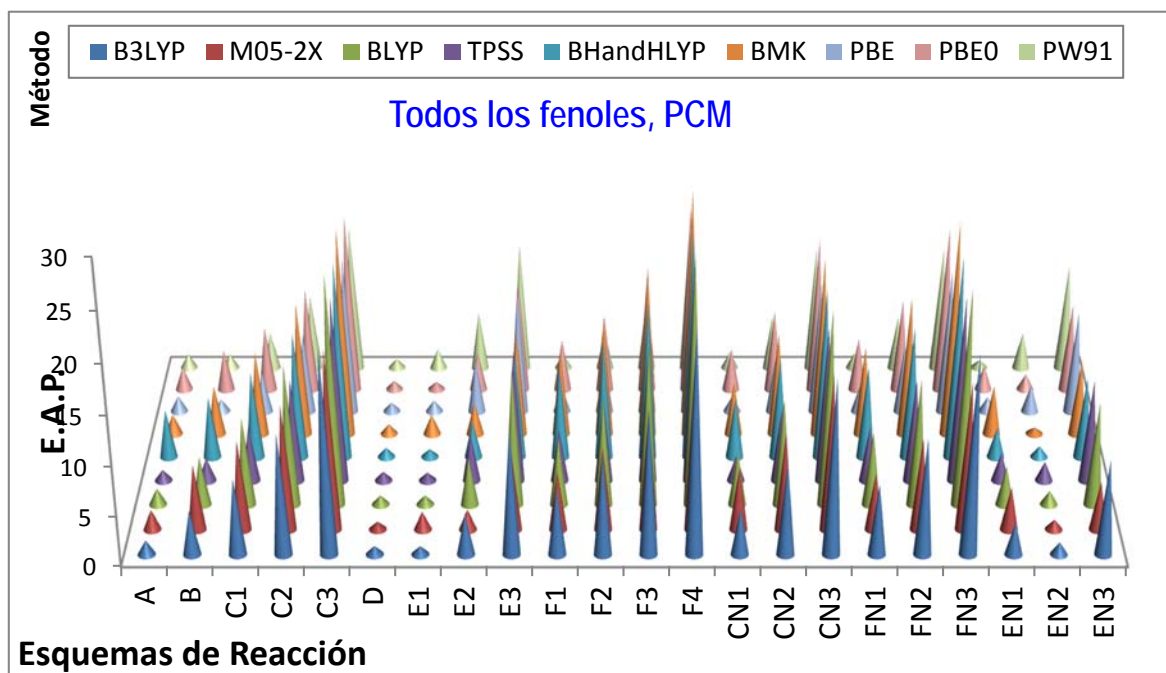


Figura 8. Errores absolutos promedio de los valores de pKa calculados para el conjunto completo de fenoles, cuando la energía de solvatación es calculada con PCM.

Tabla 11. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para el conjunto completo de fenoles, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	M05-2X	BLYP	TPSS	BH&HLYP	BMK	PBE	PBE0	PW91	EAPG
A	1.46	1.97	1.69	1.11	4.88	2.03	1.58	2.35	1.47	2.06
B	4.45	6.55	4.89	2.33	6.17	5.01	1.24	4.14	1.43	4.02
D	0.88	0.82	0.94	0.90	0.96	1.00	0.91	0.90	0.94	0.92
E1	0.86	1.77	0.95	0.82	0.82	2.02	1.15	0.77	2.02	1.24
F1	6.85	5.89	6.89	5.30	9.36	8.67	2.55	5.32	1.38	5.80
CN1	4.73	6.35	5.42	2.55	6.12	5.40	1.57	4.29	1.67	4.23
FN1	7.13	5.69	7.42	5.52	9.31	9.05	2.83	5.47	1.63	6.01
EN1	3.31	4.29	3.94	2.91	3.04	5.04	1.53	2.47	0.80	3.04
EN2	1.27	1.01	1.40	1.94	1.12	0.65	2.77	1.59	3.82	1.73

Como era de esperarse los esquemas que dan mejores resultados en este caso (**D**, **E1** y **EN2**) son consistentes con los análisis realizados anteriormente. Los valores de EAPG correspondientes son 0.92, 1.24 y 1.73. Para el esquema **D** los valores de EAP son menores a 1 unidad de pKa con todos los funcionales de la densidad, excepto para BMK que tiene EAP = 1. Los valores más bajos de EAP se obtuvieron con B3LYP (0.88), M05-2X (0.82) y PBE0 (0.90). En el caso del esquema **E1** solo BMK y PW91 dan valores de EAP superiores a 2 unidades de pKa. Los mejores resultados en este caso se obtuvieron con PBE0 (EAP=0.77), TPSS y BHandHLYP (EAP=0.82 en ambos casos). En cuanto al esquema **EN2** el único método con EAP < 1 es el BMK (EAP=0.65).

Para el conjunto completo de fenoles, usando PCM y ciclos termodinámicos para el cálculo de la energía de solvatación se encontraron 35 pares esquema de reacción / método de cálculo con valores de EAP < 2 y de ellos 15 con los EAP < 1. Los 10 mejores resultados se obtienen con:

- el esquema **D** en combinación con B3LYP, M05-2X, PBE0 y TPSS;
- el esquema **E1** en combinación con B3LYP, BHandHLYP, PBE0 y TPSS;

- el esquema *EN1* con el funcional PW91; y
- el esquema *EN2* con el funcional BMK.

6.1.1.6. Conjunto completo de fenoles con SMD

Para este caso los valores de EAP obtenidos se muestran de manera resumida en la Figura 9, los valores se reportan en el Anexo 8 y los datos para todos los esquemas de reacción con al menos un valor de $EAP < 2$ se reportan en la tabla 12. En general los valores obtenidos con SMD se encuentran en mejor acuerdo con los valores experimentales que los obtenidos con PCM, no solo el número de combinaciones esquema de reacción /método de cálculo con $EAP < 2$ es mayor, sino que también los valores específicos de EAP para combinación son más bajos.

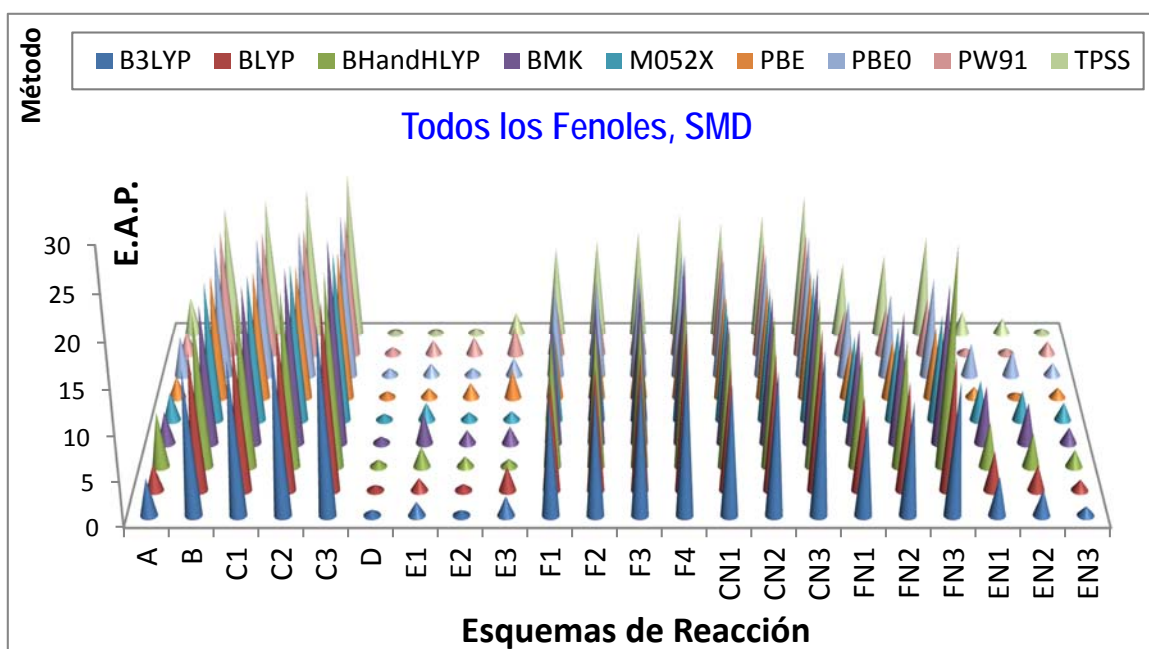


Figura 9. Errores absolutos de los valores de pKa calculados para el conjunto completo de moléculas fenólicas, cuando se usa SMD para calcular las energías de solvatación.

Tabla 12. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para el conjunto completo de fenoles, usando SMD en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BMK	BH&HLYP	M05-2X	PBE	PBE0	PW91	TPSS	EPAG
D	0.48	0.50	0.61	0.53	0.50	0.53	0.43	0.61	0.51	0.52
E1	1.55	1.38	2.05	3.18	1.74	1.01	1.00	1.60	0.52	1.56
E2	0.63	0.64	0.85	1.29	0.86	1.63	0.47	1.96	0.55	0.99
E3	1.84	2.42	0.92	1.57	0.82	2.88	1.71	2.81	2.29	1.92
EN1	4.09	4.18	4.68	6.20	4.20	0.88	3.27	0.56	2.40	3.38
EN2	2.49	2.61	3.37	4.15	3.27	0.37	2.41	0.61	1.53	2.31
EN3	0.80	1.10	1.70	1.81	1.99	1.01	1.00	1.32	0.67	1.27

Los esquemas que dan mejores resultados en este caso son *D*, *E1*, *E2* y *EN3* y sus valores de EAPG son 0.52, 1.56, 0.99 y 1.27, respectivamente. Para el esquema *D* los valores de EAP son menores a 1 unidad de pKa con todos los funcionales de la densidad y varían de 0.43 a 0.61. Los mejores resultados se obtuvieron con B3LYP (0.48) y PBE0 (0.43). Para el esquema *E1* con 7 de los 9 funcionales probados se obtuvieron valores de EAP < 2, siendo TPSS el método con el error más bajo (0.52). En el caso del esquema *E2* solo BMK y BHandHLYP dan valores de EAP superiores a 1 unidad de pKa, aunque menores a 2 unidades. Los mejores resultados en este caso se obtuvieron con PBE0 (EAP=0.47) y TPSS (EAP=0.55). Para el esquema *EN3* todos los métodos tienen EAP < 2, pero solo B3LYP (0.65) y TPSS (0.67) tienen EAP < 1.

Para el conjunto completo de fenoles, usando SMD para el cálculo de la energía de solvatación se encontraron 44 pares esquema de reacción / método de cálculo con valores de EAP < 2 y de ellos 24 con los EAP < 1. Los 10 mejores resultados se obtienen con:

- el esquema *D* en combinación con B3LYP, BLYP, BHandHLYP, M05-2X, PBE, PBE0 y TPSS;
- el esquema *E1* con el funcional TPSS;
- el esquema *E2* con el funcional PBE0; y
- el esquema *EN2* con el funcional PBE.

6.1.1.7. Método de correlaciones para fenoles

Los resultados obtenidos con el método de las correlaciones para el cálculo de pKas de compuestos fenólicos se muestran en la tabla 13. Con todos los métodos de cálculo probados se encontró que los coeficientes de correlación (R^2) son mayores que, o cercanos a, 0.9. El acuerdo con los valores experimentales es muy bueno con todos los valores de EAP siendo menores a 0.4. El valor más bajo de EAP se obtuvo con B3LYP (0.06) y el más alto con BMK (0.35). Además los valores obtenidos con los funcionales PBE0 y TPSS, que son los que más frecuentemente dan buenos resultados con la estrategia de los esquemas de reacción, también son muy buenos (0.20 y 0.24, respectivamente). De hecho estos valores son menores que los mejores obtenidos con cualquiera de los esquemas de reacción. Esto indica que, si bien esta metodología tiene la desventaja de necesitar datos experimentales para obtener los parámetros empíricos, da muy buenos resultados y podría usarse de manera confiable para predecir valores de pKas de compuestos fenólicos para los que no se tiene esta información.

Tabla 13. Parámetros m y C_0 , coeficiente de correlación lineal (R^2) y errores absoluto promedio para el cálculo de pKas de fenoles usando el método de las correlaciones.

	m	C_0	R^2	EAP
B3LYP	0.388	-102.501	0.963	0.06
BLYP	0.411	-108.302	0.885	0.16
BH&HLYP	0.417	-112.050	0.942	0.28
BMK	0.412	-109.073	0.988	0.35
M05-2X	0.443	-117.798	0.945	0.10
PBE	0.404	-106.042	0.906	0.26
PBE0	0.385	-101.995	0.960	0.20
PW91	0.365	-94.856	0.860	0.30
TPSS	0.388	-102.535	0.923	0.24

6.1.1.8. Generalidades para fenoles

Sin importar el conjunto de fenoles investigados, de los diferentes aspectos explorados en el presente trabajo, el esquema de reacción es el que tiene mayor influencia a la hora de obtener buena correspondencia entre los valores de pKa calculados y los datos experimentales. El factor clave a la hora de diseñar un esquema de reacción es que el número y tipo de especies cargadas sea idéntico a ambos lados del equilibrio ácido base.

El segundo aspecto en importancia parece ser el modelo continuo de solvente, siendo SMD mejor que PCM para el cálculo de pKas de compuestos fenólicos. Los esquemas **D** (previamente reportado en la literatura), **E1** y **EN2** (propuestos en este trabajo) son con los que se obtienen sistemáticamente los mejores resultados cuando el solvente se modela con PCM. Cuando se usa SMD para este propósito, los mejores acuerdos con los datos experimentales se obtienen con los esquemas **D**, **E1**, **E2** y **EN3** (los 3 últimos son propuestos en este trabajo). De manera que sin importar cómo se calcularon las energías de solvatación los esquemas **D** y **E1** dan buenos resultados en el cálculo de pKas de compuestos fenólicos. Sin embargo, es importante insistir en el hecho de que la selección del ácido de referencia es crucial para obtener buenos resultados cuando se usa el esquema **D**.

En tercer lugar se encuentra el método de estructura electrónica. De los funcionales de la densidad estudiados, con PBE0 y TPSS se obtienen sistemáticamente buenos resultados, frecuentemente con valores de EAP menores a 1 unidad de pKa. Sin embargo cuando se usa el esquema **D**, ya sea con PCM o con SMD, cualquier funcional de la densidad (de los investigados en este trabajo) puede usarse para el cálculo de compuestos fenólicos y obtener un buen acuerdo con los datos experimentales. Cuando se usa el esquema **E1** (con SMD), además de PBE0 y TPSS, también el funcional PBE da buenos resultados para fenoles en general. Lo mismo sucede para el funcional B3LYP cuando se usan el esquema **EN3** y SMD.

Considerando de manera conjunta la información obtenida, puede recomendarse de manera general que para cálculos de pKas de compuestos fenólicos se utilicen los esquemas **D** o **E1**, en combinación con los funcionales PBE0 o TPSS, y SMD para modelar el solvente. Esto cuando los pKas se calculan utilizando métodos teóricos en una reacción química y la expresión de pKa correspondiente que relaciona al pKa con la energía libre de Gibbs de la

reacción. Se espera que siguiendo esta metodología se obtengan desviaciones menores a 1 unidad de pKa con respecto al valor experimental, lo que se encuentra muy por debajo del límite aceptado actualmente.

Por otra parte es importante destacar que para el caso de compuestos fenólicos se cumple la hipótesis planteada, ya que fue posible encontrar metodologías igualmente confiables para especies relativamente sencillas (mono-funcionales) que para moléculas más complejas (poli-funcionales). Además, se encontró que el método de las correlaciones da excelentes resultados para el cálculo de pKas de compuestos fenólicos, con cualquiera de los funcionales de la densidad utilizados en este trabajo. En todos los casos los valores de EAP fueron menores a 0.4 unidades de pKa.

6.1.2. Ácidos Carboxílicos

Al igual que en el caso de fenoles, en los ácidos carboxílicos la especie ácida es neutra y la base conjugada es aniónica, por lo que se calcularon los valores de pKas utilizando los esquemas de reacción y las ecuaciones mostrados en la Tabla 3.

6.1.2.1 Mono-funcionales PCM

Los valores de EAP obtenidos para ácidos carboxílicos mono-funcionales con todos los métodos de cálculo y esquemas de reacción investigados se muestran en la Figura 10. Todos los valores de EAP obtenidos se reportan en el Anexo 9, mientras que en la tabla 14 se muestran todos los esquemas de reacción con los que se encontró al menos un valor de EAP menor a 2 unidades de pKa. La figura 10 muestra claramente, que al igual que en el caso de fenoles la influencia del esquema de reacción sobre los valores de EAP es mucho mayor que la del método de cálculo. Los esquemas de reacción con los que se obtienen los errores más bajos son **A**, **D**, **E2** y **EN3**. Los valores de EAPG que se obtuvieron en estos casos son: **A** (1.48), **D** (1.79), **E2** (1.57) y **EN3** (1.81).

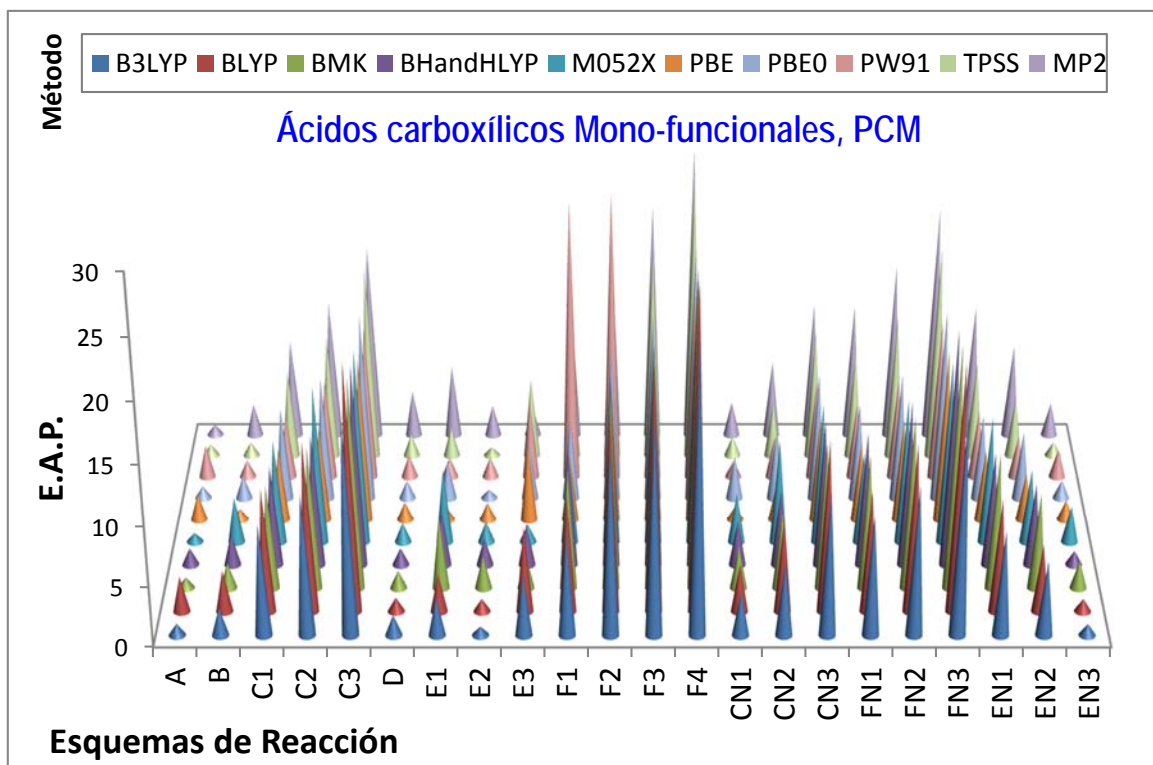


Figura 10. Errores absolutos promedio de los valores de pKa calculados ácidos carboxílicos mono-funcionales usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación.

Tabla 14. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para ácidos carboxílicos mono-funcionales, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BMK	BH&HLYP	M05-2X	PBE	PBE0	PW91	TPSS	MP2	EAPG
A	1.00	2.98	0.89	1.40	0.74	2.23	1.02	2.70	0.96	0.86	1.48
B	2.33	3.46	2.11	2.60	3.85	0.94	1.93	1.42	1.14	2.65	2.24
D	1.73	1.17	1.47	1.41	1.61	1.54	1.52	1.88	1.82	3.77	1.79
E1	3.22	3.19	5.97	4.62	6.13	1.10	3.19	1.65	2.43	5.92	3.74
E2	0.74	1.07	2.74	2.28	1.73	1.48	0.72	1.66	0.70	2.56	1.57
E3	5.20	7.08	2.94	3.47	1.50	6.49	4.63	6.40	6.58	2.39	4.67
CN1	3.68	4.05	3.57	3.79	4.18	1.18	3.63	1.33	1.67	2.85	2.99
EN3	0.97	1.06	2.31	1.16	3.09	1.69	1.53	2.35	1.13	2.76	1.81

Para el esquema **A** se encontraron 7 métodos de cálculo con valores de EAP < 2 y de ellos 4 con EAP < 1 (BMK, M05-2X, TPSS y MP2); obteniéndose el menor EAP (0.74) con M05-2X. El esquema **D** es el único caso en el que se obtienen valores de EAP < 2 para todos los métodos de cálculo utilizados (excepto MP2) pero en ningún caso se obtiene EAP < 1. Es interesante notar que en el caso de fenoles mono-funcionales con PCM se encontró la tendencia opuesta para este esquema (el mejor resultado se obtuvo con MP2). Esto indica la necesidad de buscar estrategias para el cálculo de pKas de diferentes compuestos de manera separada. Para ácidos carboxílicos el mejor resultado con el esquema **D** se obtuvo para BLYP (EAP=1.17).

Para el esquema **E2** se encontraron 7 funcionales con EAP < 2 y 3 con EAP < 1: B3LYP (0.74), PBE0 (0.72) y TPSS (0.70). Por otro lado, para el esquema **EN3** los métodos de cálculo con EAP < 2 son 6 funcionales de la densidad. De ellos solo con B3LYP se obtiene un valor de EAP < 1 (0.97), aunque con BLYP, BHandHLYP y TPSS son cercanos a la unidad (1.06, 1.16 y 1.13, respectivamente).

En total se obtuvieron 39 pares esquema de reacción / método de cálculo con valores de EAP < 2, pero solo 9 con EAP < 1. Los 10 mejores resultados individuales son:

- el esquema **A** en combinación con MP2, B3LYP, BMK, M05-2X y TPSS;
- el esquema **D** con BLYP;
- el esquema **E2** con B3LYP, PBE0 y TPSS; y
- el esquema **EN3** con B3LYP.

6.1.2.2 Mono-funcionales SMD

Para los ácidos carboxílicos mono-funcionales se muestran los valores de EAP obtenidos con todos los métodos de cálculo y esquemas de reacción investigados usando SMD en la Figura 11. Todos los valores de EAP obtenidos de este modo se reportan en el Anexo 10, mientras que en la tabla 15 se muestran todos los esquemas de reacción con los que se encontró al menos un valor de EAP menor a 2 unidades de pKa. Los esquemas de reacción con los que se obtienen los errores más bajos en este caso son **A**, **D**, **E3** y **EN3**. Los valores de EAPG correspondientes son: **A** (1.43), **D** (1.68), **E3** (1.58) y **EN3** (1.99). A diferencia de lo que ocurría para fenoles, para ácidos carboxílicos mono-funcionales no se observa una

disminución significativa de los valores de EAP cuando se usa SMD en lugar de PCM para el cálculo de la energía de solvatación.

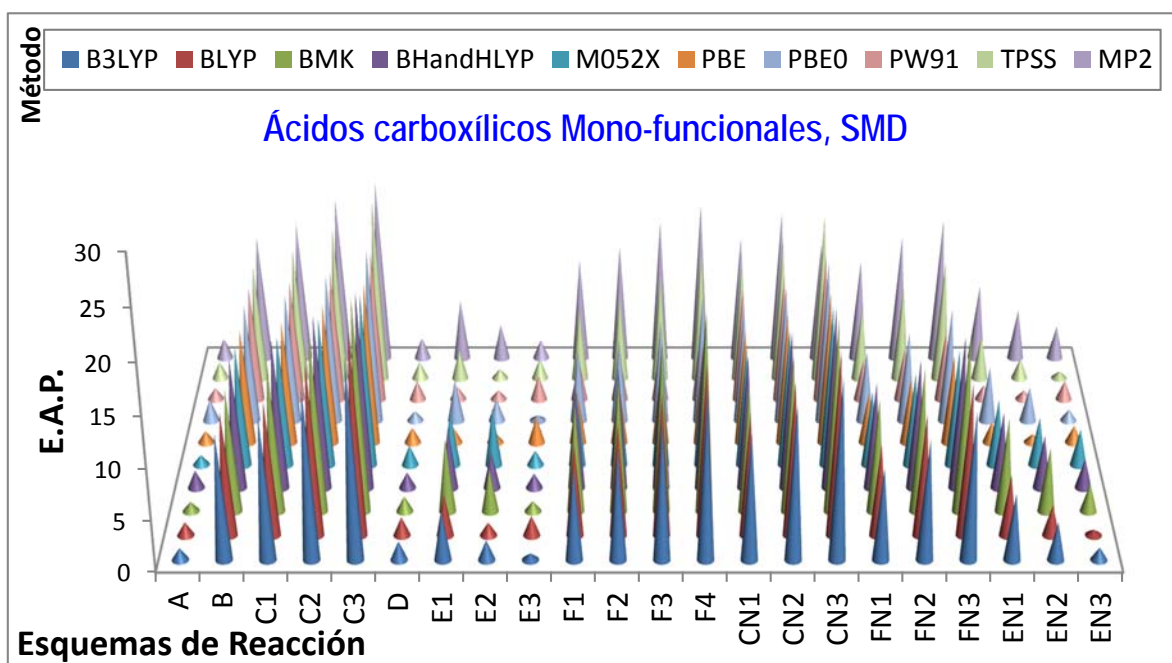


Figura 11. Errores absolutos promedio de los valores de pKa calculados con la estrategia de SMD y todo en solución para todos los ácidos carboxílicos mono-funcionales.

Tabla 15. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para ácidos carboxílicos mono-funcionales, usando SMD en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BMK	BH&HLYP	M052X	PBE	PBE0	PW91	TPSS	MP2	EAPG
A	1.17	1.29	0.93	1.74	0.99	1.16	2.15	1.14	1.68	2.05	1.43
D	1.84	1.80	1.48	1.54	1.97	1.51	0.97	1.87	1.69	2.11	1.68
E1	4.77	3.99	7.14	5.96	5.41	1.48	4.23	1.24	2.89	6.04	4.32
E2	1.96	1.28	4.23	3.65	5.70	0.99	2.17	0.91	0.94	3.55	2.54
E3	0.64	1.93	1.01	1.42	1.44	2.57	0.50	2.40	1.98	1.86	1.58
EN1	6.49	5.95	9.18	7.53	6.55	2.09	5.45	1.45	4.30	7.56	5.66
EN2	3.68	3.00	6.27	5.22	4.77	0.83	3.39	0.89	2.03	5.08	3.52
EN3	1.24	0.55	2.76	2.93	3.59	1.71	1.14	1.92	0.69	3.38	1.99

Para el esquema **A** se encontraron 8 métodos de cálculo con valores de EAP < 2, dos de ellos con EAP < 1 (BMK y M05-2X); los valores de EAP en estos dos casos fueron de 0.93 y 0.99, respectivamente. En el caso del esquema **D** se obtienen valores de EAP < 2 para todos los funcionales de la densidad, pero no para MP2. Con este esquema se obtuvo solo 1 valor de EAP < 1, que corresponde a PBE0 (0.97). Para el esquema **E3** se encontraron 8 métodos de cálculo con EAP < 2, de ellos solo dos con EAP < 1: B3LYP (0.64) y PBE0 (0.50). Para el esquema **EN3** los métodos de cálculo con EAP < 2 son 6 funcionales de la densidad. De ellos BLYP y TPSS producen valores de EAP < 1 (0.55 y 0.69, respectivamente).

En total se obtuvieron 39 pares esquema de reacción / método de cálculo con valores de EAP < 2, y 12 de ellos con EAP < 1. Los 10 mejores resultados individuales son:

- el esquema **A** en combinación con BMK;
- el esquema **D** con PBE0;
- el esquema **E2** con PW91 y TPSS;
- el esquema **E3** con B3LYP y PBE0; y
- el esquema **EN2** con PBE y PW91; y
- el esquema **EN3** con BLYP y TPSS.

Es interesante que al comparar los resultados de ácidos carboxílicos con los de fenoles hay una marcada diferencia con respecto al esquema **D**. Mientras que para fenoles con este esquema se obtenían valores de EAP < 1 con varios métodos de cálculo, para ácidos carboxílicos esto pasa solamente con un funcional de la densidad (BLYP con PCM y PBE0 con SMD).

6.1.2.3 Poli-funcionales PCM

Los valores de EAP obtenidos en este caso se muestran de manera resumida en la Figura 12, mientras que los valores individuales de EAP se reportan en el Anexo 11 y en la tabla 16 se reportan todos los esquemas de reacción con los que se encontró al menos un valor de EAP menor a 2 unidades de pKa. Los esquemas de reacción con los que se obtienen los errores más bajos en este caso son **D** y **EN2**. En estos dos casos se obtuvieron valores de EAPG iguales a 1.29 y 1.70, respectivamente. En general los valores de EAP obtenidos en el caso

de ácidos carboxílicos poli-funcionales son mayores que los de ácidos carboxílicos mono-funcionales, cuando se obtienen las energías de solvatación usando PCM y ciclos termodinámicos.

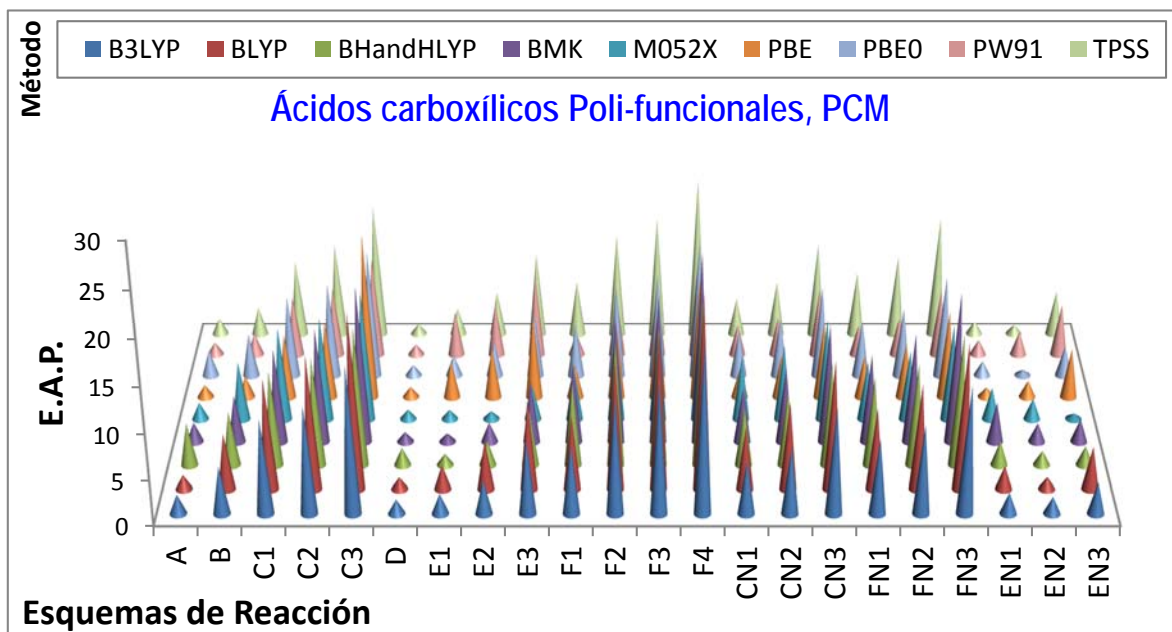


Figura 12. Errores absolutos promedio de los valores de pKa calculados para ácidos carboxílicos poli-funcionales usando PCM y ciclos termodinámicos.

Tabla 16. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para ácidos carboxílicos poli-funcionales, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BH&HLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	EAPG
A	2.11	1.54	4.62	2.17	1.86	1.36	2.91	1.37	1.87	2.20
D	1.58	1.29	1.91	1.10	1.07	1.15	1.22	1.13	1.12	1.29
E1	2.20	2.63	1.23	0.94	1.06	4.08	2.34	5.00	3.05	2.50
E2	3.67	5.31	2.71	2.17	0.88	5.59	3.79	5.93	4.96	3.89
EN1	2.32	2.53	2.65	4.37	3.55	1.09	1.64	1.68	1.46	2.37
EN2	1.87	1.33	1.57	2.07	2.32	1.87	0.65	2.62	1.04	1.70
EN3	3.44	4.64	2.10	2.57	0.61	5.50	3.38	5.86	5.03	3.68

El esquema **D** es el único caso en el que se obtienen valores de EAP < 2 para todos los métodos de cálculo utilizados, pero en ningún caso se obtiene EAP < 1. Los mejores resultados con este esquema EAP = 1.10 se obtuvieron con BMK, M05-2X, PBE, PW91 y TPSS. Como se ve, en este caso varios funcionales de la densidad conducen a valores de EAP muy cercanos entre sí. En el caso del esquema **EN2** el mejor resultado se obtuvo con PBE0 (EAP = 0.65) que es el único caso en el que se obtuvo EAP < 1 para este esquema. Sin embargo se encontraron 6 funcionales con EAP < 1

Además para el tipo de sistemas analizados en esta sección ocurre que de manera puntual (o sea con algún método de estructura electrónica específico) se obtienen muy buenos resultados con esquemas que en el resto de los casos tienen valores altos de EAP. Los dos casos más dramáticos en este sentido son los pares **EN3** / M05-2X (EAP = 0.61) y **E2** / M05-2X (EAP = 0.88). De hecho el EAP obtenido con **EN3** / M05-2X es el menor de todos, mientras que con los otros valores de EAP que se obtuvieron usando este esquema van de 2.10 a 5.86 unidades de pKa.

En total se obtuvieron 29 pares esquema de reacción / método de cálculo con valores de EAP < 2, pero solo 4 de ellos con EAP < 1. Los 10 mejores resultados individuales son:

- el esquema **D** con BMK, M05-2X, PBE, PW91 y TPSS;
- el esquema **E1** con BMK;
- el esquema **E2** con M05-2X;
- el esquema **EN2** con PBE0 y TPSS; y
- el esquema **EN3** con M05-2X.

Si bien ninguno de estas combinaciones coincide con los 10 mejores resultados para ácidos carboxílicos mono-funcionales, sí fue posible encontrar 2 pares esquema de reacción / método de cálculo con los cuales se obtienen valores de pKa calculados en acuerdo razonable con el experimento (EAP ≤ 1.5 unidades de pKa). Ambos se corresponden al esquema **D**, en un caso en combinación con BLYP y en el otro con PBE0. De manera que puede decirse que la hipótesis se cumple en este caso pero en menor medida que para fenoles.

6.1.2.4. Poli-funcionales SMD

La Figura 13 muestra de manera resumida los resultados para ácidos carboxílicos poli-funcionales usando SMD. Todos los valores de EAP obtenidos de este modo se reportan en

el Anexo 12, mientras que en la tabla 17 se muestran todos los esquemas de reacción con los que se encontró al menos un valor de EAP menor a 2 unidades de pKa. Los esquemas de reacción con los que se obtienen los errores más bajos en este caso son **A**, **D**, **E2** y **EN3**. Los valores de EAPG correspondientes son: **A** (1.79), **D** (0.68), **E2** (1.48) y **EN3** (1.95). A diferencia de lo que ocurría para ácidos carboxílicos mono-funcionales, en este caso sí se observa una disminución significativa de los valores de EAP cuando se usa SMD en lugar de PCM para el cálculo de la energía de solvatación.

Para el esquema **A** se encontraron 5 métodos de cálculo con valores de EAP < 2, pero ninguno de ellos con EAP < 1. Los valores de EAP correspondientes a estos 5 métodos (BLYP, BMK, M05-2X, PBE y PW91) son muy similares entre sí (1.10). En el caso del esquema **D**, para todos los funcionales de la densidad probados se obtienen valores de EAP que son no solo menores a 2 unidades de pKas, sino también a 1 unidad de pKa. Un comportamiento similar se encontró con este esquema (también usando SMD) para el caso de los compuestos poli-funcionales fenólicos. Los valores de EAP más bajos se obtuvieron para el esquema **D** con B3LYP (0.53), BLYP (0.60), y PW91 (0.58).

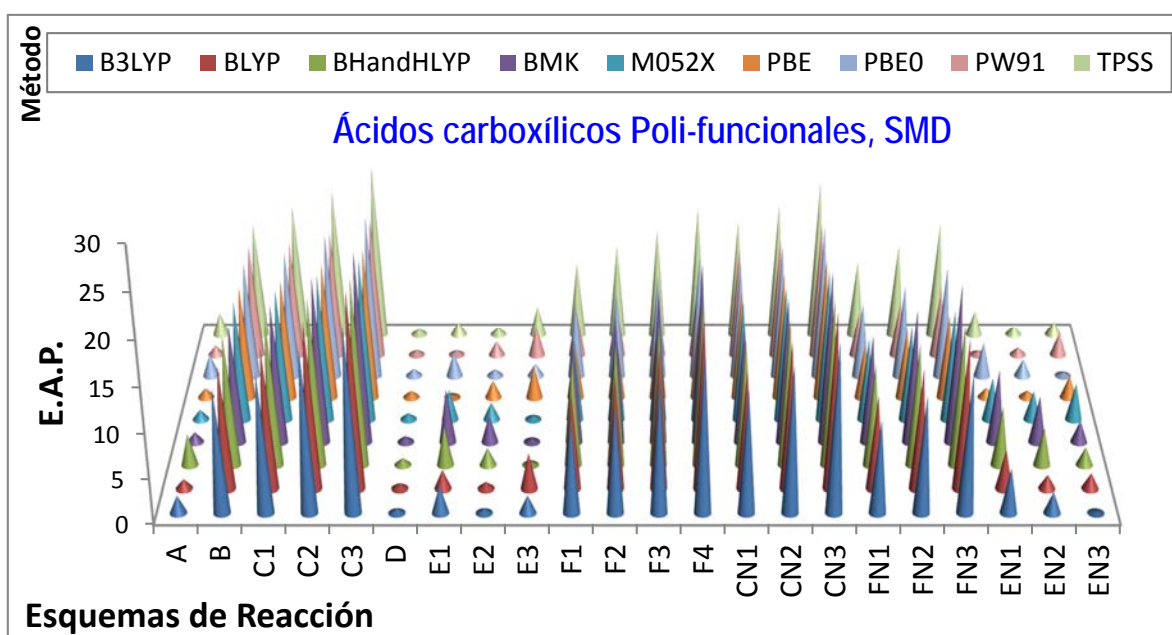


Figura 13. Errores absolutos promedio de los valores de pKa calculados usando SMD para ácidos carboxílicos poli-funcionales.

Tabla 17. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para ácidos carboxílicos poli-funcionales, usando SMD en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BH&HLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	EAPG
A	2.00	1.07	3.54	1.13	1.11	1.06	2.48	1.09	2.64	1.79
D	0.53	0.60	0.80	0.68	0.72	0.76	0.81	0.58	0.67	0.68
E1	3.11	2.14	4.35	5.99	3.44	0.26	2.67	0.55	1.49	2.67
E2	0.53	0.74	1.96	3.04	1.94	1.95	0.77	1.60	0.82	1.48
E3	2.03	4.03	0.39	0.43	0.44	3.62	1.40	3.30	3.34	2.11
EN1	5.00	4.43	6.65	8.19	4.77	1.12	3.88	0.46	2.84	4.15
EN2	2.41	1.55	4.26	5.23	3.27	0.79	1.94	0.89	0.98	2.37
EN3	0.21	1.73	2.04	2.42	4.14	2.45	0.40	2.58	1.54	1.95

Con el esquema **E2** se obtienen valores de EAP < 2 con todos los métodos de cálculo probados (excepto BMK) y EAP < 1 con 4 de ellos (B3LYP, BLYP, PBE0 y TPSS). Los mejores resultados para este esquema se obtuvieron con B3LYP (0.53), PBE0 (0.77) y TPSS (0.82). Para el esquema **EN3** los métodos de cálculo con EAP < 2 son 4 y los mejores resultados se obtuvieron con B3LYP (0.21) y PBE0 (0.40), que son además los únicos con EAP < 1.

En total se obtuvieron 40 pares esquema de reacción / método de cálculo con valores de EAP < 2, y 23 de ellos con EAP < 1. Los 10 mejores resultados individuales son:

- el esquema **D** con B3LYP y PW91;
- el esquema **E1** con PBE y PW91;
- el esquema **E2** con B3LYP;
- el esquema **E3** con BH&HLYP, BMK y M05-2X;
- el esquema **EN1** con PW91; y
- el esquema **EN3** con B3LYP y PBE0.

De manera análoga a lo que se encontró al usar PCM, en este caso tampoco ninguno los 10 mejores resultados coincide para ácidos carboxílicos poli-funcionales y mono-funcionales. Sin embargo, se encontraron numerosos casos para los cuales se obtienen valores de pKa calculados en acuerdo razonable con el experimento (en este caso con $EAP \leq 1.2$ unidades

de pKa). Ellos son: el esquema **A** con BMK, PBE y PW91; el esquema **D** con PBE0; el esquema **E2** con TPSS; el esquema **EN2** con PBE y PW91; y el esquema **EN3** con PBE0. De manera que puede decirse que en este caso se cumple claramente la hipótesis de trabajo. También puede decirse que en el caso de ácidos carboxílicos ésta se cumple de manera más evidente cuando la energía de solvatación se calcula con SMD que cuando se calcula con PCM.

6.1.2.5. Conjunto completo de ácidos carboxílicos con PCM

Al igual que se hizo para el caso de fenoles, en esta sección se investigan todos los ácidos carboxílicos estudiados como un conjunto único para poder hacer recomendaciones en cuanto a las mejores estrategias de cálculo para estos compuestos en general. Los valores de EAP obtenidos se muestran de manera resumida en la Figura 14, los valores se reportan en el Anexo 13 y la Tabla 18 contiene los datos para todos los esquemas de reacción con al menos un valor de EAP < 2. Se encontró que los esquemas que dan mejores resultados en este caso (**A**, **D** y **E2**). Los valores de EAPG correspondientes son 1.70, 1.48 y 2.19.

En el caso de ácidos carboxílicos, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación no se encontró ninguna combinación esquema de reacción / método de cálculo con EAP < 1, para el conjunto completo de moléculas estudiadas. Para el esquema **A** se encontraron 6 esquemas de reacción con valores de EAP menores a 2 unidades de pKa. Para este esquema los valores más bajos de EAP fueron obtenidos con M05-2X (1.03), BMK (1.24) y TPSS (1.26). En el caso del esquema **D** para todos los funcionales de la densidad probados se obtuvieron valores de EAP < 2. Los mejores resultados corresponden a los cálculos realizados con BMK (1.28), PBE (1.41) y PBE0 (1.40). Para el esquema **E2** se obtuvieron valores de EAP < 2 con solo tres de los métodos empleados. Estos son B3LYP (EAP=1.61), M05-2X (EAP=1.39) y PBE0 (EAP=1.62).

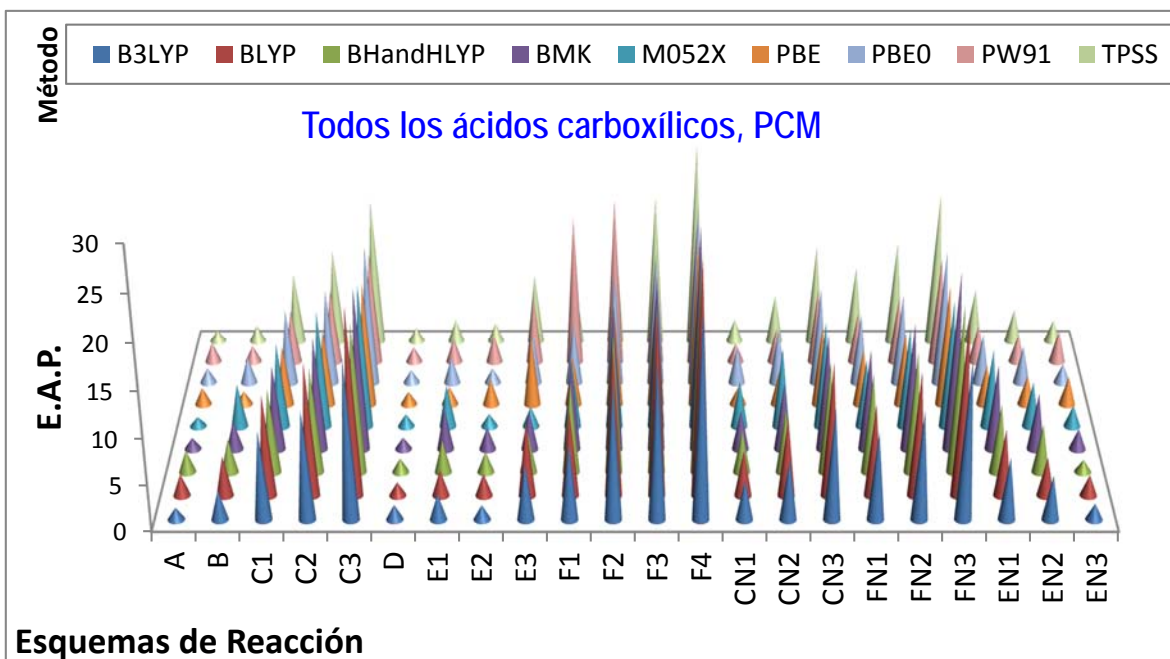


Figure 14. Errores absolutos promedio de los valores de pKa calculados usando PCM para el estudio del conjunto completo de ácidos carboxílicos.

Tabla 18. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para el conjunto completo de ácidos carboxílicos poli-funcionales, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BH&HLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	EAPG
A	1.32	2.40	2.44	1.24	1.03	1.83	1.64	2.14	1.26	1.70
B	3.25	4.37	3.67	3.08	4.67	1.45	2.87	1.78	1.85	3.00
D	1.65	1.48	1.53	1.28	1.36	1.41	1.40	1.64	1.58	1.48
E1	2.81	2.84	3.53	4.45	4.60	1.95	2.81	2.63	2.46	3.12
E2	1.61	2.44	2.26	2.43	1.39	2.81	1.62	3.04	2.07	2.19
CN1	4.34	5.02	4.66	4.35	5.11	1.94	4.13	2.09	2.56	3.80

Para el conjunto completo de ácidos carboxílicos, usando PCM y ciclos termodinámicos para el cálculo de la energía de solvatación se encontraron 22 pares esquema de reacción / método de cálculo con valores de EAP < 2. Los 10 mejores resultados se obtienen con:

- el esquema **A** con los funcionales B3LYP, BMK, M05-2X y TPSS;
- el esquema **B** en combinación con PBE;
- el esquema **D** con los funcionales BMK, M05-2X, PBE y PBE0; y
- el esquema **E2** con el funcional M05-2X.

6.1.2.6. Conjunto de ácidos carboxílicos con SMD

En este caso los valores de EAP obtenidos se muestran de manera resumida en la Figura 15, los valores se reportan en el Anexo 14 y los datos para todos los esquemas de reacción con al menos un valor de $EAP < 2$ se reportan en la tabla 19. En general los valores obtenidos con SMD se encuentran en mejor acuerdo con los valores experimentales que los obtenidos con PCM. Este comportamiento es similar al que se encontró para fenoles y se refleja tanto en el número de combinaciones esquema de reacción / método de cálculo con $EAP < 2$ (y menores a 1) como en los valores específicos de EAP para combinación son más bajos.

En el caso de ácidos carboxílicos, usando SMD en el cálculo de las energías de solvatación sí se encontraron combinaciones esquema de reacción / método de cálculo con $EAP < 1$. Los esquemas que dan mejores resultados en este caso son **A**, **D**, **E3** y **EN3**. Con estos esquemas se obtuvieron valores de EAPG iguales a 1.52, 1.35, 1.86 y 1.91, respectivamente.

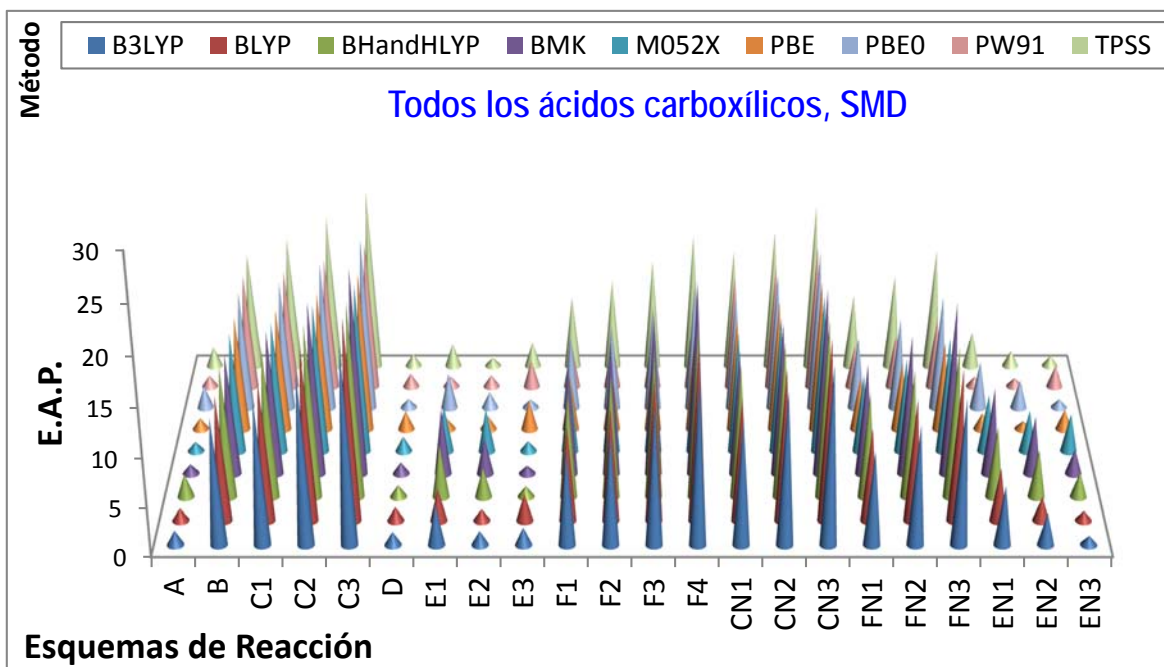


Figure 15. Errores absolutos promedio de los valores de pKa calculados con SMD para el estudio del conjunto completo de ácidos carboxílicos.

Para el esquema **A** se encontraron 6 métodos de cálculo con valores de EAP menores a 2 unidades de pKa. Para este esquema los valores más bajos de EAP fueron obtenidos con BMK y M05-2X (0.98 en ambos casos), que son los dos únicos casos en los que los valores de EAP son además menores a la unidad de pKa, para este esquema. En el caso del esquema **D** se obtuvieron valores de EAP < 2 para todos los funcionales de la densidad probados, pero solo con PBE0 el valor de EAP es menor a 1 (0.92). Para el esquema **E3** se obtuvieron valores de EAP < 2 con 5 de los métodos empleados y EAP < 1 con dos de ellos. Estos son BMK (EAP=0.84) y PBE0 (EAP=0.85). En el caso del esquema **EN3** se obtuvieron 4 valores de EAP < 2, tres de ellos son además menores a la unidad de pKa. Los mejores resultados con este esquema se corresponden a los cálculos realizados con B3LYP (0.93) y PBE0 (0.87).

Tabla 19. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para el conjunto completo de ácidos carboxílicos poli-funcionales, usando SMD en el cálculo de las energías de solvatación, y valores de EAPG.

	B3LYP	BLYP	BH&HLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	EAPG
A	1.47	1.27	2.35	0.98	0.98	1.15	2.28	1.17	2.04	1.52
D	1.34	1.39	1.22	1.18	1.54	1.90	0.92	1.39	1.29	1.35
E1	4.14	3.27	5.42	6.75	4.75	1.06	3.65	1.05	2.38	3.61
E2	1.44	1.17	3.02	3.83	4.44	1.35	1.67	1.21	0.88	2.11
E3	1.80	2.74	1.12	0.84	1.07	3.02	0.85	2.81	2.51	1.86
EN1	5.96	5.37	7.24	8.87	5.93	1.71	4.88	1.08	3.78	4.98
EN2	3.25	2.43	4.84	5.94	4.28	0.84	2.89	0.92	1.65	3.00
EN3	0.93	0.99	2.53	2.58	3.93	2.05	0.87	2.22	1.06	1.91

Para el conjunto completo de ácidos carboxílicos, usando SMD para el cálculo de la energía de solvatación se encontraron 37 pares esquema de reacción / método de cálculo con valores de EAP < 2, de ellos 11 con EAP < 1. Los 10 mejores resultados se obtienen con:

- el esquema **A** con los funcionales BMK y M05-2X;
- el esquema **D** en combinación con PBE0;
- el esquema **E2** con el funcional TPSS;
- el esquema **E3** con los funcionales BMK y PBE0;
- el esquema **EN2** con los funcionales PBE y PW91; y
- el esquema **EN3** con los funcionales B3LYP y PBE0.

6.1.2.7. Método de correlaciones para ácidos carboxílicos

Los resultados obtenidos con el método de las correlaciones para el cálculo de pKas de compuestos fenólicos se muestran en la tabla 20. Con todos los métodos de cálculo probados se encontró que los coeficientes de correlación (R^2) son mayores que, o cercanos a, 0.8. De manera general los valores de R^2 obtenidos en este caso son más bajos que para fenoles. A pesar de esto, el acuerdo con los valores experimentales es muy bueno, con todos los valores de EAP siendo menores a 0.45 unidades de pKa.

Tabla 20. Parámetros m y C_0 , coeficiente de correlación lineal y errores absoluto promedio para el cálculo de pKas de ácidos carboxílicos usando el método de las correlaciones.

	m	C_0	R^2	EAP
B3LYP	0.231	-59.831	0.840	0.32
BLYP	0.223	-57.501	0.864	0.26
BH&HLYP	0.260	-68.336	0.870	0.41
BMK	0.273	-71.478	0.841	0.29
M05-2X	0.257	-66.694	0.762	0.31
PBE	0.227	-58.702	0.792	0.23
PBE0	0.240	-62.765	0.788	0.26
PW91	0.223	-57.488	0.768	0.28
TPSS	0.224	-58.266	0.792	0.29

El valor más bajo de EAP se obtuvo con PBE (0.23) y el más alto con BHandHLYP (0.41). Además los valores obtenidos con los funcionales PBE0 y TPSS, que son los que más frecuentemente dan buenos resultados con la estrategia de los esquemas de reacción, también son muy buenos (menores a 0.3 unidades de pKa). Al igual que ocurría para fenoles, estos valores son menores que los obtenidos con cualquiera de los esquemas de reacción. Esto soporta el uso del método de correlaciones para predecir, de manera confiable, valores de pKas de compuestos fenólicos para los que no se tiene esta información.

6.1.2.8 Generalidades para ácidos carboxílicos

Al igual que en el caso de fenoles, para ácidos carboxílicos el aspecto que tiene mayor influencia a la hora de obtener buena correspondencia entre los valores de pKa calculados y los datos experimentales es el esquema de reacción. Mientras que el segundo aspecto en importancia sigue siendo el modelo continuo de solvente, obteniéndose mejores resultados con SMD que con PCM para el cálculo de pKas. En tercer lugar se encuentra el método de estructura electrónica.

En el caso de ácidos carboxílicos cuando se usa PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, solo hay un esquema que da resultados en buen acuerdo con los experimentos tanto para moléculas mono-funcionales como para moléculas polifuncionales (el esquema *D*). Por el contrario, en el caso de SMD hay 3 esquemas de reacción que dan buenos resultados, comparados con los valores experimentales, sin importar el tipo de ácidos carboxílicos investigados (mono-funcionales, poli-funcionales, o conjunto completo). Estos son los esquemas *A*, *D* y *EN3*. Según estos resultados solamente el esquema *D* puede aplicarse de manera general para el cálculo de pKas tanto de compuestos fenólicos como de ácidos carboxílicos, si no se considera escoger SMD para modelar el solvente.

Cuando se usa el esquema *D*, con PCM (y se consideran tanto especies mono-funcionales, como poli-funcionales, o incluso el conjunto completo de ácidos carboxílicos) se obtienen buenos acuerdos con los datos experimentales con 5 funcionales de la densidad. Estos son BLYP, BMK, PBE, PBE0 y TPSS. Con este mismo esquema y SMD solo el funcional PBE0 da valores de EAP significativamente menores a las 2 unidades de pKa. Sin embargo en este caso los valores de EAP son además menores a 1. Cuando se usa el esquema *EN3* (con SMD), además de PBE0 también los funcionales B3LYP y TPSS dan buenos resultados para ácidos carboxílicos en general. Por el contrario, los buenos resultados obtenidos con BMK, M05-2X y PW91 para los pKas de ácidos carboxílicos usando el esquema *A* (con SMD) parecen ser una particularidad y no una generalidad.

Considerando de manera conjunta la información obtenida, puede recomendarse de manera general que para cálculos de pKas de ácidos carboxílicos se utilice el esquema *D*, en combinación con el funcional de la densidad PBE0, ya sea que se modele el solvente con SMD o con PCM. Esto cuando los pKas se calculan utilizando una reacción química y la expresión de pKa correspondiente que relaciona al pKa con la energía libre de Gibbs de la reacción. Se espera que siguiendo esta metodología se obtengan desviaciones menores a 1 unidad de pKa con respecto al valor experimental, lo que se encuentra muy por debajo del límite aceptado actualmente.

También es importante destacar que, al igual que en el caso de compuestos fenólicos, para los ácidos carboxílicos se cumple la hipótesis planteada, ya que fue posible encontrar metodologías igualmente confiables para especies relativamente sencillas (mono-

funcionales) que para moléculas más complejas (poli-funcionales). Además, se encontró que el método de las correlaciones da excelentes resultados para el cálculo de pKas de ácidos carboxílicos, con cualquiera de los funcionales de la densidad utilizados en este trabajo. En este caso todos los valores de EAP fueron menores a 0.45 unidades de pKa.

6.1.3 Aminas

En el contexto de los cálculos de pKas, este tipo de moléculas es diferente a las dos familias anteriormente analizadas por dos razones principales. La primera es que en el proceso de desprotonación de las aminas el ácido es una especie catiónica y la base conjugada es la especie neutra. Por lo tanto, los valores de pKas para las aminas se obtendrán teniendo como base los esquemas de reacción y ecuaciones mostrados en la Tabla 4. En segundo lugar, la naturaleza química de ácidos carboxílicos y fenoles en el sitio de desprotonación es idéntica para todos los miembros de la familia. Por el contrario, las aminas se dividen en primarias, secundarias y terciarias, dependiendo del número de sustituyentes que tenga en N protonado. Esto puede hacer aún más complicado el encontrar metodologías únicas que permitan obtener valores calculados de pKas en buen acuerdo con los datos experimentales. Por lo tanto, además de realizar análisis similares a los hechos para fenoles y ácidos carboxílicos, para aminas se harán, cuando sea necesario, análisis adicionales incluyendo la posible influencia que tiene la naturaleza de los sitios ácidos (aminas primarias, secundarias y terciarias) en el cálculo de pKas.

También vale la pena destacar que en los resultados reportados no se incluyen los esquemas *E* y *EN*, ya que en el caso de aminas (a diferencia de fenoles y ácidos carboxílicos) con estos esquemas se obtienen las mayores desviaciones con respecto a los datos experimentales. Esto es lógico ya que mientras para las otras dos familias de compuestos estudiados con los esquemas *E* y *EN* se igualaba el número y tipo de especies cargadas a ambos lados del equilibrio ácido base, en el caso de amina sucede lo contrario (estos son los esquemas que concentran a todas las especies iónicas del mismo lado del equilibrio).

Este punto se demuestra en la figura 16. Los resultados en esta figura se muestran solamente con el funcional BHandHLYP ya que las tendencias obtenidas son las mismas para todos los

métodos de cálculo. En ella se puede apreciar que sin importar el tipo de amina estudiado (amonio, aminas primarias, secundarias y terciarias) los valores de EAP que se obtienen con los esquemas *E* y *EN* son mayores a 8 unidades de pKa. Esto confirma que estos esquemas no son adecuados para calcular valores de pKas de aminas, por lo que no se han incluido en la discusión de resultados de ninguna de las secciones 6.1.3.

Además como el interés principal de esta tesis es poder hacer generalizaciones para el cálculo de pKas, y el caso de aminas es particularmente complicado, para esta familia de compuestos se redujo el número de cálculos para compuestos poli-funcionales. La estrategia empleada para ello se describe a continuación. Los esquemas de reacción a incluir en esta parte se escogieron en base a dos criterios. El primero fue incluir los esquemas más empleados en la literatura (*A*, *B* y *D*). El segundo fue escoger esquemas de reacción, pero de los propuestos en este trabajo, que dieran buenos resultados para aminas mono-funcionales. En el caso de cálculos PCM estos esquemas son *CI*, *CNI* y *F2*, mientras que en el caso de SMD fueron *F1*, *F2*, *F3*, *FN1*, *FN2* y *FN3*. También en base a los valores de EAP obtenidos para aminas mono-funcionales, los métodos de estructura electrónica empleados para el cálculo de pKa de aminas poli-funcionales fueron: BHandHLYP, PBE0 y M05-2X (con PCM) y PBE, PBE0 y PW91 (con SMD).

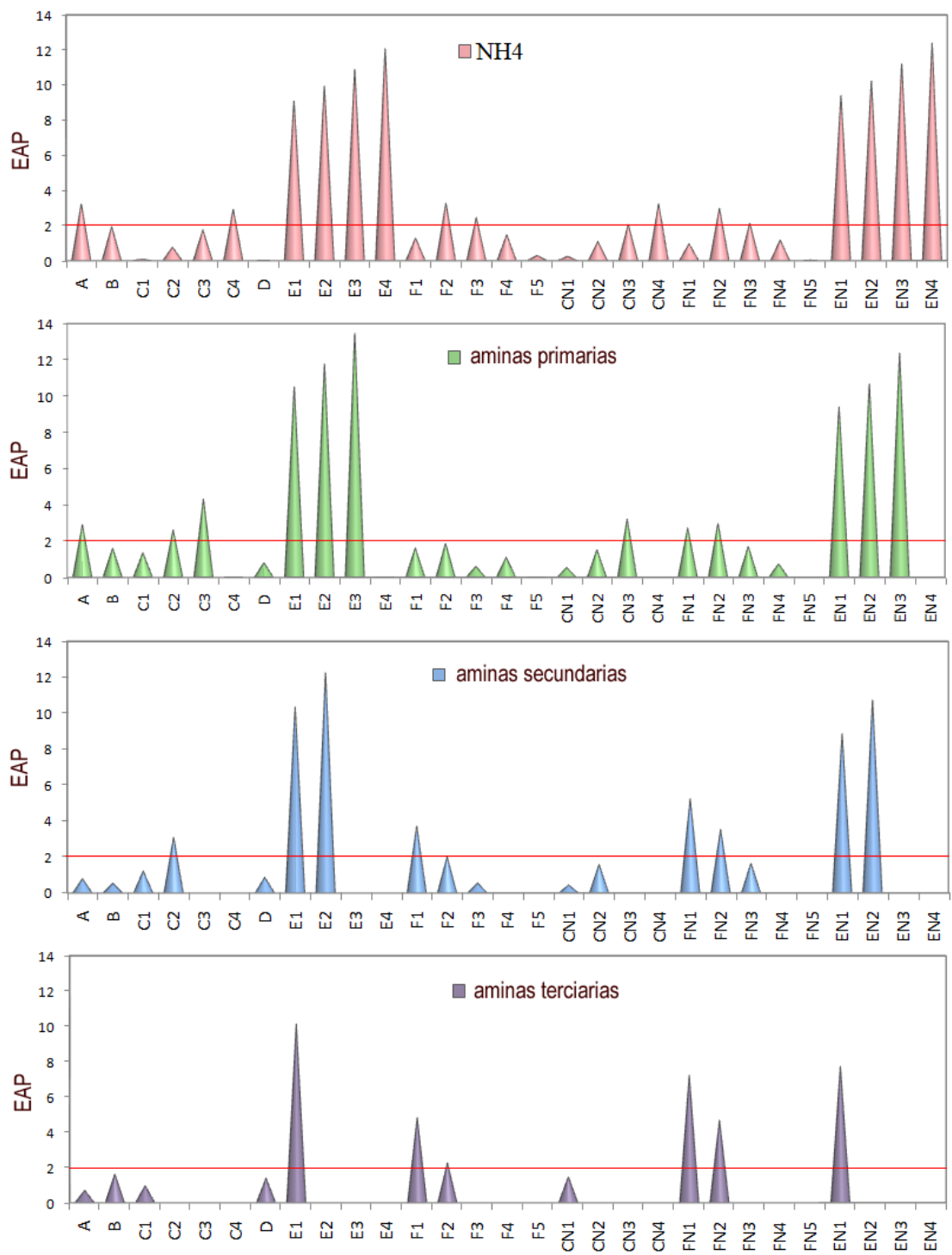


Figura 16. Errores absolutos promedio para los pKa calculados con PCM y BHandHLYP para aminas mono-funcionales.

6.1.3.1 Mono-funcionales PCM

En este caso los valores de EAP obtenidos se muestran de manera resumida en la Figura 17, los valores se reportan en el Anexo 15 y los datos para todos los esquemas de reacción con al menos un valor de $EAP < 2$ se reportan en la tabla 21. La figura 17 muestra que en general, comparando con fenoles y ácidos carboxílicos, los EAP obtenidos en este caso son relativamente bajos (cuando no se consideran los esquemas *E* y *EN*). De hecho se encontraron 9 esquemas de reacción con EAPG por debajo de 2 unidades de pKa (*B*, *CI*, *D*, *F2*, *F3*, *CNI*, *CN2*, *FN3* y *FN4*). De ellos 7 esquemas (*B*, *CI*, *D*, *F2*, *F3*, *FN3*, y *CNI*) tienen valores de EAPG < 1.6 . Los valores de EAPG en estos casos son 1.46, 1.20, 1.02, 1.54, 1.41, 1.20 y 1.55, respectivamente.

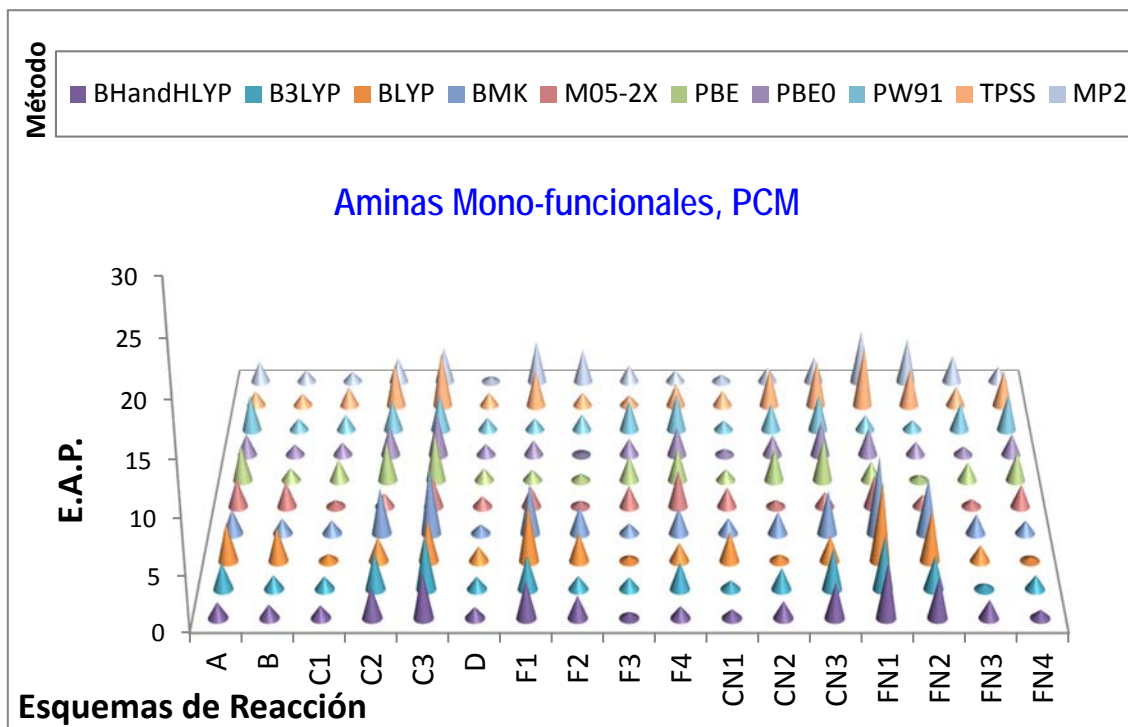


Figura 17. Errores absolutos promedio de los valores de pKa calculados con PCM y ciclos termodinámicos para aminas mono-funcionales.

Para los esquemas *D* y *CI* los valores de EAP obtenidos con todos los métodos de cálculo probados son menores a 2 unidades de pKa y en ambos casos 3 de ellos cumplen con $EAP < 1$. En el caso del esquema *D* son los métodos BMK, M05-2X y MP2, mientras que para el esquema *CI* son BLYP, M05-2X y MP2. Sin embargo es importante destacar que para el

esquema **D** con todos los métodos (excepto BLYP) se obtiene valores de EAP < 1.3, lo cual representa un buen acuerdo con los datos experimentales. En el caso del esquema **CI** este valor es sobrepasado por los funcionales PBE, PW91 y TPSS.

Tabla 21. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para aminas mono-funcionales, usando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, y valores de EAPG.

	BH&HLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2	EAPG
A	1.45	2.41	3.43	2.21	2.53	3.29	1.94	3.23	1.36	2.03	2.39
B	1.24	1.28	3.01	1.39	2.25	1.15	0.97	1.07	1.12	1.07	1.46
C1	1.17	1.19	0.65	1.20	0.61	1.94	1.15	1.41	1.74	0.92	1.20
C2	2.84	3.33	2.18	4.12	1.24	3.57	2.68	2.68	3.92	2.37	2.89
D	1.00	1.14	1.33	0.83	0.96	1.14	1.01	1.14	1.22	0.42	1.02
F1	3.38	3.04	5.01	4.46	1.84	1.00	1.36	1.10	3.29	3.88	2.84
F2	2.03	1.21	2.57	2.45	0.54	0.63	0.18	1.47	1.23	3.12	1.54
F3	0.57	1.09	0.46	0.93	1.91	2.25	1.51	2.73	0.95	1.65	1.41
F4	1.10	2.45	1.66	2.32	3.27	2.97	2.55	3.36	2.05	1.07	2.28
CN1	0.81	0.80	2.68	1.42	1.78	1.05	0.56	0.88	1.38	0.64	1.20
CN2	1.53	1.93	0.60	1.96	0.59	3.01	1.85	2.45	3.47	1.43	1.88
CN3	3.19	3.68	2.27	3.90	1.36	3.95	3.10	3.28	4.25	2.44	3.14
FN1	5.05	4.85	7.12	6.99	2.90	1.80	2.52	1.39	5.64	4.90	4.32
FN2	3.71	3.03	4.68	4.99	1.13	0.48	1.23	0.93	3.59	4.14	2.79
FN3	1.66	0.52	1.44	1.78	0.70	1.69	0.79	2.50	1.79	2.59	1.55
FN4	0.74	1.28	0.37	1.18	2.02	2.63	1.93	3.33	3.26	1.58	1.83

En el caso del esquema **B** se encontraron 8 métodos de cálculo con valores de EAP < 2, pero de ellos solo uno con EAP < 1 (PBE0). Sin embargo con los métodos MP2, PW91, TPSS y PBE se obtienen valores de EAP ~1. Para el esquema **F3** también se encontraron 8 métodos de cálculo con valores de EAP < 2 y de ellos 4 con EAP < 1 (BHandHLYP, BLYP, BMK y TPSS), siendo los mejores resultados para este esquema los obtenidos con BHandHLYP (0.57) y BLYP (0.46). En el caso del esquema **F2** se encontraron 6 métodos de cálculo con

valores de $EAP < 2$ y de ellos 3 con $EAP < 1$ (M05-2X, PBE y PBE0). Los mejores resultados para este esquema se obtuvieron con M05-2X (0.54) y PBE0 (0.18).

Para el esquema **CNI** con todos los métodos de cálculo (excepto BLYP) se obtienen valores de $EAP < 2$. De ellos 5 tienen $EAP < 1$ (BHandHLYP, B3LYP, PBE0, PW91 y MP2). En este caso los valores más bajos de EAP se obtuvieron con PBE0 (0.56) y MP2 (0.64). En el caso del esquema **FN3** se encontraron 7 métodos de cálculo con valores de $EAP < 2$ y dos de ellos con $EAP < 1$, B3LYP (0.52) y M05-2X (0.70).

De manera resumida, los resultados más relevantes en el caso de aminas mono-funcionales, usando PCM y ciclos termodinámicos para el cálculo de la energía de solvatación, son que se encontraron 89 pares esquema de reacción / método de cálculo con valores de $EAP < 2$, de ellos 25 con $EAP < 1$ (y varios con $EAP < 0.5$). Los 10 mejores resultados se obtienen con:

- el esquema **D** con el método MP2;
- el esquema **F2** con los funcionales M05-2X y PBE0;
- el esquema **F3** con los funcionales BHandHLYP y BLYP;
- el esquema **CNI** con el funcional PBE0;
- el esquema **CN2** con el funcional M05-2X;
- el esquema **FN2** con el funcional PBE;
- el esquema **FN3** con el funcional B3LYP; y
- el esquema **FN4** con el funcional BLYP.

Sin embargo es importante destacar que no todos los esquemas pueden utilizarse para todas las aminas estudiadas. Por ejemplo, el esquema **FN4** solo puede usarse en el caso del amonio, el esquema **FN3** para amonio y aminas terciarias; y el esquema **FN2** para amonio, aminas terciarias y secundarias. De hecho los únicos esquemas de la tabla 4 que pueden usarse para todas las aminas mono-funcionales reportadas en la tabla 1 son los esquemas **A**, **B**, **CI**, **CNI**, **D**, **F1**, **F2**, **FN1** y **FN2** (esto excluyendo a los esquemas **E1** y **EN1** por las razones expuestas anteriormente).

Para este sub-conjunto de esquemas también hay muchos casos en acuerdo razonablemente bueno con los experimentos. Si tomamos por ejemplo todos los casos en los que $EAP < 1.3$, estos serían:

- esquema **B** con BHandHLYP, B3LYP, PBE, PBE0, PW91, TPSS y MP2
- esquema **CI** con BHandHLYP, B3LYP, BLYP, BMK, M05-2X, PBE0 y MP2
- esquema **D** con todos los métodos (excepto BLYP)
- esquema **FI** con PBE y PW91
- esquema **F2** con B3LYP, M05-2X, PBE, PBE0 y TPSS
- esquema **CNI** con BHandHLYP, B3LYP, PBE, PBE0, PW91 y MP2
- esquema **FN2** con M05-2X, PBE, PBE0 y PW91

De manera que cualquiera de estas combinaciones podría usarse exitosamente para calcular pKas de aminas mono-funcionales, cuando se emplea PCM y ciclos termodinámicos en el cálculo de las energías de solvatación. De ellas solamente las que involucran al esquema **D** serían igualmente adecuadas para usarse en el cálculo de pKas de fenoles y ácidos carboxílicos. Además, considerando que de los esquemas propuestos en este trabajo **CI**, **F2** y **CNI** son con los que se obtienen buenos resultados para un mayor número de métodos de estructura electrónica, éstos fueron los que se seleccionaron para el estudio de aminas poli-funcionales (cuando el solvente se modela con PCM).

6.1.3.2. Mono-funcionales SMD

Los valores de EAP obtenidos se muestran de manera resumida en la Figura 18, los valores se reportan en el Anexo 16 y los datos para todos los esquemas de reacción con al menos un valor de $EAP < 2$ se reportan en la tabla 22. Comparando la figura 18 con la 17 se aprecia que, en el caso de aminas mono-funcionales, al utilizar SMD para modelar el solvente los valores de EAP se incrementan significativamente, en especial para los esquemas **B**, **CI**, **C2**, **C3**, **F4**, **CNI**, **CN2**, **CN3** y **FN4**. Este comportamiento es el contrario del observado para fenoles y ácidos carboxílicos (para estas dos familias de compuestos los valores de EAP eran menores con SMD que con PCM). A pesar de esto, para aminas mono-funcionales, usando SMD para modelar el solvente, también se encontraron numerosos casos en buen acuerdo

con los experimentos. En este caso, los esquemas de reacción con valores más bajos de EAPG son **A** (0.91), **D** (1.05) y **F3** (1.84).

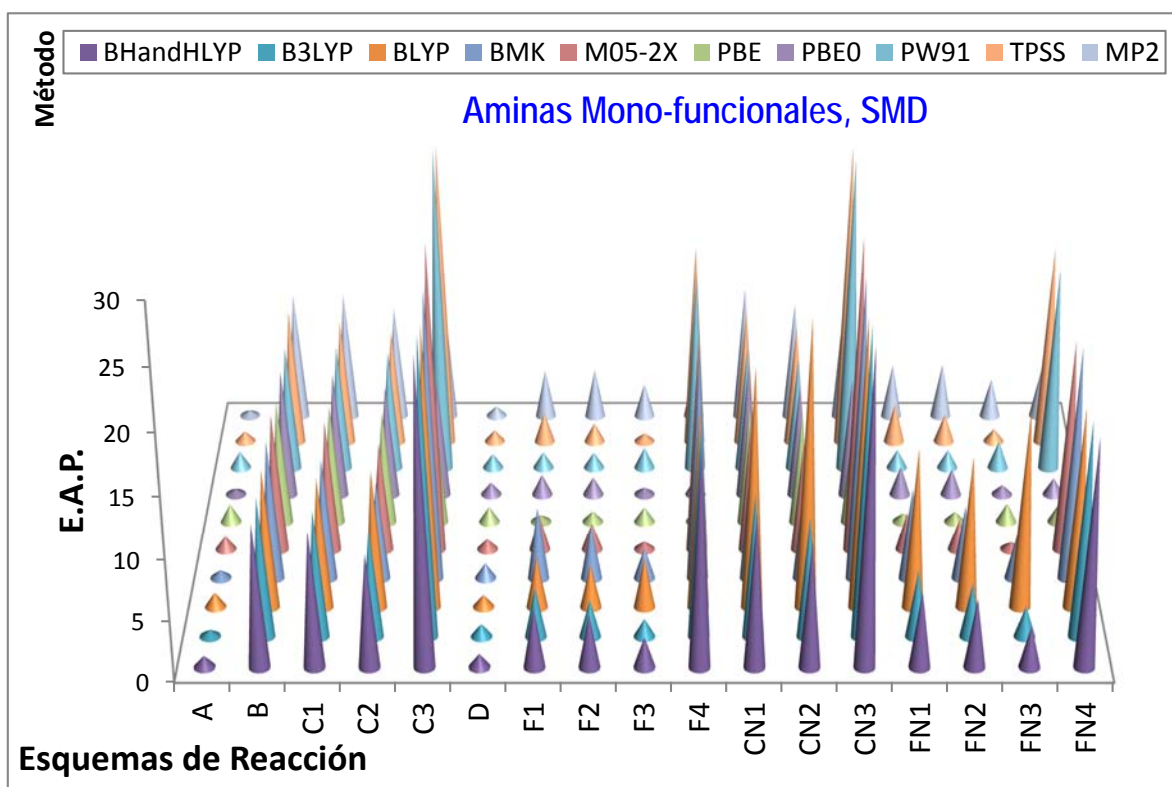


Figura 18. Errores absolutos promedio de los valores de pKa calculados con SMD para aminas mono-funcionales.

Para los esquemas **A** y **D** los valores de EAP obtenidos con todos los métodos de cálculo probados son menores a 2 unidades de pKa. En el caso del esquema **A**, además hay 6 métodos de cálculos con los que se obtuvieron valores de EAP < 1 (BHandHLYP, B3LYP, BKM, PBE0, TPSS y MP2). Los mejores resultados para este esquema se encontraron con MP2 (EAP = 0.43), PBE0 (EAP = 0.44) y B3LYP (EAP = 0.56). En el caso del esquema **D**, los valores de EAP obtenidos no solo son menores a 2 con todos los métodos de cálculo, sino también menores a 1.3. Esto sugiere que usar este esquema (con SMD) se pueden estimar pKas de aminas primarias en general, en buen acuerdo con los valores experimentales, con cualquier método de cálculo (de los empleados en este trabajo). Los valores más bajos de

EAP con el esquema **D** se obtuvieron para M05-2X (0.87), TPSS (0.95) y MP2 (0.80). En el caso del esquema **F3** se encontraron valores de EAP < 2 para 6 funcionales de la densidad, de ellos 3 con EAP < 1. Estos son M05-2X (0.64), PBE0 (0.58) y TPSS (0.69).

Tabla 22. Combinaciones esquema de reacción / método de cálculo con al menos un valor de EAP < 2 unidades de pKa para aminas mono-funcionales, usando SMD en el cálculo de las energías de solvatación, y valores de EAPG.

	BH&HLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2	EAPG
A	0.92	0.56	1.24	0.65	1.11	1.49	0.44	1.37	0.86	0.43	0.91
D	1.09	1.06	1.04	1.25	0.87	1.26	1.01	1.17	0.95	0.80	1.05
F1	5.00	4.14	4.29	5.99	2.84	0.55	1.70	1.33	2.48	3.96	3.23
F2	4.39	3.07	3.63	4.51	2.25	0.88	1.48	1.29	1.58	4.04	2.71
F3	2.51	1.50	4.20	2.54	0.64	1.24	0.58	1.70	0.69	2.77	1.84
F4	18.21	17.13	16.36	18.72	17.61	1.00	1.45	16.46	16.73	4.05	12.77
FN1	6.36	5.60	13.19	7.57	3.16	0.71	2.39	1.52	3.25	4.45	4.82
FN2	5.75	4.54	12.53	6.10	2.57	0.85	2.18	1.53	2.36	4.54	4.30
FN3	3.47	2.58	16.32	3.76	0.61	1.57	0.80	2.27	1.07	3.20	3.57
FN4	18.70	17.66	16.45	19.03	17.36	1.52	1.40	16.81	16.59	4.07	12.96

Resumiendo de manera similar a como se hizo para los resultados obtenidos con PCM, podemos decir que al usar SMD en el cálculo de la energía de solvatación de aminas mono-funcionales se encontraron 45 pares esquema de reacción / método de cálculo con valores de EAP < 2, de ellos 18 con EAP < 1. Los 10 mejores resultados se obtuvieron con:

- el esquema **A** con los métodos MP2, PBE0, B3LYP y BMK;
- el esquema **F1** con el funcional PBE;
- el esquema **F3** con los funcionales M05-2X, PBE0 y TPSS;
- el esquema **FN1** con el funcional PBE; y
- el esquema **FN3** con el funcional M05-2X.

Considerando ahora solamente los esquemas aplicables a todas las aminas mono-funcionales estudiadas, los casos en los que EAP < 1.3 serían:

- esquema **A** con BHandHLYP, B3LYP, BLYP, BMK, M05-2X, PBE0, TPSS y MP2

- esquema **D** con todos los métodos probados
- esquema **F1** con PBE
- esquema **F2** con PBE y PW91
- esquema **FN1** con PBE
- esquema **FN2** con PBE

De manera que cualquiera de estas combinaciones podría usarse exitosamente para calcular pKas de aminas mono-funcionales, cuando se emplea SMD en el cálculo de las energías de solvatación. Al igual que ocurre en el caso de los cálculos con PCM, en este caso de las combinaciones mencionadas anteriormente las que involucran al esquema **D** serían igualmente adecuadas para usarse en el cálculo de pKas de fenoles y ácidos carboxílicos. Además, considerando que con los esquemas propuestos en este trabajo **F1**, **F2**, **FN1** y **FN2** se obtienen buenos resultados cuando se consideran solo los esquemas aplicables a todas las aminas mono-funcionales, éstos se seleccionaron para el estudio de aminas poli-funcionales (cuando el solvente se modela con SMD). Además se incluirán también **F3** y **FN3** ya que estos también dan buenos resultados y pueden aplicarse a todas las aminas poli-funcionales estudiadas (todas son primarias o secundarias, no hay ninguna terciaria entre ellas).

6.1.3.3 Poli-funcionales PCM

Los valores de EAP obtenidos se muestran de manera resumida en la Figura 19 para aminas poli-funcionales, utilizando PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, los valores se reportan además en el Anexo 17. En este caso la figura es más sencilla que en los casos anteriores, porque (como ya se mencionó anteriormente) los cálculos se realizaron solamente con un subconjunto de esquemas de reacción y métodos de cálculo.

Para todos los esquemas de reacción estudiados se obtuvieron valores de EAP menores a dos unidades de pKa con algún método de estructura electrónica, excepto con el esquema **D**. Este fue un resultado sorprendente dado los buenos resultados obtenidos con este esquema en todos los casos analizados anteriormente (más adelante regresaremos a discutir este esquema en más detalle).

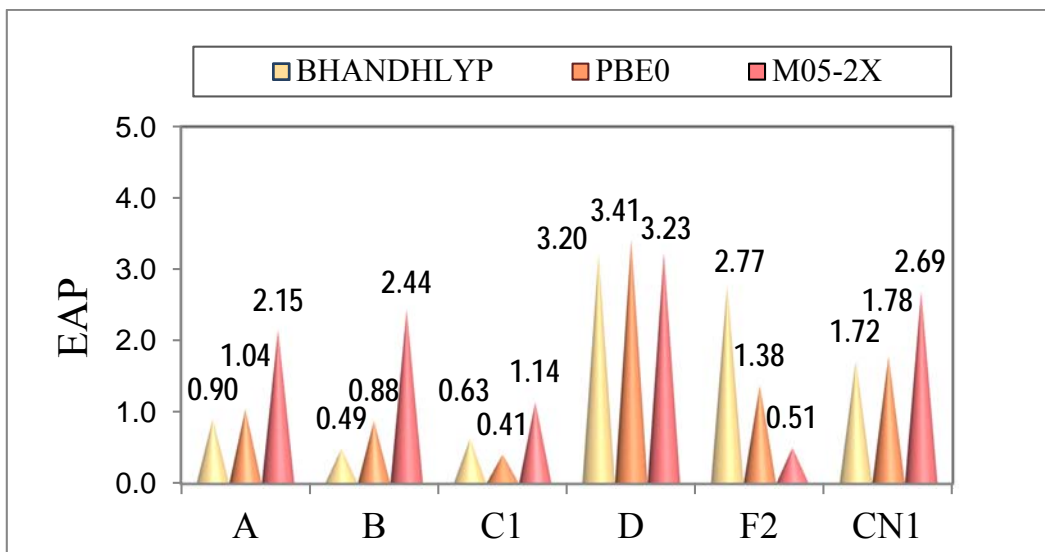


Figura 19. Errores absolutos promedio de los valores de pKa calculados, usando PCM y ciclos termodinámicos, para aminas poli-funcionales.

Los mejores resultados en general se obtuvieron con el esquema *C1*, con buenos valores de EAP para los tres funcionales de la densidad estudiados: PBE0 (0.41), BHandHLYP (0.63) y M05-2X (1.14). Para los esquemas *A*, *B* y *CN1* los mayores valores de EAP también se obtuvieron con el funcional M05-2X. En el caso del esquema *A* se obtuvieron valores de EAP menores (o cercanos) a la unidad de pKa tanto con BHandHLYP (0.90) como con PBE0 (1.04), mientras que para el esquema *B* con ambos funcionales se obtuvieron valores de EAP < 1 (0.49 y 0.88 respectivamente).

En este punto vale la pena regresar al esquema *D*, en particular a analizar en detalle el papel del ácido de referencia en los resultados obtenidos con este esquema. Los resultados mostrados en la figura 19 se obtuvieron con HRef = metil amina, que es una amina primaria (HP1). Esto se hizo considerando que el sitio de desprotonación más probable en tres de los cuatro neurotransmisores investigados pertenece a la misma categoría. Sin embargo, luego de ver los resultados obtenidos con esta selección, se decidió probar otras aminas como HRef. Estas son amonio (NH₄), una amina secundaria (HS1, dimetil amina), una amina terciaria (HT1, trimetil amina) y un neurotransmisor diferente al investigado (HN1, serotonina para la dopamina, y dopamina para el resto).

El análisis de los resultados obtenidos en cada caso (figura 20) muestra claramente lo sensible que es el esquema *D* al ácido de referencia utilizado. Cuando se usa amonio como HRef, los valores de EAP obtenidos son mayores a 4 unidades de pKa con todos los métodos de estructura electrónica empleados. O sea el amonio es incluso peor que la metil amina como HRef, para la cual los valores de EAP eran mayores a 3 pero menores que 4. Al usar una amina secundaria como HRef los valores de EAP se reducen significativamente. En este caso son menores, o cercanos, a 1 unidad de pKa con los tres funcionales de la densidad empleados. Cuando se usa una amina terciaria como HRef, los resultados obtenidos fueron sorprendentemente buenos (considerando que ninguno de los neurotransmisores investigados se corresponde a esta clasificación). En este caso tanto con BHandHLYP como con PBE0 se obtienen valores de EAP <1, mientras que con M05-2X este valor es significativamente mayor (1.81). Finalmente, los menores valores de EAP se obtienen (con los 3 funcionales de la densidad) cuando se usa otro neurotransmisor como HRef. Todos los valores de EAP obtenidos en este caso son menores a 1 unidad de pKa: BHandHLYP (0.47), PBE0 (0.56) y M05-2X (0.84). Sin embargo, también en este caso, PBE0 y BHandHLYP dan mejores resultados que M05-2X.

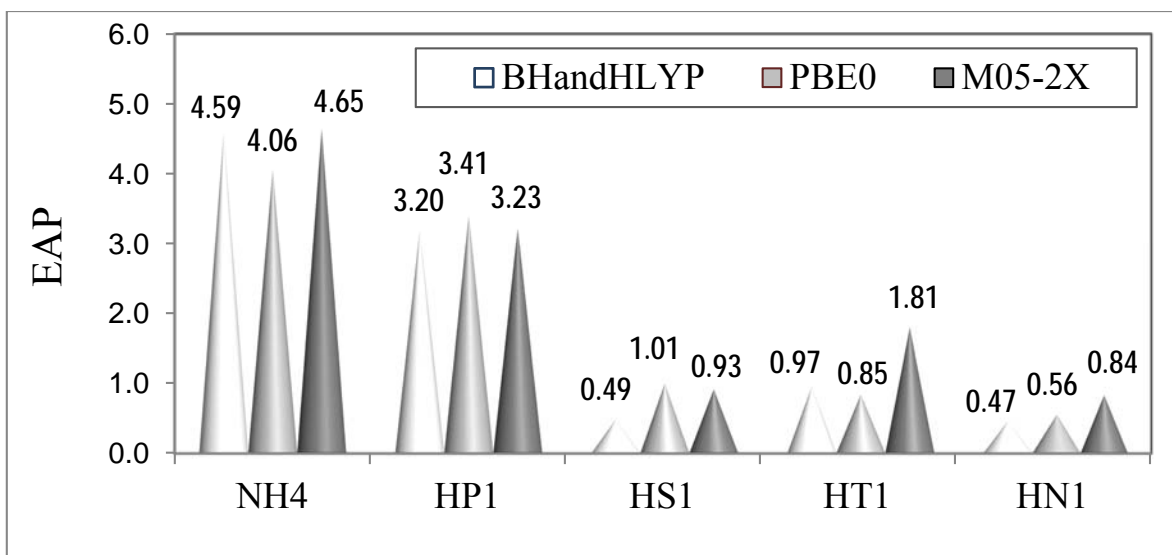


Figure 20. Errores absolutos promedio de los valores de pKa calculados aminas poli-funcionales usando el esquema *D* con diferentes ácidos de referencia.

De manera resumida puede decirse que para aminas poli-funcionales, cuando la energía de solvatación se calcula con PCM y ciclos termodinámicos, los mejores resultados (en base a EAP) se obtienen con:

- esquema **D** con todos los funcionales probados (BHandHLYP, PBE0 y M05-2X) siempre y cuando se use como HRef una especie muy parecida estructuralmente a la molécula de interés
- esquema **A** con BHandHLYP y PBE0
- esquema **B** con BHandHLYP y PBE0
- esquema **CI** con BHandHLYP, PBE0 y M05-2X
- esquema **F2** con PBE0 y M05-2X

Comparando estos resultados con los obtenidos para aminas mono-funcionales (cuando el solvente también se modela con PCM), se encuentra que con estas combinaciones esquema de reacción / método de cálculo se llega a buen acuerdo con los datos experimentales para ambos tipos de aminas (mono-funcionales y poli-funcionales). De manera que puede decirse que la hipótesis se cumple también en este caso.

6.1.3.4 Poli-funcionales SMD

La figura 21 muestra, de manera resumida, los valores de EAP obtenidos para aminas poli-funcionales, utilizando SMD en el cálculo de las energías de solvatación, los valores se reportan además en el Anexo 18. Con base a lo ya discutido en la sección anterior, para el esquema **D** se usó HN1 como HRef. De manera general puede decirse que los esquemas de reacción **A** y **F3** fueron los que dieron mejores resultados en este caso. Para ambos los valores de EAP fueron menores a 1 unidad de pKa con los tres funcionales de la densidad probados (PBE0, PBE y PW91). Además se obtuvieron valores de EAP < 0.5, para el esquema **A** con PW91 (0.34) y PBE (0.38) y para **F3** con PBE (0.34).

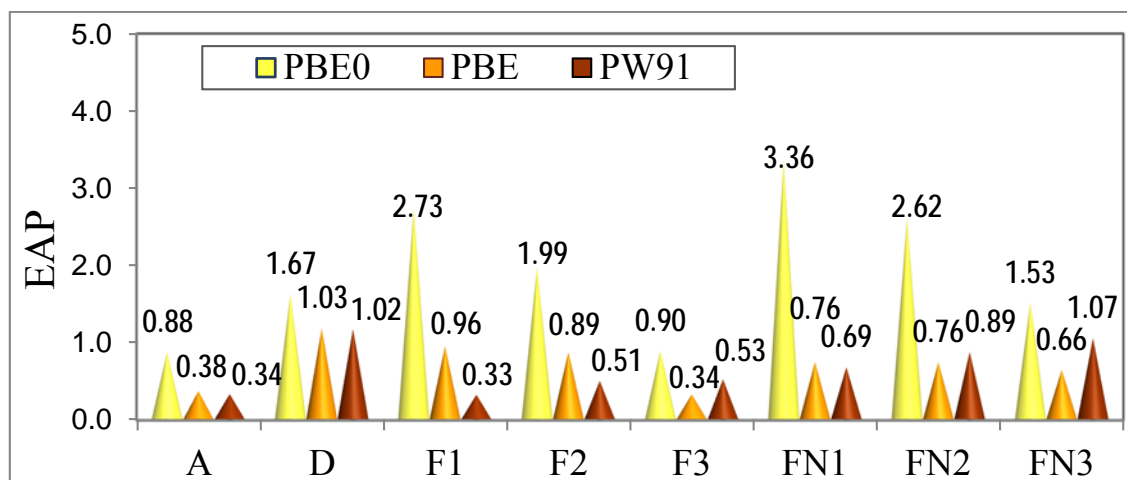


Figure 22. Errores absolutos de los valores de pKa calculados con SMD para las aminas poli-funcionales.

Para el esquema **D** los valores más bajos de EAP (~ 1.0) se obtuvieron con PBE y PW91. En el caso de los esquemas **F1**, **F2** y **FN1** los mejores resultados se obtuvieron con el funcional PW91 (0.33, 0.51 y 0.69, respectivamente), mientras que para los esquemas **F3**, **FN2** y **FN3** los mejores resultados se obtuvieron con el funcional PBE (0.34, 0.76 y 0.66, respectivamente). Es interesante notar que aminas poli-funcionales (específicamente neurotransmisores), cuando se usa SMD para modelar el solvente, es el único caso en el que PBE0 no se encuentra en el subconjunto de métodos de cálculo que producen los menores de EAP. Sin embargo, también es importante destacar que a pesar de esto se obtienen valores de pKas calculados con este funcional en buen acuerdo con la data experimental y significativamente inferiores al límite de error considerado actualmente como aceptable.

De manera resumida puede decirse que para aminas poli-funcionales, cuando la energía de solvatación se calcula con PCM y ciclos termodinámicos, los mejores resultados (en base a valores de EAP) se obtienen con:

- esquema **A** con PBE0, PBE y PW91
- esquema **D** con PBE0, PBE y PW91
- esquema **F1** con PBE y PW91
- esquema **F2** con PBE y PW91

- esquema ***FNI*** con PBE y PW91
- esquema ***FN2*** con PBE y PW91
- esquema ***FN3*** con PBE y PW91

Varios de estos también se encuentran entre los que dan buen acuerdo con el experimento para aminas mono-funcionales (***A*** con PBE0; ***D*** con PBE0, PBE y PW91; ***F2*** con PBE y PW91; y ***F1***, ***FNI*** y ***FN2*** con PBE). O sea que la hipótesis planteada también se cumple en este caso.

6.1.3.5 Conjunto completo de aminas con PCM

Los valores de EAP obtenidos para el conjunto completo de aminas estudiadas (mono-funcionales y poli-funcionales) se muestran en la Figura 23, los valores se reportan en el Anexo 19. Los esquemas de reacción con menores valores de EAP son el ***D*** y el ***CI***, lo cual es consistente con el análisis separado de moléculas mono-funcionales y poli-funcionales. Para ambos esquemas se obtienen valores de EAP menores a una unidad de pKa con los tres funcionales de la densidad probados. En el caso del esquema ***D*** los valores de EAP fueron 0.74 (BHandHLYP), 0.79 (PBE0) y 0.90 (M05-2X), mientras que para el esquema ***CI*** fueron 0.90 (BHandHLYP), 0.78 (PBE0) y 0.88 (M05-2X).

Para el esquema ***A*** no se obtuvo ningún valor de EAP < 1, pero sí dos valores con EAP < 2. Estos fueron los correspondientes a los funcionales BHandHLYP (1.18) y PBE0 (1.49). Para el esquema ***B*** el valor de EAP obtenido con M05-2X es también superior a las dos unidades de pKa. Sin embargo con BHandHLYP y PBE0 se obtuvieron valores de EAP < 1 (0.87 y 0.93, respectivamente). También se obtuvieron muy buenos resultados con el esquema ***F2***, si bien no para todos los funcionales de la densidad. Con este esquema fue el funcional BHandHLYP el que dio un valor de EAP > 2, mientras que con PBE0 y M05-2X se obtuvieron valores de EAP < 1 (0.78 y 0.52, respectivamente). Para el esquema ***CNI*** todos los valores de EAP fueron mayores a 1 unidad de pKa, aunque los obtenidos con BHandHLYP y PBE0 fueron menores a 2 (1.26 y 1.17, respectivamente)

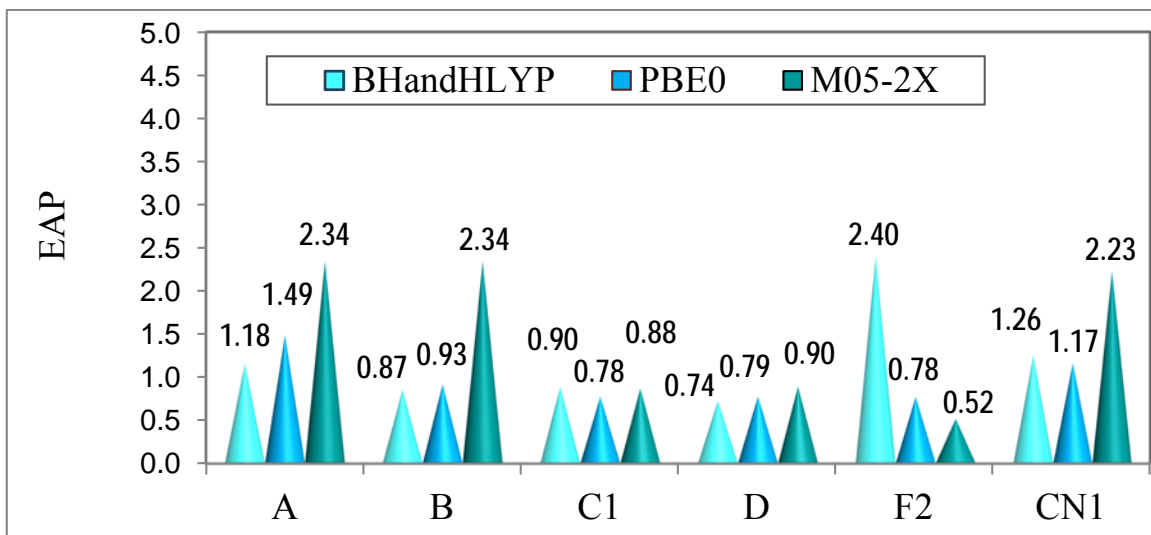


Figure 23. Errores absolutos de los valores de pKa calculados con PCM y ciclos termodinámico para el conjunto completo de aminas.

De manera resumida puede decirse que para el conjunto completo de aminas, usando PCM y ciclos termodinámicos para el cálculo de la energía de solvatación, los valores de pKa calculados en mejor acuerdo con los datos experimentales se obtuvieron para:

- el esquema **B** con BHandHLYP y PBE0;
- el esquema **CI** con BHandHLYP, PBE0 y M05-2X;
- el esquema **D** con BHandHLYP, PBE0 y M05-2X; y
- el esquema **F2** con PBE0 y M05-2X.

Además, analizando estos resultados con los obtenidos para fenoles y ácidos carboxílicos, puede decirse que (al menos con PCM) la única metodología que podría usarse sistemáticamente para obtener valores calculados de pKas -razonablemente buenos- para cualquiera de estas familias de compuestos es el esquema **D** con el funcional PBE0.

6.1.3.6 Conjunto completo de aminas con SMD

La Figura 24 muestra los resultados obtenidos para el conjunto completo de aminas estudiadas, cuando se modela el solvente con SMD. Además los valores se reportan en el Anexo 20. Los esquemas de reacción **A** y **F3** son los que, de manera general, dan los mejores

resultados. Para el esquema *A* se obtuvieron valores de EAP menores a una unidad de pKa con los tres funcionales de la densidad probados. Estos valores fueron 0.88 (PBE0), 0.38 (PBE) y 0.34 (PW91). Para el esquema *F3* se obtuvieron valores de EAP < 1 con PBE0 (0.74) y PBE (0.79), mientras que con PW91 el valor es superior a 1, aunque solo ligeramente (1.12).

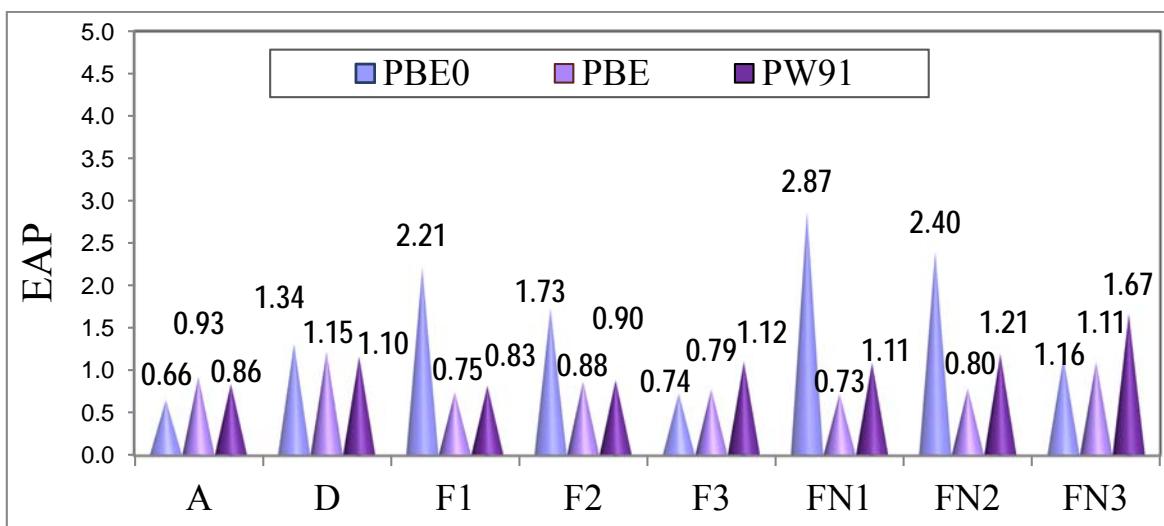


Figure 25. Errores absolutos de los valores de pKa calculados con SMD para el conjunto completo de aminas.

Para el esquema *D* no se obtuvo ningún valor de EAP < 1, aunque en todos los casos los valores de EAP son significativamente menores que 2 (1.34, 1.15 y 1.10 con PBE0, PBE y PW91, respectivamente). En el caso del esquema *F1* el valor de EAP obtenido con PBE0 es superior a las dos unidades de pKa, mientras que con PBE y PW91 se obtuvieron valores de EAP < 1 (0.75 y 0.83, respectivamente). Lo mismo sucede con el esquema *F2*, en este caso los valores de EAP obtenidos fueron 1.73, 0.88 y 0.90 con PBE0, PBE y PW91, respectivamente.

Para los esquemas *FN1*, *FN2* y *FN3* el funcional PBE0 fue el que mostró el peor desempeño. Sin embargo es interesante notar que este funcional fue el único para el cual los resultados con los esquemas *FN* mejoran de manera sistemática al incrementar el número de moléculas de agua explícitas. Esta tendencia también se observó al modelar el solvente con PCM. Para

los esquemas *FN1* y *FN2* los valores más bajos de EAP se obtuvieron con PBE y en ambos casos fueron menores a 1 unidad de pKa (0.73 y 0.80, respectivamente). Los valores de EAP obtenidos para estos esquemas con PW91 fueron mayores a 1 pero menores a 2 (1.11 y 1.21, respectivamente). En el caso del esquema *FN3* todos los valores de EAP se encuentran entre 1 y 2. Estos son 1.16, 1.11 y 1.67 con PBE0, PBE y PW91, respectivamente. Es interesante notar que de este es el único de los esquemas *FN* para el que el menor valor de EAP se obtiene con PBE0.

De manera resumida puede decirse que para el conjunto completo de aminas, usando SMD para el cálculo de la energía de solvatación, los valores de pKa calculados en mejor acuerdo con los datos experimentales se obtuvieron para:

- el esquema *A* con PBE0, PBE y PW91;
- el esquema *F1* con PBE y PW91;
- el esquema *F2* con PBE y PW91;
- el esquema *F3* con PBE0 y PBE;
- el esquema *FN1* con PBE; y
- el esquema *FN2* con PBE.

Sin embargo, al comparar los resultados obtenidos para aminas con los de fenoles y ácidos carboxílicos, puede decirse que (al igual que con PCM) la única metodología que podría usarse sistemáticamente con SMD para obtener valores calculados de pKas -razonablemente buenos- para cualquiera de estas familias de compuestos es el esquema *D* con el funcional PBE0.

6.1.3.7. Método de correlaciones para aminas

Los resultados obtenidos con el método de las correlaciones para el cálculo de pKas de aminas se muestran en la tabla 23. Con todos los métodos de cálculo probados se encontró que los coeficientes de correlación (R^2) son mayores cercanos a 0.6. Estos valores de R^2 son significativamente más bajos que para fenoles y ácidos carboxílicos. Lo que puede deberse a la diferente naturaleza del sitio ácido, dependiendo de que las aminas sean primarias,

secundarias o terciarias. Esto hace que el acuerdo con los valores experimentales sea peor en este caso que para las otras dos familias de compuestos estudiadas. Para aminas, si bien todos los valores de EAP están muy por debajo de las 2 unidades de pKa, son bastante más altos que para fenoles y ácidos carboxílicos (en este caso se encuentran alrededor de 1 unidad de pKa). El valor más bajo de EAP se obtuvo con PBE0 (1.01) y el más alto con PW91 (1.19).

Tabla 23. Parámetros m y C_0 , coeficiente de correlación lineal y errores absoluto promedio para el cálculo de pKas de aminas usando el método de las correlaciones.

	m	C_0	R^2	EAP
PBE	0.426	-110.101	0.643	1.18
PBE0	0.421	-109.389	0.604	1.01
PW91	0.434	-112.286	0.626	1.19

Además los valores obtenidos con el funcional PBE0, que también da buenos resultados con la estrategia de los esquemas de reacción y para fenoles y ácidos carboxílicos, son razonablemente buenos. Sin embargo, contrario a lo que ocurría para los otros tipos de compuestos, estos valores son mayores que los obtenidos con algunos de los esquemas de reacción y mayores a la unidad de pKa.

Para esta familia de compuestos en particular, considerando la variedad de sitios ácidos posibles (aminas primarias, secundarias y terciarias, así como con sustituyentes alifáticos y aromáticos) es posible que el conjunto de moléculas usadas en la correlación no sea suficiente. Por esta razón se extendió este conjunto de moléculas incluyendo methoxylamina, allyl-amina, aminoacetonitrilo, morfolina, piperidina, aniline, *p*-amino-anilina, *o*-cloro-anilina, *m*-hidroxi-anilina, *p*-cloro-anilina, piridina y pirazina. Esto se hizo solamente para el funcional PBE0 ya que es de manera general, el que da mejores resultados en todos los casos estudiados. Al ampliar el conjunto de aminas usadas el coeficiente de correlación mejoró significativamente (de 0.604 a 0.968).

Los parámetros obtenidos de esta correlación fueron $m = 0.482$ y $C_0 = -126.917$, o sea también muy diferentes a los obtenidos con el conjunto reducido de moléculas. Al usar estos parámetros en el cálculo de los pKa de las aminas poli-funcionales el valor de EAP disminuyó

de 1.01 a 0.82. Esto indica, no solo que el método de las correlaciones puede usarse para calcular pKa de aminas y obtener un buen acuerdo con el experimento, sino que en el caso particular de esta familia de compuestos es importante usar un conjunto suficientemente amplio de moléculas para obtener los parámetros m y C_0 .

6.1.3.8 Generalidades para aminas

Al igual que en el caso de fenoles y ácidos carboxílicos, para aminas el aspecto que tiene mayor influencia a la hora de obtener buena correspondencia entre los valores de pKa calculados y los datos experimentales es el esquema de reacción. El segundo aspecto en importancia sigue siendo el modelo continuo de solvente y en tercer lugar se encuentra el método de estructura electrónica. Sin embargo, a diferencia de lo que ocurría con fenoles y ácidos carboxílicos, en caso de aminas, usar SMD en el cálculo de energías de solvatación no mejora significativamente los resultados obtenidos con respecto a PCM.

En el caso de aminas cuando se usa PCM y ciclos termodinámicos en el cálculo de las energías de solvatación, los esquemas **F2** y **D** son los que de manera general dan resultados en mejor acuerdo con los experimentos tanto para moléculas mono-funcionales como para moléculas poli-funcionales. Con estos esquemas se obtienen valores de EAP < 1 con todos los funcionales de la densidad probados en este caso (BHandHLYP, PBE0 y M05-2X). En el caso de SMD esto se logra con los esquemas **A** y **F3** y los funcionales probados fueron PBE0, PBE y PW91.

Es importante destacar que, en el caso de aminas, el acuerdo con los datos experimentales para el esquema **D** resultó ser muy dependiente de la selección de HRef, especialmente para moléculas poli-funcionales (neurotransmisores, en este trabajo). De hecho el único modo de obtener valores de pKa para estos compuestos con desviaciones pequeñas con respecto al experimento es usar como HRef una molécula muy parecida a la de interés. No es suficiente que sea del mismo tipo de aminas (primaria, secundaria o terciaria), sino que tiene que tener más características estructurales en común, como por ejemplo que los grupos (\neq H) enlazados al nitrógeno sean similares.

Comparando los resultados obtenidos para aminas con los de fenoles y ácidos carboxílicos, puede decirse que, tanto con PCM como con SMD, la única metodología que podría usarse sistemáticamente para obtener valores calculados de pKas en acuerdo razonable con los experimentales para cualquiera de estas familias de compuestos es el esquema **D** con el funcional PBE0. En el caso de aminas (con SMD), a pesar de que la combinación esquema **D** / PBE0 no se encuentra entre las que tienen los valores más bajos de EAP, estos valores son cercanos a 1, lo que se encuentra muy por debajo del límite considerado como aceptable actualmente.

También es importante destacar que, al igual que para las otras familia de compuestos estudiadas, para aminas se cumple la hipótesis planteada, ya que fue posible encontrar metodologías en buen acuerdo con los experimentos tanto para especies relativamente sencillas (mono-funcionales) como para moléculas más complejas (poli-funcionales).

Otra particularidad de la familia de aminas es que el método de las correlaciones puede dar buenos resultados para el cálculo de pKas, pero para que eso ocurra el conjunto de moléculas usadas para obtener los parámetros m y C_0 tiene que ser suficientemente amplio. De otro modo la correlación lineal no es buena (valores relativamente bajos de R^2) y los valores de EAP superan la unidad de pKa. Al hacer esto, se logra un buen acuerdo con los valores experimentales. Por ejemplo usando el funcional PBE0 se obtuvo un valor de $EAP = 0.71$.

6.1.4. Efectos del conjunto de funciones de base

Teniendo en cuenta que el funcional de la densidad PBE0 en combinación con el esquema **D**, o usando el método de las correlaciones, parece ser la recomendación general para el cálculo de pKas de compuestos orgánicos, el efecto del conjunto de funciones de base se probó para este funcional. Además como, tanto para el esquema **D** como para el método de las correlaciones, el valor de energía que determina la calidad de los resultados es la diferencia de energía entre cada ácido y su base conjugada ($\Delta G_{BA} = G_B - G_A$), este fue el valor empleado para investigar el efecto del conjunto de funciones de base. Además, dado el gran número de cálculos realizados en este trabajo, este análisis se hizo para un conjunto

reducido de moléculas (2 fenoles, 5 ácidos carboxílicos y 5 aminas). Los resultados obtenidos se muestran en la tabla 24.

Tabla 24. Valores de ΔG_{BA} calculados con PBE0 y diferentes conjuntos de funciones de base. Todos los valores se reportan en Kcal/mol.

	6-31+G(d)	6-311+G(d)	6-311+G(d,p)	6-311++G(d,p)
4-metil fenol	288.36	287.76	291.66	291.60
3-metoxi fenol	286.90	286.26	290.08	290.01
ácido fórmico	273.76	273.51	277.45	277.47
ácido acético	277.47	277.33	281.03	281.15
ácido cloroacético	271.83	271.53	275.25	275.28
ácido benzoico	275.80	275.54	279.19	279.23
ácido crotonico	277.37	277.13	280.87	280.89
metil-amina	283.46	283.81	284.48	284.51
dimetil-amina	283.07	282.99	284.17	284.20
piperidina	284.71	284.67	285.62	285.65
trimetil-amina	281.57	281.16	282.94	282.96
N-alil-piperidina	282.96	282.55	284.03	284.04

En todos los casos se usaron conjuntos de funciones de base con funciones difusas (al menos para los átomos \neq H). Esta decisión se tomó teniendo en cuenta que hay aniones involucrados en los equilibrios ácido-base de fenoles y ácidos carboxílicos. Por otra parte, se fue aumentando el tamaño de los conjuntos de funciones de base paulatinamente, para poder analizar separadamente el efecto de las diferentes modificaciones.

Al aumentar el conjunto de funciones de base de 6-31+G(d) a 6-311+G(d), las diferencias en ΔG_{BA} son en todos los casos menores a 1 kcal/mol. Las mayores diferencias se encontraron para fenoles, con variaciones de alrededor de 0.6 kcal/mol, para los otros 2 tipos de compuestos este valor fue menor a 0.5 kcal/mol. De modo que puede decirse el usar una base de valencia desdoblada triple Z, en lugar de doble Z, en el cálculo de pKa tiene efectos pequeños, pero no despreciables en los resultados.

La mayor diferencia en ΔG_{BA} aparece al aumentar el conjunto de funciones de base de 6-311+G(d) a 6-311+G(d,p). Este es un resultado esperado ya que la diferencia radica en las funciones de polarización para H, y estos son precisamente los átomos involucrados en equilibrios ácido-base. Es por ello que no se recomienda usar conjuntos de funciones de base con simple polarización en el estudio de estos equilibrios.

Por el contrario, la menor diferencia ocurre al cambiar de 6-311+G(d,p) a 6-311++G(d,p). Este también es un resultado lógico ya que el agregar funciones difusas a átomos de H, para sistemas donde no hay hidruros, no debe mejorar significativamente la descripción del sistema. En este caso la máxima diferencia encontrada se corresponde al ácido acético y es de solo 0.11 kcal/mol. Este valor está muy por debajo de los errores discutidos para el cálculo de pKas de fenoles, ácidos carboxílicos y aminas. Por lo tanto puede decirse que cualquiera de estos dos conjuntos de funciones de base podría usarse de manera igualmente confiable para calcular estos valores de pKa.

6.1.5. Generalidades pKas.

Hacer generalizaciones en cuanto a estrategias de cálculo de pKa de compuestos orgánicos en buen acuerdo con los valores experimentales es una ardua tarea. Esto se debe primeramente a que las estrategias con menores desviaciones respecto al experimento no son las mismas para diferentes familias de compuestos.

En segundo lugar hay numerosos factores que influyen en la calidad de resultado, incluyendo el esquema de reacción utilizado, la forma en que se modela el solvente (tanto en lo referente al modelo continuo, como al número de moléculas de agua explícitas consideradas), el método de estructural electrónica empleado y la naturaleza del sitio ácido. Este último factor se refiere al tipo de sitio (en nuestro caso fenoles, ácidos carboxílicos y aminas) pero también a detalles más específicos como por ejemplo si las aminas son primarias, secundarias o terciarias o si los compuestos son mono o poli-funcionales.

A pesar de estas complicaciones, el estudio realizado permitió hacer varias generalizaciones para el cálculo de pKas de compuestos orgánicos y en base a ellas poder hacer también

recomendaciones en este sentido. Probablemente la más importante de ellas es que utilizando el funcional PBE0 con el esquema *D*, o el método de correlaciones, (y SMD para modelar el solvente) podrían obtenerse valores calculados de pKas en acuerdo razonable con los datos experimentales. Sin embargo es importante destacar el acuerdo con los datos experimentales para este esquema es muy dependiente de la selección de HRef. Aunque esto es cierto de forma general, se hizo especialmente evidente para aminas poli-funcionales (neurotransmisores, en este trabajo). De hecho el único modo de obtener valores de pKa para estos compuestos con desviaciones pequeñas con respecto al experimento fue usar como HRef una molécula muy parecida a la de interés.

En cuanto a la importancia relativa de los diferentes aspectos investigados, en la metodología de cálculo, también puede hacerse una generalización importante. Esta es que el aspecto que más influye para obtener buena correspondencia entre los valores de pKa calculados y los datos experimentales es el esquema de reacción. El segundo aspecto en importancia es la estrategia usada para modelar el solvente y en tercer lugar se encuentra el método de estructura electrónica. En cuanto a los esquemas de reacción, aquellos con idéntico número y tipo de iones a ambos lados del equilibrio son los que conducen a mejores resultados. Se asume que esto se debe a que es en estos casos donde ocurre la máxima cancelación de errores.

Además, se encontró que el método de las correlaciones puede usarse para calcular valores de pKa en buen acuerdo con los datos experimentales tanto para fenoles, como para ácidos carboxílicos y aminas. Sin embargo, para lograr esto en el caso particular las aminas es crucial usar un conjunto suficientemente amplio de moléculas a la hora de obtener los parámetros empíricos que se usaran en el cálculo de pKas.

Con respecto al método de las correlaciones, también vale la pena mencionar que este método debe usarse separadamente para cada familia de compuestos. Esto se concluye en base a los resultados obtenidos en este trabajo para fenoles, ácidos carboxílicos y aminas. Como se muestra en la figura 25, al graficar de manera conjunta los valores de pKa y las diferencia en energía libre de Gibbs para cada par ácido – base conjugad (ΔG_{AB}) de las tres familias de compuestos se observa un agrupamiento por familia. Esto quiere decir que los valores óptimos de m y C_0 son diferentes para cada una de estas familias, e intentar obtener una

correlación lineal única para todos los compuestos estudiados empeoraría la correspondencia entre valores calculados y experimentales.

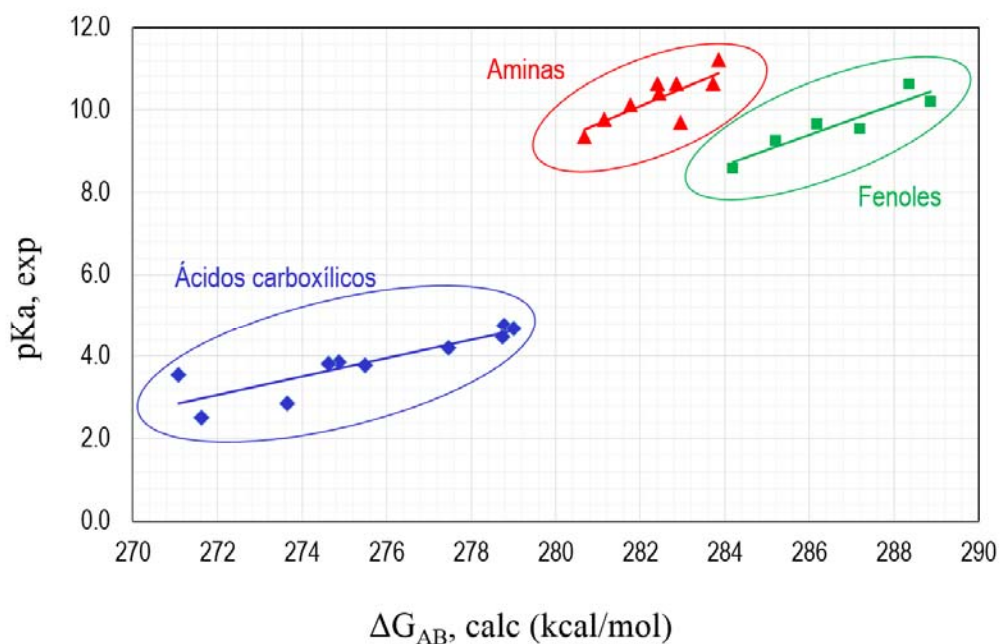


Figura 25. Correlaciones entre valores experimentales de pKa y diferencia en energía libre de Gibbs para cada par ácido – base conjugad (ΔG_{AB}).

Otra recomendación, basada en la investigación realizada, es que el conjunto de funciones de bases a usar en el cálculo de pKas tenga funciones difusas al menos para átomos \neq H, y funciones de polarización tanto para H como para átomos más pesados. De los conjuntos de funciones de bases probados se recomiendan 6-311+G(d,p) o 6-311++G(d,p) para este propósito.

Finalmente, para todas las familias de compuestos estudiadas (fenoles, ácidos carboxílicos y aminas), se cumple la hipótesis planteada, ya que fue posible encontrar metodologías que estuvieran en buen acuerdo con los experimentos tanto para especies relativamente sencillas (mono-funcionales) como para moléculas más complejas (poli-funcionales).

6.2 Constantes de Acomplejamiento

6.2.1 Geometría del Cu(II) en solución acuosa

Es importante conocer el ambiente de coordinación alrededor del Cu(II) en solución acuosa antes de realizar modelaciones que incluyan a esta especie. Como la estructura de este ión en sólidos es octaédrica, se podría asumir que este es también el caso en solución acuosa. Sin embargo está descrito que el cobre hexa-coordinado tiene una estructura octaédrica severamente distorsionada. Esta distorsión consiste en que las distancias en uno de los ejes de simetría (2 enlaces) es mucho mayor que en los otros dos ejes (4 enlaces). Se ha reportado que esta diferencia es de aproximadamente 1 Å.⁷⁴ De modo que se puede asumir que en solución acuosa para los dos ligantes más débilmente enlazados las interacciones con el solvente pueden ser suficientemente fuertes para que la geometría del Cu(II) se vuelva tetra-coordinada plana. Esto ha generado controversia en la literatura y se han realizado estudios teóricos al respecto, tratando de dar una respuesta definitiva a cómo es la geometría de este ión en solución acuosa.

En 2008 Bryantsev y colaboradores⁵¹ (usando B3LYP y el modelo continuo de solvente COSMO) investigaron la estructura de $[\text{Cu}(\text{H}_2\text{O})_n]^{2+}$ con $n = 4-18$ y para números de coordinación 4, 5 y 6. Encontraron que en solución acuosa la estructura de menor energía es penta-coordinada, con geometría de pirámide cuadrada. Sin embargo las diferencias de energía que fueron pequeñas (~ 1.4 kcal/mol), e inferiores al error del método (3.5 kcal/mol) por lo que propusieron que varias especies pueden coexistir.

Un año más tarde, en otro estudio del mismo grupo se investigó la estructura de $[\text{Cu}(\text{CH}_2\text{NH}_2)(\text{H}_2\text{O})_{n-1}]^{2+}$ y $[\text{Cu}(\text{OH})_x(\text{H}_2\text{O})_{n-x}]^{2-x}$ con el mismo método de cálculo para $n = 4-8, 18$, $x = 1-3$ y números de coordinación 4, 5 y 6.²² Esta vez los resultados fueron que para las especies con $n = 4-8$ la geometría tetra-coordinada cuadrado plana era la más estable en todos los casos excepto para $[\text{Cu}(\text{OH})_3(\text{H}_2\text{O})]^-$. Por el contrario para $n = 18$ la estructura más estable fue la penta-coordinada, con geometría de pirámide cuadrada, para $[\text{Cu}(\text{CH}_2\text{NH}_2)(\text{H}_2\text{O})_3]^{2+}$ y $[\text{Cu}(\text{OH})(\text{H}_2\text{O})_3]^+$, mientras que para $[\text{Cu}(\text{OH})_2(\text{H}_2\text{O})_2]$ y $[\text{Cu}(\text{OH})_3(\text{H}_2\text{O})]^-$ fue la geometría tetra-coordinada cuadrado plana. Además las diferencias

de energía fueron pequeñas nuevamente (0.23-2.65 kcal/mol) por lo que repitieron la propuesta de que varias especies pueden coexistir.

Por su parte Ortega-Castro y colaboradores,⁷⁵ investigaron las estructuras más probable para los complejos de Cu(II) con varios ligandos (amino-guanidina, piridoxamina, compuesto de Amadori y ácido ascórbico). Este estudio se realizó también con B3LYP, pero usando PCM como modelo continuo de solvente. En este caso se encontró que la forma cuadrada plana era la más estable para todos los ligandos estudiados. Las diferencias de energía en este caso fueron mayores a las 10 kcal/mol.

De manera que puede decirse que, aunque la mayoría de los estudios teóricos realizados parece soportar la idea de que el Cu(II) tenga una estructura tetra-coordinada cuadrada plana, esto no ha sido demostrado definitivamente. Además, aparentemente el modelo continuo usado para modelar el solvente tiene una influencia en las energías relativas de las diferentes conformaciones. No hay estudios previos reportados usando SMD, que es el modelo usado en este trabajo. Es por todo esto que aquí incluimos un estudio conformacional para escoger la geometría del Cu(II) que se usará en el resto del trabajo.

Para ello se escogió el funcional M05 (y todos los cálculos se realizaron en solución con SMD). Como los resultados parecen ser dependientes del ligante escogido, se modelaron tres sistemas diferentes: $\text{Cu}(\text{H}_2\text{O})_6^{2+}$, $\text{Cu}(\text{NH}_3)_4(\text{H}_2\text{O})_2^{2+}$ y $\text{Cu}(\text{CH}_3\text{COO})_2(\text{H}_2\text{O})_2$. Estos se escogieron por tener los mismos ligantes que se estudiarán más adelante.

Para $\text{Cu}(\text{H}_2\text{O})_6^{2+}$ se encontró que la estructura de mínima energía es la tetra-coordinada plana, o sea con cuatro moléculas de agua coordinadas a Cu(II) y las otras dos interaccionan con las moléculas de agua de la esfera de coordinación. Dos estructuras diferentes se encontraron para este caso, dependiendo de la ubicación de las aguas de solvatación, y estas tienen energías muy similares entre sí, se diferencian solo en 0.22 kcal/mol. Por el contrario la estructura octaédrica tiene una energía ~14 kcal/mol mayor. En la estructura octaédrica obtenida en el presente trabajo dos de las aguas de coordinación tienen enlaces más largos con el Cu que las otras cuatro. La diferencia es menor que la reportada en estado sólido, pero también significativa (alrededor de 0.8 Å).

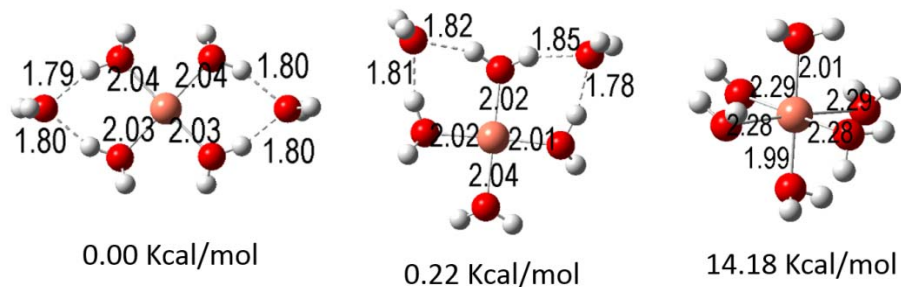


Figura 26. Confórmeros estudiados para el sistema $\text{Cu}(\text{H}_2\text{O})_6^{2+}$ y sus energías relativas con respecto al confórmero más estable.

Para $\text{Cu}(\text{NH}_3)_4(\text{H}_2\text{O})_2^{2+}$ se encontraron resultados similares en el sentido de que las dos estructuras con geometría cuadrada plana fueron de menor energía que la octaédrica (figura 27). Sin embargo en este caso esta última fue solamente ~ 3 kcal/mol mayor que la menor de todas. Por otra parte las dos estructuras tetra-coordinadas se diferencian en 0.55 kcal/mol. Lo que sí fue similar al caso anterior, con H_2O como ligantes, es que la geometría octaédrica es distorsionada con 2 enlaces significativamente más largos que el resto.

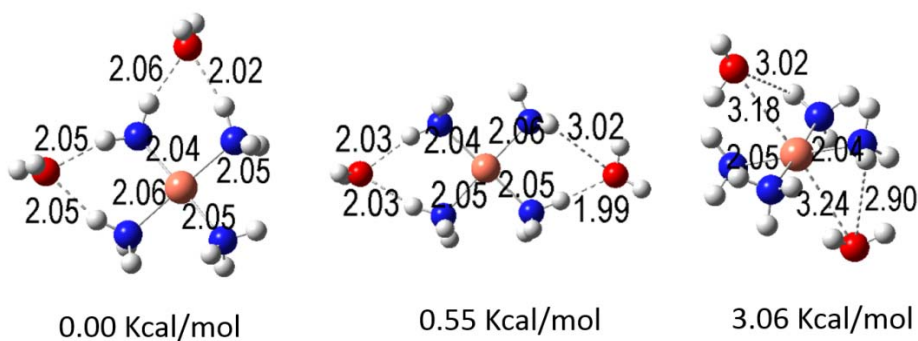


Figura 27. Confórmeros estudiados para el sistema $\text{Cu}(\text{NH}_3)_4(\text{H}_2\text{O})_2^{2+}$ y sus energías relativas con respecto al confórmero más estable.

En el caso de $\text{Cu}(\text{CH}_3\text{COO})_2(\text{H}_2\text{O})_2$ solamente se pudo encontrar un confórmero (figura 28). A pesar de que se realizaron múltiples intentos de localizar una estructura octaédrica esto no fue posible. En todos los casos las dos moléculas de agua migraban a solvatar los oxígenos

de los acetatos. La estructura encontrada presenta una geometría cuadrado plana, al igual que fue el caso de las estructuras de mínima energía para ligantes H₂O y NH₃.

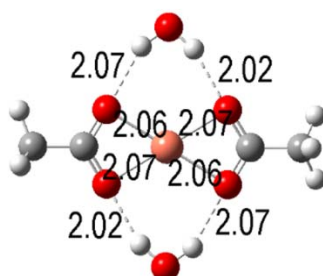


Figura 28. Único conformero encontrado para Cu(CH₃COO)₂(H₂O)₂.

De modo que para cualquiera de los ligandos investigados, los resultados obtenidos aquí concuerdan con los obtenidos por Ortega-Castro y colaboradores con B3LYP y PCM. Con base a lo encontrado, el ambiente de coordinación alrededor del Cu(II) en solución acuosa que se usará en el resto del trabajo corresponde a la estructura tetra-coordinada cuadrado plana.

6.2.2 Validación de estrategia de puntos simples

Como se mencionó en la sección de metodología, se exploró la posibilidad de hacer optimizaciones de geometría y cálculos de frecuencia con un solo método de cálculo y obtener solamente energías de punto simple para el resto de los métodos. Esto con la idea de seguir una estrategia que permitiera minimizar el tiempo de cómputo. Sin embargo, para poder usar esta estrategia debe comprobarse primeramente que las geometrías obtenidas, al realizar optimizaciones con los diferentes métodos de cálculo, son similares. Para ello se seleccionó la molécula [Cu(H₂O)₂(NH₃)₂]²⁺ ya que es sencilla y presenta dos ligandos diferentes (Figura 2).

Las diferencias de parámetros geométricos (distancias y ángulos) se calcularon tomando como referencia la geometría obtenida con el funcional M05 y se expresaron como porcentaje para poder comparar justamente valores diferentes. Los resultados obtenidos para distancias

de enlace se muestran en la tabla 25, mientras que los correspondientes a ángulos de enlace se reportan en la tabla 26.

Tabla 25. Porcentaje de desviación para distancias de enlace

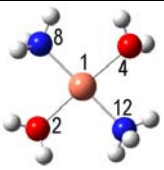
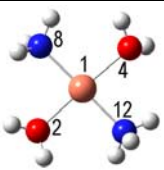
	d(1,8)	d(1,12)	d(1,2)	d(1,4)	Promedio
PBE	0.472	0.441	0.590	0.979	0.621
PBE0	0.558	0.534	1.326	2.527	1.236
BLYP	0.612	0.594	0.913	0.467	0.646
B3LYP	0.254	0.441	0.767	1.530	0.748
PW91	0.678	0.528	1.117	1.446	0.942
M06	1.114	1.167	1.181	1.714	1.294
BMK	0.291	0.267	0.704	1.756	0.754
BH&HLYP	0.363	0.243	0.988	1.572	0.791
TPSS	0.313	0.325	1.600	2.343	1.145

Tabla 26. Porcentaje de desviación para ángulos de enlace

	A(8,1,4)	A(8,1,2)	A(12,1,2)	A(12,1,4)	Promedio
PBE	0.895	1.710	0.975	1.741	1.330
PBE0	0.482	0.390	1.040	1.341	0.813
BLYP	0.940	2.186	2.120	1.071	1.579
B3LYP	0.544	0.421	1.010	1.358	0.833
PW91	0.632	0.067	0.219	0.879	0.449
M06	0.229	0.222	1.259	1.379	0.772
BMK	0.874	0.032	0.948	2.075	0.982
BH&HLYP	1.436	0.132	0.152	1.585	0.826
TPSS	0.258	0.017	0.638	0.191	0.276

En el caso de las distancias el menor valor obtenido es 0.24% y el mayor 2.53%. Además en la mayoría de los casos (25 de 36) estas diferencias son menores al 1%. Curiosamente, el método que sistemáticamente da mayores diferencias en distancias es el M06, a pesar de ser el más parecido a M05 en cuanto a su formalismo. Los otros métodos con desviaciones porcentuales promedio superiores al 1%, en el caso de distancias de enlace, fueron PBE0 y TPSS.

Para los ángulos de enlace se encontró que las desviaciones porcentuales varían de 0.02% a 2.19%. En este caso los dos únicos métodos con desviaciones porcentuales promedio superiores al 1% fueron PBE y BLYP. Además, al igual que para distancias, en la mayoría de los casos (22 de 36) las desviaciones fueron son menores al 1%.

Considerando que tanto para distancias como para ángulos de enlace las diferencias porcentuales de los otros 9 funcionales de la densidad probados, con respecto a M05, son menores al 2.6% en todos los casos, se concluyó que las diferencias son suficientemente pequeñas como para que pueda usarse la estrategia de puntos simples. Por lo tanto ésta se utilizó en el resto de la investigación de constantes de complejamiento.

Para los resultados que se describen a continuación las geometrías y los cálculos de frecuencia se realizaron solamente con el funcional M05, mientras que para el resto de los métodos se obtuvieron las energías a partir de cálculos de punto simple. La energía libre de Gibbs se obtuvo para cada especie química, y método de cálculo, sumando la energía electrónica obtenida con dicho método para esa especie más la corrección termodinámica correspondiente (obtenida con M05).

6.2.3 Complejos de Cu (II) con amoniac

En el caso de este ligante fue necesario hacer una exploración adicional del espacio conformacional para sistemas con 6 o más moléculas de agua explícitas. Para explicar los detalles de esto tomemos como ejemplo los complejos de Cu(II) involucrados en el esquema *EFS6*, o sea $[\text{Cu}(\text{H}_2\text{O})_3(\text{NH}_3)]^{+2} \cdot 2\text{H}_2\text{O}$, $[\text{Cu}(\text{H}_2\text{O})_2(\text{NH}_3)_2]^{+2} \cdot 2\text{H}_2\text{O}$ y $[\text{Cu}(\text{H}_2\text{O})(\text{NH}_3)_3]^{+2} \cdot 2\text{H}_2\text{O}$. Primeramente las 2 aguas de solvatación se ubicaron de manera que pudieran formar

el mayor número de interacciones de puentes de H, o sea cuatro. Sin embargo, aun así estas moléculas de agua pueden estar ubicadas de manera diferente, dependiendo de que su interacción tipo puente de H sea con las aguas o los NH₃ de coordinación. Esto da lugar a 4 posibles arreglos para las aguas de solvatación en cada uno de estos tres complejos. En todos ellos las interacciones tipo puente de H involucran al O en las aguas de solvatación y a los H en los grupos coordinados al cobre. Además en todos los casos existen 4 de este tipo de interacciones.

En el caso del complejo $[\text{Cu}(\text{H}_2\text{O})_3(\text{NH}_3)]^{+2} \cdot 2\text{H}_2\text{O}$ (figura 29) la estructura de menor energía presenta interacciones en las que los 4 H involucrados pertenecen a las aguas de coordinación. Cuando uno de estos H pertenece al grupo NH₃ la energía relativa se incrementa. Para el complejo $[\text{Cu}(\text{H}_2\text{O})_2(\text{NH}_3)_2]^{+2} \cdot 2\text{H}_2\text{O}$ (figura 30) en la estructura de mínima energía tres de los H involucrados en las interacciones pertenecen a las aguas de coordinación y uno a uno de los grupos NH₃. Cuando un segundo H de NH₃ participa, la energía se incrementa. Finalmente, para el complejo $[\text{Cu}(\text{H}_2\text{O})(\text{NH}_3)_3]^{+2} \cdot 2\text{H}_2\text{O}$ (figura 31) la estructura de mínima energía es aquella en la que de los cuatro H involucrados en las interacciones 2 pertenecen a la molécula de agua coordinada al cobre y los otros dos a los grupos NH₃. En la medida en que el número de H de NH₃ que participa aumenta, también lo hace la energía.

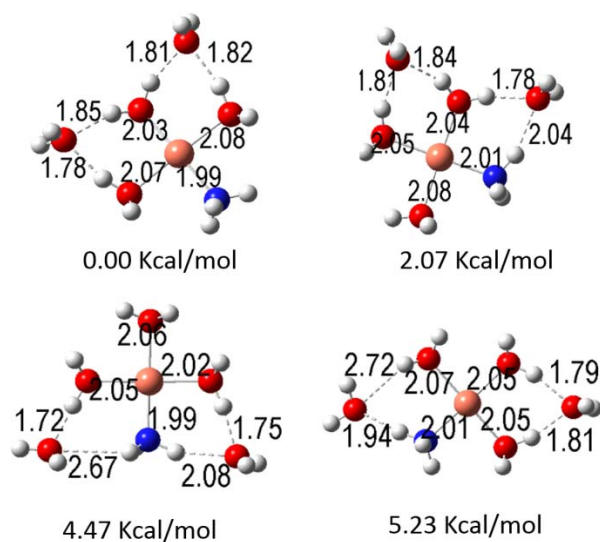


Figura 29. Confórmeros estudiados para el complejo $[\text{Cu}(\text{H}_2\text{O})_3(\text{NH}_3)]^{+2} \cdot 2\text{H}_2\text{O}$ y sus energías relativas con respecto al confórmero más estable.

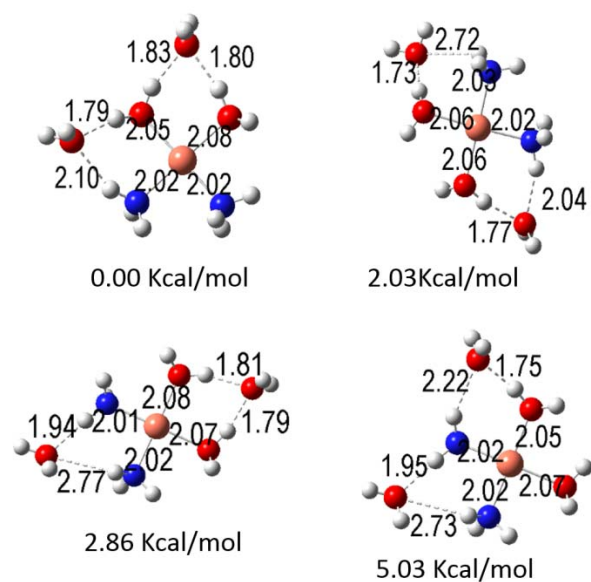


Figura 30. Conformeros estudiados para el complejo $[\text{Cu}(\text{H}_2\text{O})_2(\text{NH}_3)_2]^{+2} \cdot 2\text{H}_2\text{O}$ y sus energías relativas con respecto al conformero más estable.

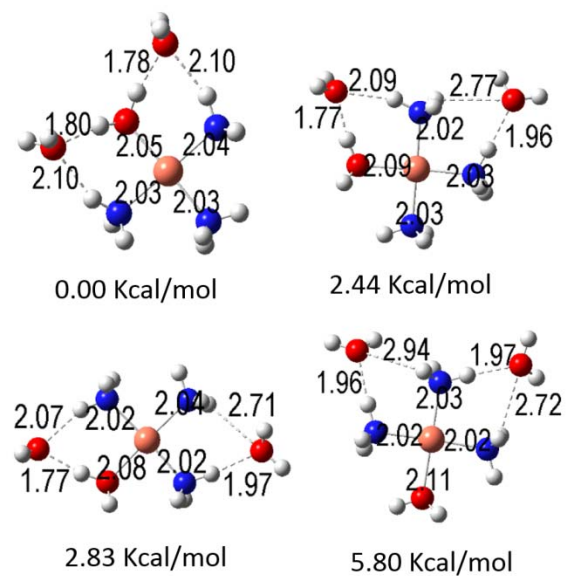


Figura 31. Conformeros estudiados para el complejo $[\text{Cu}(\text{H}_2\text{O})(\text{NH}_3)_3]^{+2} \cdot 2\text{H}_2\text{O}$ y sus energías relativas con respecto al conformero más estable.

Una vez identificadas las estructuras de menor energía para cada complejo, estas fueron las que se utilizaron en el cálculo de las constantes de complejamiento. Además, en base a la información obtenida de este análisis puede concluirse que en todos los casos las estructuras de menor energía son aquella en las que las aguas de solvatación forman el mayor número posible de puentes de H con las aguas de coordinación y el menor número posible con los grupos amonio coordinados al cobre. Con esta información en mente se diseñaron los complejos involucrados en los esquemas *EFS8* y *EFS12*. Sus estructuras se muestran en la figuras 33 y 34.

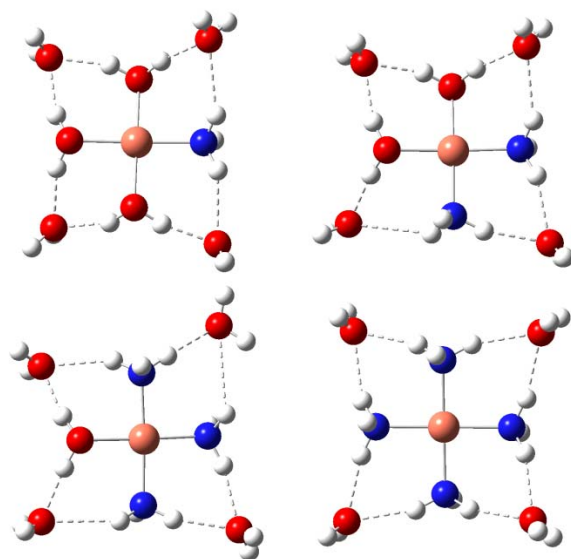


Figura 33. Estructuras de los complejos formados entre Cu(II) y NH₃, esquema de reacción *EFS8*.

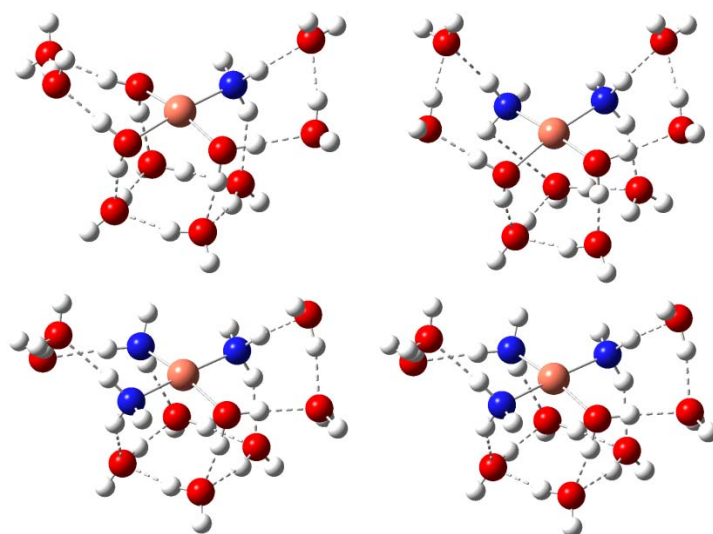


Figura 34. Estructuras de los complejos formados entre Cu(II) y NH₃, esquema de reacción *EFS12*.

Los valores de EAP obtenidos para las reacciones de complejamiento de Cu(II) con amonio usando los diferentes esquemas se reportan en la tabla 27. Se ve claramente que la desviación con respecto a los valores experimentales es altamente dependiente del esquema de reacción empleado y en menor medida del funcional de la densidad empleado. Para estas reacciones lo único que se modifica entre un esquema y otro es el número de moléculas de agua explícitas incluido.

Tabla 27. Valores de EAP para las reacciones de complejamiento de Cu(II) con amonio. Se usaron los equilibrios de formación sucesivos y logK.

	M05	M06	TPSS	PBE	PBE0	BH&HLYP	PW91	BLYP	BMK	B3LYP	EAPG
<i>EFS0</i>	28.72	37.18	41.95	47.40	33.24	25.78	48.89	42.93	34.74	32.91	37.37
<i>EFS4</i>	3.64	4.73	4.69	5.09	4.22	2.56	5.14	4.46	3.57	3.79	4.19
<i>EFS6</i>	2.10	3.17	2.85	3.17	2.52	1.00	3.17	2.65	1.99	2.09	2.47
<i>EFS8</i>	0.63	1.44	1.01	1.31	0.74	0.67	1.26	0.93	0.55	0.43	0.90
<i>EFS12</i>	1.04	2.21	1.69	1.65	1.36	1.50	1.64	1.60	0.95	1.25	1.49

Se encontró que modelar al ión del cobre sin aguas explícitas lleva a grandes errores (valores de EAP mayores a 20) aunque se use un modelo de solvente continuo (SMD en este caso). Esto evidencia que para constantes de complejamiento en solución acuosa modelar el solvente de esta manera no es suficiente para describir adecuadamente los sistemas de interés. El acuerdo con el experimento mejora significativamente al agregar al menos 4 moléculas de agua explícita, que se corresponden a la esfera de coordinación del cobre. Para este modelo los valores de EAP van de 2.56 a 5.14, lo que sigue siendo bastante alto. Disminuyen un poco más al incluir otras dos moléculas de agua (EAP = 1.00 – 3.17).

Los mejores valores se obtuvieron con los esquemas *EFS8* y *EFS12*. Los valores de EAPG para estos esquemas fueron 0.90 y 1.49, respectivamente. En el caso del esquema *EFS8* con todos los funcionales de la densidad utilizados se obtuvieron valores de EAP < 2 y con 6 de ellos EAP < 1. Los mejores resultados en este caso fueron los obtenidos con los funcionales BMK (EAP = 0.55) y B3LYP (EAP = 0.43). Curiosamente al pasar del esquema *EFS8* al esquema *EFS12* los valores de EAP se incrementan, aunque no mucho. Esto se atribuye a

las interacciones que aparecen entre aguas de solvatación, o sea interacciones adicionales a las que involucran al complejo. En este caso todos los funcionales, excepto M06 dan valores de EAP menores a 2 y de ellos $EAP < 1$ solo con BMK (0.95).

Además el esquema *EFS12* tiene 2 desventajas con respecto al esquema *EFS8*: mayor costo computacional y un mayor número de arreglos posibles para las moléculas de agua implícitas. Este último aspecto hace que sea necesaria una búsqueda conformacional exhaustiva (lo que por tiempo fue imposible durante el desarrollo de esta tesis). De modo que considerando los resultados obtenidos se puede decir que la metodología óptima para estudiar los complejos de Cu(II) con amonio es usar el esquema de reacción *EFS8*. Esto usando los funcionales BMK o B3LYP, aunque también puede obtenerse un buen acuerdo con los experimentos usando el mismo esquema y los funcionales M05, BHandHLYP, PBE0 y BLYP.

6.2.4 Complejos de Cu (II) con acetato (CH_3COO^-)

En el caso del ligante acetato el arreglo que maximiza el número de puentes de H (también 4 en cada complejo) es diferente para $[\text{Cu}(\text{H}_2\text{O})_2(\text{CH}_3\text{COO})]^+ \cdot 2\text{H}_2\text{O}$ que para $[\text{Cu}(\text{H}_2\text{O})_2(\text{CH}_3\text{COO})_2] \cdot 2\text{H}_2\text{O}$. Usaremos nuevamente los complejos involucrados en el esquema de reacción *EFS6* para ejemplificar (figura 35). En el complejo con un solo acetato cada una de las aguas de solvatación se ubican de manera que su átomo de oxígeno interactúa con uno de los átomos de H en una de las aguas de coordinación y uno de sus H interactúa con uno de los O en el ligante. En el complejo con dos acetatos, por otra parte, cada una de las aguas de solvatación se ubican de manera que sus átomos de H interactúan con uno de los átomos de H un átomo de O de cada ligante.

Las estructuras correspondientes a los complejos involucrados en los esquemas *EFS8* y *EFS12* se muestran en las figuras 36 y 37, respectivamente. En todos los casos las interacciones tipo puente de H son similares a las descritas en el párrafo anterior, aunque para estos dos esquemas además aparecen interacciones entre aguas de solvatación.

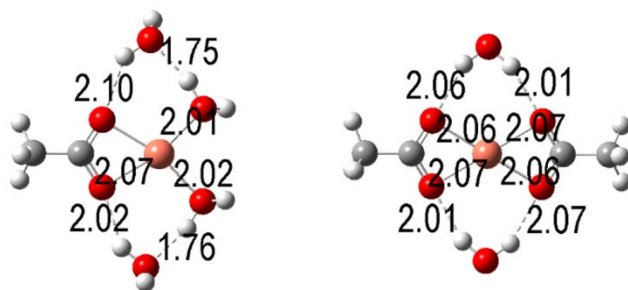


Figura 35. Estructuras de los complejos formados entre Cu(II) y acetatos en el esquema de reacción *EFS6*, $[\text{Cu}(\text{H}_2\text{O})_2(\text{CH}_3\text{COO})]^+ \cdot 2\text{H}_2\text{O}$ y $[\text{Cu}(\text{H}_2\text{O})_2(\text{CH}_3\text{COO})_2] \cdot 2\text{H}_2\text{O}$.

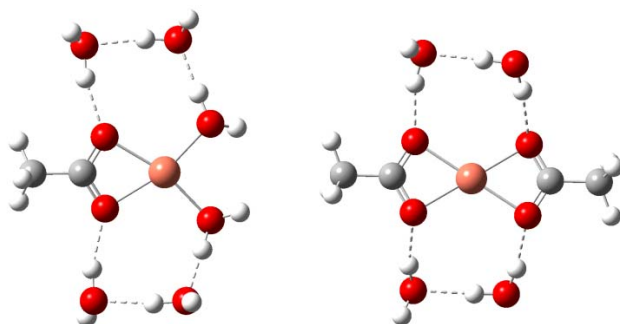


Figura 36. Estructuras de los complejos formados entre Cu(II) y acetatos en el esquema de reacción *EFS8*.

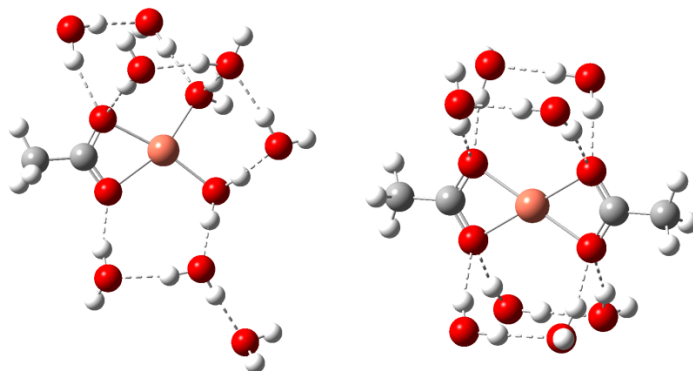


Figura 37. Estructuras de los complejos formados entre Cu(II) y acetatos en el esquema de reacción *EFS12*.

Los valores de EAP obtenidos para las reacciones de complejamiento de Cu(II) con acetatos se muestran en la tabla 28. Al igual que ocurría con el ligante amonio, también en este caso el esquema de reacción empleado tiene mayor influencia en la desviación con respecto a los valores experimentales que el funcional de la densidad. Sin embargo, para las reacciones con

acetato se encontró que hay 3 esquemas de reacción con al menos 1 caso en el que se obtienen valores de EAP < 2 y para NH₃ solo había 2. Estos esquemas son *EFS6*, *EFS8* y *EFS12*; y sus valores de EAPG fueron 2.72, 1.76 y 1.60, respectivamente. Además en el caso de acetatos el esquema *EFS12* da resultados similares (aunque ligeramente mejores) que *EFS8*.

Tabla 28. Valores de EAP para las reacciones de complejamiento de Cu(II) con acetatos. Se usaron los equilibrios de formación sucesivos.

	M05	M06	TPSS	PBE	PBE0	BH&HLYP	PW91	BLYP	BMK	B3LYP	EAPG
<i>EFS0</i>	27.24	30.64	40.61	44.79	30.30	22.28	45.95	41.93	29.07	31.64	34.45
<i>EFS4</i>	5.42	4.66	7.91	7.84	5.14	3.23	7.78	7.66	3.79	5.41	5.88
<i>EFS6</i>	2.85	1.80	4.09	3.92	1.95	0.96	3.74	4.14	1.35	2.36	2.72
<i>EFS8</i>	1.79	1.35	2.86	2.37	1.61	0.62	2.18	2.85	0.75	1.18	1.76
<i>EFS12</i>	0.63	3.82	1.73	1.48	1.18	0.87	1.37	2.04	1.40	1.46	1.60

Las interacciones con el solvente involucran a los H del amonio y a los O de los acetatos. Esto hace que la primera capa de solvatación para este último necesite un menor número de moléculas de agua. Esto parece causar que en el caso de acetatos se obtengan buenos resultados con *EFS6*. Para este esquema se obtuvieron valores de EAP < 2 con 4 funcionales de la densidad y uno de ellos con EAP < 1 (BHandHLYP, EAP = 0.96). En el caso del esquema *EFS8* se obtuvieron valores de EAP < 2 con 6 funcionales de la densidad y 2 de ellos con EAP < 1. Estos fueron BHandHLYP y BMK, con valores de EAP igual 0.62 y 0.75, respectivamente. Para el esquema *EFS12* se obtuvieron valores de EAP < 2 con 8 funcionales de la densidad y 2 de ellos con EAP < 1. Estos fueron M05-2X (0.63) y BHandHLYP (0.87).

En base a los resultados obtenidos puede decirse que la metodología óptima para estudiar los complejos de Cu(II) con acetato es, al igual que para NH₃, usar el esquema de reacción *EFS8* con los funcionales BHandHLYP y BMK, aunque también puede obtenerse un acuerdo aceptable con los experimentos usando el mismo esquema y el funcional B3LYP. Esto es considerando que la mejoría al pasar del esquema *EFS8* al *EFS12* es pequeña pero el incremento de tiempo de computo puede ser grande (dependiendo del sistema).

6.2.5. Generalidades de ambos ligantes mono-funcionales

Aunque el análisis de los dos ligantes presentados anteriormente (amonio y acetato) ya indica que la mejor metodología de las estudiadas para constantes de complejamiento de Cu(II) parece ser usar el esquema de reacción *EFS8* con los funcionales BHandHLYP o BMK, también se analizaron los dos ligantes de manera conjunta. Esto con la idea de obtener datos cuantitativos independientes del tipo de ligante.

Los valores de EAP obtenidos en este caso se muestran en la tabla 29. Como era de esperarse el menor valor de EAPG se obtuvo al usar el esquema *EFS8* (1.33). Con este esquema los valores de EAP son menores a 2, independientemente del funcional utilizado. Además con 3 de ellos se obtienen valores de EAP < 1. Estos son BMK, BHandHLYP y B3LYP. Sus valores de EAP fueron 0.64, 0.65 y 0.81, respectivamente.

Tabla 29. Valores de EAP para las reacciones de complejamiento de Cu(II) con amonio y acetatos, analizados conjuntamente. Se usaron los equilibrios de formación sucesivos.

	M05	M06	TPSS	PBE	PBE0	BH&HLYP	PW91	BLYP	BMK	B3LYP	EAPG
<i>EFS0</i>	27.98	33.91	41.28	46.09	31.77	24.03	47.42	42.43	31.90	32.28	35.91
<i>EFS4</i>	4.53	4.69	6.30	6.47	4.68	2.89	6.46	6.06	3.68	4.60	5.04
<i>EFS6</i>	2.48	2.48	3.47	3.55	2.23	0.98	3.46	3.39	1.67	2.22	2.59
<i>EFS8</i>	1.21	1.40	1.93	1.84	1.18	0.64	1.72	1.89	0.65	0.81	1.33
<i>EFS12</i>	0.83	3.01	1.71	1.57	1.27	1.19	1.50	1.82	1.17	1.36	1.54

Considerando los resultados obtenidos para los dos ligantes mono-funcionales, se escogieron los funcionales de la densidad BMK, BHandHLYP y B3LYP para el estudio de las constantes de complejamiento de Cu con un ligante polifuncional (dopamina). Para ello se incluyeron los esquemas *EFS0*, *EFS4*, *EFS6* y *EFS8*, el esquema *EFS12* ya no se incluyó en esta parte por ser computacionalmente más costoso y haberse demostrado con anterioridad que no necesariamente mejora los resultados obtenidos con *EFS8*.

6.2.6. Complejos de Cu(II) con dopamina

El complejamiento de Cu(II) con dopamina (H_nD^n) es un proceso mucho más complejo que los previamente analizados para NH_3 y acetatos. Considerando la estructura de este

compuesto son posibles tanto complejos mono-dentados como bi-dentados. Los primeros podrían formarse con el grupo amino (sitio de complejamiento N), con el O del grupo catecol que se encuentra en posición *para* con respecto al grupo aminoetil (sitio de complejamiento Op), o con el O del grupo catecol que se encuentra en posición *meta* con respecto al grupo aminoetil (sitio de complejamiento Om). Además puede formarse un complejo bi-dentado que involucre a los dos átomos de oxígeno del grupo catecol (sitio de complejamiento OO).

Adicionalmente, al tener la dopamina 3 sitios ácidos, dependiendo del pH, la especie ácido-base que predomina puede ser diferente: protonada (H_3D^+), neutra (H_2D), mono-aniónica (HD^-), o di-aniónica (D^{2-}). Debido a esto, primeramente hubo que identificar el orden de desprotonación para poder usar las estructuras correctas para cada especie. Estas se muestran en la figura 38.

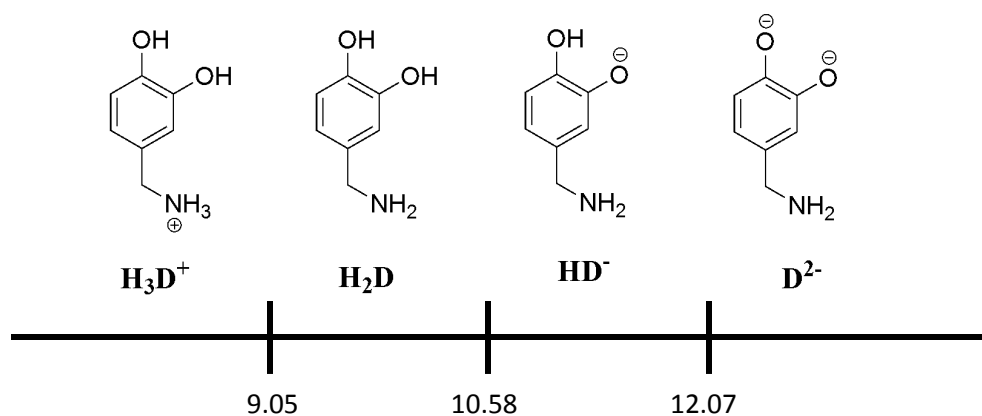


Figura 38. Diagrama de zona de predominio de las estructuras de las diferentes especies ácido-base de la dopamina.

Una vez identificadas sus estructuras hubo que considerar las posibles reacciones para formar complejos de estas cuatro especies, formando complejos por los cuatro tipo de sitios mencionados anteriormente. Esto llevaría en principio a 16 diferentes rutas para formar complejos con el Cu(II) con dopamina. Sin embargo, considerando que cuando el N está protonado este sitio no puede participar y que para la dopamina di-aniónica se sabe que el complejo más probable es el bidentado (sitio OO), el número de rutas posibles puede reducirse a 12. Estas son:

- H_3D^+ , sitios Om, Op, OO

- H₂D, sitios N, Om, Op, OO
- HD⁻, sitios N, Om, Op, OO
- D²⁻, sitio OO

Los datos obtenidos para estas 12 rutas, con los tres funcionales seleccionados en la sección anterior (BHandHLYP, BMK y B3LYP) se reportan en los Anexos 21, 22 y 23, respectivamente. Se encontró que la energía libre de Gibbs más baja se corresponde sistemáticamente a la formación del complejo bidentado (sitio OO). Este fue el caso para todas las combinaciones esquema / funcional de la densidad / especie ácido-base de la dopamina. De modo que se propone que la dopamina se acompleja con Cu(II) como ligante bi-dentado, con el sitio de anclaje involucrando a los dos átomos de O del grupo catecol. Este fue por lo tanto el complejo utilizado para hacer las comparaciones con el único dato disponible en la literatura que es el que corresponde a la dopamina como ligante en su forma D²⁻. Las estructuras usadas en este caso para los diferentes esquemas se muestran en la figura 39. Las estructuras correspondientes a los complejos H₃D⁺, H₂D y HD⁻ se muestran en el Anexo 24.

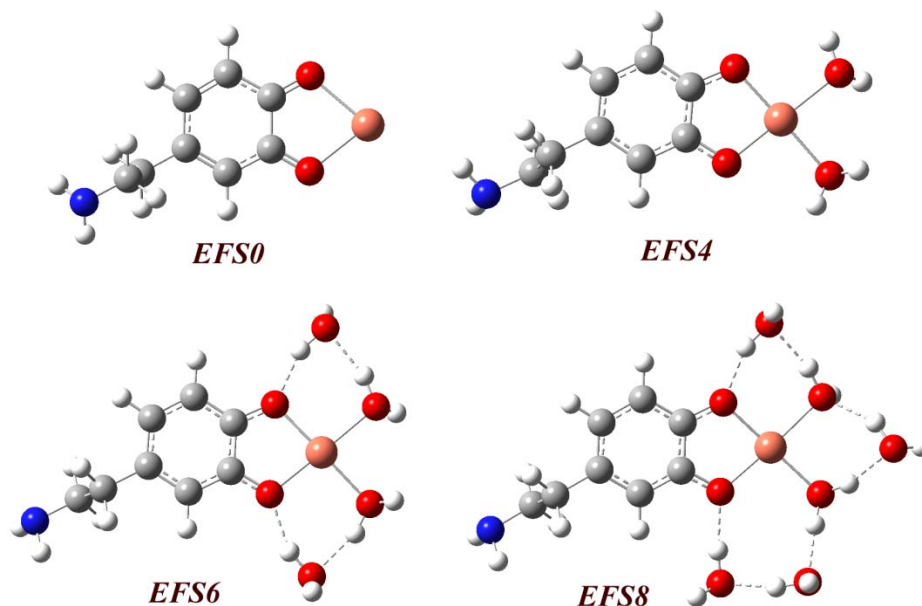


Figura 39. Estructuras del complejo dopamina (D²⁻) con Cu(II) para los diferentes esquemas de reacción.

Los valores de los errores absolutos obtenidos (EA) se reportan en la tabla 30 (en este caso no es posible obtener valores de EAP porque hay un solo valor experimental disponible). De manera análoga a lo que sucedió cuando el ligante era amonio o acetato, en el caso de la dopamina también se encontró que el acuerdo de los valores calculados con el experimental depende marcadamente del número de moléculas de agua explícitas que se incluyan el modelo. El esquema de reacción *EFS8* fue el que dio resultados en mejor acuerdo con el experimento, con valor de EA igual a 0.60 para cuando los cálculos se realizan con el funcional BMK. Además, con el funcional BHandHLYP el valor de EA fue menor a 2 unidades de logK.

Tabla 30. Errores absolutos (EA) para las reacciones de complejos de Cu(II) con dopamina di-aniónica (D^{2-}).

	BH&HLYP	BMK	B3LYP	EAPG
<i>EFS0</i>	9.40	14.95	26.64	17.00
<i>EFS4</i>	5.62	5.83	8.23	6.56
<i>EFS6</i>	3.79	3.70	5.65	4.38
<i>EFS8</i>	1.42	0.60	3.13	1.72

Estos resultados no solo son muy buenos, sino que además son consistentes con los obtenidos para ligantes mono-funcionales. De modo que podemos decir que la hipótesis planteada también se cumple para los equilibrios de formación de complejos. Además se puede decir que usando SMD se disminuye considerablemente el número de moléculas de agua explícitas necesarias para lograr un buen acuerdo con el experimento, en comparación a los resultados obtenidos con COSMO (se necesitaron 18 moléculas de agua).²²

Habiendo identificado una metodología que reproduce adecuadamente los resultados experimentales, esta se empleó para estimar los valores de energías libres de Gibbs y constantes de equilibrio para las especies de dopamina diferentes a D^{2-} . Estos valores se proponen aquí por primera vez. No hay datos previos para estos equilibrios ni experimentales ni teóricos. Considerando que los resultados obtenidos para la dopamina di-aniónica estas estimaciones se hicieron con el esquema *EFS8* y el funcional de la densidad BMK (tabla 31).

En esta tabla también se incluyeron los valores correspondientes a D^{2-} para poder analizar el conjunto completo.

Tabla 31. Energías libres de Gibbs (ΔG , kcal/mol) y valores de $\log K$ para los equilibrios de complejamiento de Cu(II) con dopamina en sus diferentes grados de protonación, obtenidos con el esquema *EFSS* y el funcional de la densidad BMK.

	ΔG	$\log K$
D^{2-}	-28.87	21.16
HD^-	-5.63	4.12
H_2D	9.14	-6.70
H_3D^+	10.30	-7.55

Se encontró que el grado de desprotonación de la dopamina tiene un fuerte efecto en el proceso de complejamiento con Cu(II). Mientras más desprotonada la dopamina, más favorecida la reacción. De hecho, los valores de ΔG se hacen más bajos con el grado de desprotonación y solamente son negativos cuando la dopamina está como especie di- o mono-aniónica. Esto indica que estas son las únicas formas ácido-base en las que puede estar la dopamina para actuar de forma eficiente como ligante de Cu(II).

Es de esperarse que la metodología aquí recomendada (esquema *EFSS*, con SMD y el funcional de la densidad BMK) pueda usarse para calcular otras constantes de complejamiento de las que no haya valores reportados anteriormente. Al menos cuando el metal sea Cu(II). Queda por demostrar si este sería también el caso para otros metales, y/o para otros estados de oxidación del cobre, o si el cambiar el centro metálico modifica en algún aspecto la metodología que da mejor acuerdo con los experimentos.

7- CONCLUSIONES

En este proyecto se propusieron 29 esquemas de reacción para calcular los valores de pKas de tres familias orgánicas diferentes: aminas, fenoles y ácidos carboxílicos. Hacer generalizaciones en cuanto a estrategias de cálculo de constantes de equilibrio en solución acuosa en buen acuerdo con los valores experimentales es una tarea muy difícil, ya que las estrategias con menores errores no son las mismas para diferentes familias de compuestos, ni para diferentes tipos de equilibrios.

Se encontraron numerosos factores que influyen en la calidad de los resultados, siendo los más importantes el esquema de reacción utilizado, la forma en que se modela el solvente y el método de estructura electrónica empleado (en este orden de importancia). En el caso de los esquemas de reacción, aquellos con idéntico número y tipo de iones a ambos lados del equilibrio son los que conducen a mejores resultados.

Para equilibrios de acidez se concluye que utilizando el funcional PBE0 con el esquema **D**, (y SMD para modelar el solvente) pueden obtenerse valores calculados de pKas en acuerdo razonable con los datos experimentales para cualquiera de las familias de compuestos estudiadas (fenoles, ácidos carboxílicos y aminas). Sin embargo, también se pueden hacer recomendaciones específicas dependiendo del tipo de compuesto.

Además se demostró que el acuerdo con los datos experimentales para el esquema **D** es muy dependiente de la selección de HRef, lo es especialmente marcado para aminas poli-funcionales (neurotransmisores, en este trabajo).

El método de las correlaciones también puede usarse para calcular valores de pKa en buen acuerdo con los datos experimentales tanto para fenoles, como para ácidos carboxílicos y aminas. Sin embargo, para lograr esto en el caso particular las aminas es crucial usar un conjunto suficientemente amplio de moléculas a la hora de obtener los parámetros empíricos que se usaran en el cálculo de pKas.

El conjunto de funciones de bases a usar en el cálculo de pKas debe tener funciones difusas, al menos para átomos \neq H, y funciones de polarización tanto para H como para átomos más pesados. Se recomiendan 6-311+G(d,p) o 6-311++G(d,p) para este propósito.

Tanto usando el esquema *D* (conjuntamente con el funcional PBE0 y SMD para modelar el solvente), como con el método de las correlaciones pueden obtenerse valores de pKa con errores (respecto a los valores experimentales) mucho menores que lo que se considera aceptable en la literatura actualmente. En la mayoría de los casos los errores obtenidos con estas metodologías es menor a 1 unidad de pKa, y en muchos casos menor también a 0.5.

Para las constantes de complejamiento el factor que tiene mayor efecto en los resultados obtenidos es también el esquema de reacción empleado, en particular el número de aguas explícitas.

En este caso usar modelos continuos de solvente no es suficiente para describir adecuadamente los sistemas de interés. El número óptimo de moléculas de agua a emplear fue 8, o sea el esquema *EFSS*, cuando se usa además SMD.

La metodología recomendada para equilibrios de formación de complejos con Cu(II) consiste en usar este esquema con el funcional de la densidad BMK y se utilizó para proponer nuevas constantes de equilibrio para el sistema dopamina – Cu(II).

Finalmente, para los diferentes tipos de equilibrio, y para todas las familias de compuestos estudiadas, se cumplió la hipótesis planteada. Fue posible encontrar metodologías que estuvieran en buen acuerdo con los experimentos tanto para especies relativamente sencillas (mono-funcionales) como para moléculas más complejas (poli-funcionales).

8 - PERSPECTIVAS

Extender el estudio del método de las correlaciones, para el cálculo de pKas, a un mayor número de métodos de estructura electrónica, e investigar la influencia de usar conjuntos de moléculas más amplios a la hora de obtener los parámetros *m* y *C₀*.

Probar la posibilidad de emplear en el cálculo de constantes de complejamiento una metodología similar al método de las correlaciones usado para constantes de acidez.

Explorar el efecto del tamaño del conjunto de funciones de base en el cálculo de constantes de complejamiento.

Ampliar el estudio de constantes en la formación de complejos de sistemas con cobre y otros neurotransmisores y también a sistemas con otros metales diferentes al cobre.

9- REFERENCIAS

- ¹ Villuendas-Rey, Y.; Alvarez-Idaboy, J. R.; Galano, A. *J. Chem. Inf. Model.* **2015**, *55*, 2552.
- ² Galano, A.; Alvarez-Idaboy, J. R. *J. Comput. Chem.* **2013**, *34*, 2430.
- ³ León-Carmona, J. R.; Alvarez-Idaboy, J. R.; Galano, A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12534.
- ⁴ Galano, A.; Mazzone, G.; Alvarez-Diduk, R.; Marino, T.; Alvarez-Idaboy, J. R.; Russo, N. *Annu. Rev. Food Sci. Technol.* **2016**, *7*, 335.
- ⁵ Galano, A.; Medina, M. E.; Tan, D. X.; Reiter, R. J. *J. Pineal Res.* **2015**, *58*, 107.
- ⁶ Uauy, R.; Olivarez, M.; Gonzalez M. *Am. J. Clin. Nutr.* **1998**, *67*, 952S.
- ⁷ Brown, T. S.; Wallece P. M. "Psicología Fisiologica", *Mc Graw Hill Mexico* **1989**.
- ⁸ Brady, R. J. "Sistema nervioso", *Limusa* **1991**
- ⁹ Expert Group on Vitamins and Minerals Secretariat, **2002**.
- ¹⁰ Sayre, L. M.; Perry, G.; Smith, M. A. *Chem. Res. Toxicol.* **2008**, *21*, 172.
- ¹¹ Tesis para obtener el grado de Doctor de *María del Rosario Moya Hernandez, UAM-I* **2003**, "Estudio de especiación química de los fármacos antiinflamatorios Tenoxicam y Piroxicam con cationes metálicos de interés biológico".
- ¹² Rickard, G. A.; Gomez-Balderas, R.; Brunelle, P.; Raffa, D. F.; Rauk, A. *J. Phys. Chem.* **2009**, *113*, 9559.
- ¹³ Crichton, R. R.; Ward, R. J. *Met. Ions Biol. Syst.* **2004**, *41*, 185.
- ¹⁴ Wang, T.; Guo Z. *Curr. Med. Chem.* **2006**; *13*, 525.
- ¹⁵ Albert, A.; Serjeant, E. P. *The Determination of Ionization Constants; Chapman and Hall: New York*, **1984**.
- ¹⁶ Dissanayake, D. P.; Senthilnithy, R. *J. Mol. Struct. Theochem* **2009**, *910*, 93.
- ¹⁷ Uudsemaa, M.; Kanger, T.; Lopp, M.; Tamm, T. *Chem. Phys. Lett.* **2010**, *485*, 83.

-
- ¹⁸ Rayne, S.; Forest, K. *J. Mol. Struct. Theochem* **2010**, *949*, 60.
- ¹⁹ Casasnovas, R.; Fernandez, D.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. *Theor. Chem. Acc.* **2011**, *130*, 1.
- ²⁰ Ghalami-Choobar, B.; Dezhampannah, H.; Nikparsa, P.; Ghiami-Shomami, A. *Int J Quantum Chem* **2012**, *112*, 2275.
- ²¹ Ho, J.; Coote, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 295.
- ²² Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. *J. Phys. Rev.* **2009**, *113*, 9559.
- ²³ Claverie, P.; Daudey, J. P.; Langlet, J.; Pullman, B.; Piazzola, D.; Huron, M. J. *J. Phys. Chem.* **1978**, *82*, 405.
- ²⁴ Freitas, L. C. G.; Longo, R. L.; Simas, A. M. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 189.
- ²⁵ Pliego, J. R.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241.
- ²⁶ Zhang, S.; Baker, J.; Pulay, P. *J Phys Chem A* **2010**, *114*, 425.
- ²⁷ Zhang, S.; Baker, J.; Pulay, P. *J Phys Chem A* **2010**, *114*, 432.
- ²⁸ Zhang, S. *J. Comput. Chem.* **2012**, *33*, 517.
- ²⁹ Zhang, S. *J. Comput. Chem.* **2012**, *33*, 2469.
- ³⁰ Matsui, T.; Baba, T.; Kamiya, K.; Shigeta, Y. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4181.
- ³¹ Álvarez-Diduk, R.; Ramírez-Silva, M. T.; Galano, A.; Merkoçi, A. *J. Phys. Chem. B* **2013**, *117*, 12347.
- ³² Baba, T.; Matsui, T.; Kamiya, K.; Nakano, M.; Shigeta, Y. *Int. J. Quantum Chem.* **2014**, *114*, 1128.
- ³³ Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421.
- ³⁴ Dissanayake, D. P.; Senthilnithy, R. *J. Mol. Struct. Theochem* **2009**, *910*, 93.
- ³⁵ Pliego, J. R.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.
- ³⁶ Casasnovas, R.; Fernandez, D.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. *Theor. Chem. Acc.* **2011**, *130*, 1.
- ³⁷ Pliego, J. R. *Chem. Phys. Lett.* **2014**, *118*, 5730.
- ³⁸ Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413.
- ³⁹ Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963.
- ⁴⁰ Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. *J. Phys. Chem. A* **2003**, *107*, 9380.

-
- ⁴¹ Saracino, G. A. A.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, 373, 411.
- ⁴² Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., *J. Phys. Chem. A* **2004**, 108, 166.
- ⁴³ Magill, A. M.; Cavell, K. J.; Yates, B. F. *J. Am. Chem. Soc.* **2004**, 126, 8717.
- ⁴⁴ Alvarado-González, M.; Orrantia-Borunda, E.; Glossman-Mitnik, D. *J. Mol. Struct. Theochem* **2008**, 869, 105.
- ⁴⁵ Namazian, M.; Zakery, M.; Noorbala, M. R.; Coote, M. L. *Chem. Phys. Lett.* **2008**, 451, 163.
- ⁴⁶ Brown, T. N.; Mora-Diez, N. *J. Phys. Chem. B* **2006**, 110, 9270.
- ⁴⁷ Vianello, R.; Maksic, Z. B. *Eur. J. Org. Chem.* **2005**, 16, 3571.
- ⁴⁸ Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, 123, 7314-7319.
- ⁴⁹ Camaioni, D. M.; Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, 109, 10795.
- ⁵⁰ Pliego, J. R. *Chem. Phys. Lett.* **2003**, 367, 145.
- ⁵¹ Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A. III *J. Phys. Chem. B* **2008**, 112, 9709.
- ⁵² Hohenberg, P.; Kohn, W. *Phys. Rev.*, **1964**, 136, B864.
- ⁵³ Kohn, W.; Sham, L. *J. Phys. Rev.*, **1965**, 140, A1133.
- ⁵⁴ Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.*, **2003**, 91, 146401.
- ⁵⁵ Kurth, S.; Perdew, J. P.; Blaha, P. *Int. J. Quant. Chem.*, **1999**, 75, 889.
- ⁵⁶ Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *J. Chem. Phys.*, **2005**, 123, 062201.
- ⁵⁷ Becke, A. D. *J. Chem. Phys.* **1992**, 98, 1372
- ⁵⁸ Becke, A. D. *J. Chem. Phys.* **1993**, 98, 1372. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, 37, 785.
- ⁵⁹ Becke, A. D. *J. Chem. Phys.* **1997**, 107, 8554.
- ⁶⁰ Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B.* **1996**, 54, 16533.
- ⁶¹ Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77, 3865.
- ⁶² Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, 110, 6158.
- ⁶³ Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, 91, 146401.
- ⁶⁴ Boese, A. D.; Martín, J. M. L. *J. Chem. Phys.* **1993**, 121, 3405.
- ⁶⁵ Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.*, **2006**, 2, 364
- ⁶⁶ (a) Cancès M. T.; Mennucci B.; Tomasi J. *J. Chem. Phys.* **1997**, 107, 3032. (b) Mennucci B.; Tomasi J. *J. Chem. Phys.* **1997**, 106, 5151. (c) Mennucci B.; Cancès E.; Tomasi J. *J. Phys. Chem.*

B1997, 101, 10506. (d) Tomasi J.; Mennucci B.; Cancès E. *J. Mol. Struct (TEOCHEM)*. **1999**, 464, 211.

⁶⁷ Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, 107, 3210.

⁶⁸ *Gaussian 03, Revision E.01*; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian, Inc.: Wallingford, CT*, 2003.

⁶⁹ Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B*, **2009**, 113, 6378.

⁷⁰ *Gaussian 09, Revision A.08*, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R. E., Stratmann, O., Yazyev, A. J., Austin, R., Cammi, C., Pomelli, J. W., Ochterski, R., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J., *Gaussian, Inc., Wallingford CT*, 2009.

⁷¹ Gail A. Rickard, Rodolfo Gomez-Balderas, Patrick Brunelle, Duilio F. Raffa, and Arvi Rauk *J. Phys. Chem. A* **2005**, 109, 8361-8370

⁷² Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, 123, 161103.

⁷³ Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 120, 215.

⁷⁴ Cotton y Wilkinson *Química Inorgánica Avanzada* **2006**

⁷⁵ Ortega-Castro, J.; Adrover, M.; Frau, J.; Donoso, J.; Muñoz, F.; *Chem. Phys. Lett.* **2009**, 277.

ANEXOS

Anexo 1

Valores experimentales de pKa para las moléculas estudiadas

Fenoles	pKa	Ácidos	pKa	Aminas	pKa
4-metilfenol	10.19	acético	4.76	amonio	9.25
3-metoxifenol	9.65	fórmico	3.77	metil amina	10.64
3-hydroxibenzaldehido	9.24	láctico	3.86	etil amina	10.63
4-metiltiofenol	9.53	pirúvico	2.5	dimetil amina	10.64
2-4dimetilfenol	10.61	crotónico	4.69	trimetil amina	9.76
3-cianofenol	8.58	glicérico	3.55	alilmetil amina	10.11
		cloro acético	2.86	dimetilsecbutil amina	10.4
		glicólico	3.82	piperidina	11.22
		benzoico	4.2	N-alil piperidina	9.69
		protocatequico	4.48	benzil amina	9.34
paracetamol	9.5	naproxeno	4.2	dopamina	8.97
profadol	10.27	aspirina	3.45	epinefrina	8.75
tapentadol	10.09	ketoprofeno	4.5	serotonina	9.83
ketobemidona	9.96	ibuprofeno	4.4	norepinefrina	8.67

Anexo 2

Valores experimentales de constantes de acomplejamiento con Cu(II) para las moléculas estudiadas

Fenoles	log K
NH ₃ (K1)	3.99
NH ₃ (K1)	7.33
NH ₃ (K1)	10.06
NH ₃ (K1)	12.03
acetato (K1)	2.23
acetato (K1)	3.63
Dopamina dianiónica (K1)	20.56

Anexo 3

Errores absolutos promedio para fenoles mono-funcionales con ciclos termodinámicos y PCM.

	B3LYP	M05-2X	BLYP	TPSS	BHandHLYP	BMK	PBE	PBE0	PW91	MP2
A	1.46	2.04	1.72	1.22	4.92	1.96	1.63	2.35	1.49	2.74
B	4.45	6.63	4.96	2.41	6.21	4.94	1.38	4.15	1.59	4.58
C1	7.61	8.69	8.99	5.54	8.64	8.15	4.13	6.68	3.95	7.24
C2	12.34	12.31	14.58	10.70	13.06	13.81	8.57	10.86	8.24	11.00
C3	20.63	18.20	23.37	18.93	20.01	21.04	16.15	18.28	15.20	17.17
D	1.17	1.16	1.24	1.14	1.25	1.36	1.14	1.21	1.17	0.80
E1	1.20	2.14	1.15	1.08	1.20	2.46	1.31	1.08	2.16	2.62
E2	4.30	1.84	5.22	5.24	3.91	3.38	5.61	4.09	6.46	1.13
E3	12.59	7.55	14.00	13.48	10.86	10.61	13.19	11.51	13.41	7.31
F1	6.85	5.97	6.96	5.38	9.41	8.59	2.68	5.33	1.54	8.60
F2	10.01	8.02	10.98	8.51	11.83	11.80	5.45	7.85	3.89	11.25
F3	14.73	11.65	16.58	13.67	16.26	17.47	9.89	12.03	8.19	15.02
F4	23.03	17.53	25.37	21.90	23.21	24.70	17.46	19.46	15.14	21.19
CN1	4.66	6.02	5.49	2.58	5.90	4.93	1.67	4.28	1.81	5.25
CN2	9.39	9.65	11.09	7.74	10.33	10.60	6.01	8.46	6.04	9.01
CN3	17.67	15.50	19.87	15.96	17.27	17.82	13.58	15.87	12.99	15.18
FN1	7.06	5.35	7.49	5.55	9.10	8.59	2.89	5.45	1.78	9.27
FN2	11.78	8.97	13.09	10.70	13.52	14.25	7.32	9.63	5.98	13.02
FN3	20.07	14.86	21.87	18.93	20.47	21.48	14.90	17.05	12.94	19.19
EN1	3.38	4.63	3.88	2.88	3.25	5.50	1.52	2.49	1.12	4.62
EN2	1.45	1.10	1.77	2.29	1.22	0.86	3.05	1.71	4.26	0.85
EN3	9.64	4.88	10.51	10.51	8.12	7.39	10.62	9.11	11.21	5.31

Anexo 4

Errores absolutos promedio para fenoles mono-funcionales con SMD.

	B3LYP	BLYP	BMK	BH	M052X	PBE	PBO	PW91	TPSS	MP2
A	3.90	2.23	3.21	5.81	3.00	2.13	4.01	2.32	4.01	5.60
B	15.75	14.91	15.30	16.88	15.41	13.79	14.65	14.02	14.43	16.20
C1	17.02	16.44	17.38	18.06	16.15	14.22	15.29	14.05	15.19	16.92
C2	18.42	18.00	19.73	19.16	17.21	15.00	16.00	14.52	16.32	17.41
C3	21.10	20.67	23.19	21.30	18.70	16.58	17.99	15.99	18.70	18.15
D	0.93	0.88	0.92	0.94	0.89	0.86	1.24	0.89	0.87	0.79
E1	1.65	1.45	4.19	2.62	2.18	0.99	1.74	1.40	0.94	3.11
E2	0.70	0.57	1.85	1.52	1.12	1.54	1.21	1.87	0.72	2.62
E3	2.43	2.77	1.97	0.72	0.85	3.12	2.12	3.34	2.56	1.88
F1	12.48	11.87	13.96	14.14	11.03	7.42	10.02	6.53	9.79	13.79
F2	13.75	13.40	16.04	15.32	11.76	7.85	10.65	6.56	10.55	14.52
F3	15.15	14.95	18.39	16.42	12.82	8.63	11.36	7.04	11.63	15.01
F4	17.83	17.62	21.85	18.56	14.31	10.21	14.05	8.50	14.06	15.75
CN1	14.09	13.19	13.98	15.36	13.59	12.03	12.91	12.29	12.43	14.77
CN2	15.49	14.75	16.32	16.46	14.64	12.81	13.62	12.77	13.52	15.26
CN3	18.17	17.41	19.78	18.60	16.14	14.39	15.61	14.23	15.94	14.49
FN1	10.82	10.15	12.63	12.62	9.20	5.66	8.27	4.81	7.80	12.37
FN2	12.22	11.70	14.98	13.72	10.26	6.44	8.98	5.28	8.88	12.85
FN3	14.91	14.37	18.44	29.85	11.75	8.02	10.98	6.74	11.30	13.60
EN1	4.58	4.70	7.60	5.32	4.74	1.43	4.12	0.83	3.70	5.26
EN2	3.18	3.15	5.25	4.22	3.69	0.67	3.41	0.69	2.62	4.78
EN3	1.17	0.82	1.79	2.08	2.24	1.15	1.69	1.58	0.61	4.03

Anexo 5

Errores absolutos promedio para fenoles poli-funcionales con ciclos termodinámicos y PCM.

	B3LYP	M05-2X	BLYP	TPSS	BHandHLYP	BMK	PBE	PBE0	PW91
A	1.46	1.85	1.64	0.94	4.81	2.14	1.51	2.34	1.44
B	4.46	6.44	4.80	2.19	6.10	5.12	1.04	4.14	1.20
C1	7.74	9.47	8.72	5.44	8.92	9.08	3.81	6.64	3.60
C2	11.94	12.90	13.61	9.82	12.60	13.39	7.69	10.49	7.08
C3	20.25	18.16	22.84	18.33	19.72	21.49	15.53	17.89	14.44
D	0.45	0.32	0.49	0.55	0.52	0.45	0.57	0.43	0.59
E1	0.34	1.21	0.64	0.43	0.23	1.35	0.91	0.29	1.81
E2	3.90	2.26	4.24	4.36	3.44	2.96	4.73	3.72	5.29
E3	12.21	7.51	13.47	12.88	10.57	11.06	12.57	11.12	12.66
F1	6.86	5.77	6.80	5.16	9.30	8.78	2.36	5.31	1.14
F2	10.13	8.80	10.72	8.40	12.11	12.73	5.12	7.81	3.54
F3	14.34	12.24	15.61	12.79	15.79	17.05	9.01	11.66	7.03
F4	22.65	17.49	24.84	21.30	22.92	25.14	16.85	19.06	14.39
CN1	4.84	6.85	5.32	2.50	6.44	6.09	1.42	4.31	1.46
CN2	9.05	10.29	10.21	6.88	10.12	10.41	5.31	8.17	4.95
CN3	17.35	15.54	19.44	15.39	17.24	18.50	13.14	15.56	12.31
FN1	7.24	6.19	7.32	5.47	9.64	9.75	2.74	5.49	1.41
FN2	11.44	9.62	12.21	9.85	13.31	14.06	6.62	9.33	4.89
FN3	19.75	14.87	21.44	18.36	20.44	22.16	14.46	16.74	12.26
EN1	3.20	3.79	4.04	2.95	2.71	4.34	1.54	2.45	0.32
EN2	1.01	0.89	0.85	1.43	0.97	0.34	2.35	1.40	3.17
EN3	9.31	4.89	10.07	9.94	8.09	8.07	10.18	8.79	10.52

Anexo 6

Errores absolutos promedio para fenoles poli-funcionales en solución junto con SMD.

	B3LYP	BLYP	BH	BMK	M052X	PBE	PBO	PW91	TPSS
A	4.45	2.90	6.57	3.84	3.45	2.74	5.10	2.92	4.57
B	16.30	15.59	17.64	15.94	15.86	14.40	15.74	14.61	14.99
C1	17.22	16.79	18.90	18.33	16.67	14.71	16.47	14.53	16.06
C2	18.91	18.47	20.24	20.19	17.93	15.30	17.24	14.76	16.88
C3	20.30	20.24	21.78	22.74	19.11	16.61	18.94	15.52	18.48
D	0.27	0.18	0.59	0.29	0.42	0.44	0.23	0.58	0.44
E1	1.45	1.10	1.78	3.24	1.66	1.25	0.80	1.88	0.38
E2	0.47	0.67	0.68	1.39	0.53	1.84	0.37	2.11	0.74
E3	1.64	2.34	1.10	1.30	1.21	3.15	1.91	2.87	2.34
F1	13.03	12.54	14.90	14.60	11.48	8.03	11.11	7.13	10.35
F2	13.95	13.75	16.15	16.99	12.28	8.34	11.83	7.04	11.43
F3	15.65	15.42	17.50	18.85	13.54	8.93	12.61	7.27	12.24
F4	17.04	17.19	19.03	21.40	14.72	10.24	14.30	8.03	13.84
CN1	14.70	13.85	16.11	15.30	14.26	12.63	13.99	12.98	13.51
CN2	16.39	15.52	17.46	17.15	15.52	13.22	14.77	13.21	14.32
CN3	17.78	17.29	18.99	19.71	16.71	14.53	16.46	13.97	15.92
FN1	11.43	10.80	13.37	13.96	9.87	6.26	9.35	5.50	8.87
FN2	13.12	12.48	14.71	15.81	11.13	6.85	10.13	5.72	9.68
FN3	14.51	14.25	16.25	18.37	12.32	8.16	11.82	6.48	11.28
EN1	3.97	4.05	4.57	6.27	4.07	0.83	3.04	0.40	2.63
EN2	2.28	2.37	3.22	4.42	2.81	0.24	2.26	0.56	1.81
EN3	0.89	1.91	1.69	1.86	9.96	1.07	1.00	1.32	0.60

Anexo 7

Errores absolutos promedio del conjunto completo de moléculas fenólicas con ciclos termodinámicos y PCM.

	B3LYP	M05-2X	BLYP	TPSS	BHandHLYP	BMK	PBE	PBE0	PW91
A	1.46	1.97	1.69	1.11	4.88	2.03	1.58	2.35	1.47
B	4.45	6.55	4.89	2.33	6.17	5.01	1.24	4.14	1.43
C1	7.67	9.00	8.88	5.50	8.75	8.52	4.01	6.67	3.81
C2	12.18	12.55	14.19	10.34	12.88	13.65	8.22	10.71	7.78
C3	20.48	18.18	23.16	18.69	19.90	21.22	15.90	18.12	14.90
D	0.88	0.82	0.94	0.90	0.96	1.00	0.91	0.90	0.94
E1	0.86	1.77	0.95	0.82	0.82	2.02	1.15	0.77	2.02
E2	4.14	2.01	4.83	4.89	3.72	3.21	5.26	3.94	5.99
E3	12.44	7.54	13.79	13.24	10.74	10.79	12.94	11.35	13.11
F1	6.85	5.89	6.89	5.30	9.36	8.67	2.55	5.32	1.38
F2	10.06	8.33	10.87	8.46	11.94	12.17	5.32	7.84	3.75
F3	14.58	11.88	16.19	13.31	16.07	17.30	9.54	11.88	7.72
F4	22.88	17.52	25.16	21.66	23.09	24.88	17.22	19.30	14.84
CN1	4.73	6.35	5.42	2.55	6.12	5.40	1.57	4.29	1.67
CN2	9.25	9.90	10.74	7.40	10.25	10.52	5.73	8.34	5.61
CN3	17.54	15.52	19.70	15.73	17.26	18.09	13.40	15.75	12.72
FN1	7.13	5.69	7.42	5.52	9.31	9.05	2.83	5.47	1.63
FN2	11.64	9.23	12.73	10.36	13.43	14.18	7.04	9.51	5.55
FN3	19.94	14.87	21.70	18.70	20.46	21.75	14.72	16.92	12.66
EN1	3.31	4.29	3.94	2.91	3.04	5.04	1.53	2.47	0.80
EN2	1.27	1.01	1.40	1.94	1.12	0.65	2.77	1.59	3.82
EN3	9.51	4.88	10.33	10.28	8.11	7.66	10.44	8.98	10.93

Anexo 8

Errores absolutos promedio del conjunto completo las moléculas fenólicas en solución junto con SMD.

	B3LYP	BLYP	BHandHLYP	BMK	M052X	PBE	PBE0	PW91	TPSS
A	4.12	3.01	6.11	3.52	3.18	2.37	4.44	2.56	4.23
B	15.97	15.44	17.19	15.57	15.59	14.03	15.09	14.25	14.65
C1	17.10	16.72	18.39	17.61	16.36	14.42	15.76	14.24	15.54
C2	18.62	18.28	19.59	19.51	17.50	15.12	16.49	14.62	16.54
C3	20.78	20.86	21.49	22.60	18.86	16.59	18.37	15.80	18.61
D	0.66	0.61	0.80	0.74	0.70	0.69	0.84	0.77	0.41
E1	1.57	1.41	2.28	3.38	1.97	1.09	1.36	1.59	0.43
E2	0.61	0.73	1.18	1.48	0.89	1.66	0.87	1.97	0.60
E3	2.11	2.73	0.87	1.87	0.99	3.13	2.03	3.15	2.48
F1	12.70	12.33	14.45	13.84	11.21	7.67	10.45	6.77	10.01
F2	13.83	13.60	15.65	15.89	11.97	8.05	11.12	6.76	10.90
F3	15.35	15.17	16.85	17.79	13.11	8.75	11.86	7.13	11.88
F4	17.51	17.75	18.75	20.87	14.48	10.22	14.15	8.31	13.97
CN1	14.33	13.69	15.66	14.53	13.86	12.27	13.34	12.57	12.86
CN2	15.85	15.25	16.86	16.43	15.00	12.97	14.08	12.94	13.84
CN3	18.02	17.83	18.76	19.52	16.36	14.45	15.95	14.13	15.93
FN1	11.07	10.57	12.92	12.81	9.47	5.90	8.71	5.08	8.22
FN2	12.58	12.14	14.12	14.71	10.61	6.61	9.44	5.46	9.20
FN3	14.75	14.72	24.41	17.79	11.98	8.08	11.32	6.64	11.29
EN1	4.33	4.44	5.02	6.46	4.47	1.19	3.69	0.66	2.60
EN2	2.82	2.88	3.82	4.56	3.33	0.50	2.95	0.64	1.78
EN3	1.06	1.23	1.92	1.78	9.53	1.12	1.41	1.48	0.57

Anexo 9

Errores absolutos promedio para ácidos carboxílicos mono-funcionales con ciclos termodinámicos y PCM.

	B3LYP	BLYP	BMK	BHandHLYP	M052X	PBE	PBE0	PW91	TPSS	MP2
A	1.00	2.98	0.89	1.40	0.74	2.23	1.02	2.70	0.96	0.86
B	2.33	3.46	2.11	2.60	3.85	0.94	1.93	1.42	1.14	2.65
C1	9.31	10.33	8.63	8.42	8.68	6.08	7.74	5.30	7.19	8.11
C2	11.83	14.18	11.85	11.04	13.08	7.98	10.26	8.28	10.21	11.47
C3	17.66	20.60	17.53	16.79	16.22	13.72	15.56	12.74	16.20	16.42
D	1.73	1.17	1.47	1.41	1.61	1.54	1.52	1.88	1.82	3.77
E1	3.22	3.19	5.97	4.62	6.13	1.10	3.19	1.65	2.43	5.92
E2	0.74	1.07	2.74	2.28	1.73	1.48	0.72	1.66	0.70	2.56
E3	5.20	7.08	2.94	3.47	1.50	6.49	4.63	6.40	6.58	2.39
F1	8.64	9.62	10.05	9.91	7.35	5.64	6.84	23.35	7.87	10.83
F2	15.62	16.50	16.45	15.78	12.18	11.56	13.08	24.01	14.32	16.29
F3	18.14	20.34	19.67	18.40	16.58	13.47	15.60	13.32	17.34	19.65
F4	23.81	26.77	25.22	24.47	19.75	19.20	20.90	17.38	23.60	24.47
CN1	3.68	4.05	3.57	3.79	4.18	1.18	3.63	1.33	1.67	2.85
CN2	6.21	7.89	6.60	6.41	8.59	3.26	5.60	3.44	4.69	6.32
CN3	12.03	14.32	12.28	12.15	11.72	8.82	10.58	8.20	10.68	11.27
FN1	9.99	10.21	11.19	11.15	7.68	6.66	8.10	6.30	8.81	11.14
FN2	12.51	14.05	14.42	13.77	12.08	8.56	10.61	8.47	11.83	14.50
FN3	18.34	20.48	20.10	19.51	15.22	14.30	15.92	13.24	17.82	19.45
EN1	8.78	9.48	11.32	9.53	10.63	5.95	7.14	5.61	7.94	11.07
EN2	6.25	5.64	8.00	6.91	6.22	4.04	5.66	3.44	4.92	7.71
EN3	0.97	1.06	2.31	1.16	3.09	1.69	1.53	2.35	1.13	2.76

Anexo 10

Errores absolutos promedio para ácidos carboxílicos mono-funcionales con SMD.

	B3LYP	BLYP	BMK	BHandHLYP	M052X	PBE	PBE0	PW91	TPSS	MP2
A	1.17	1.29	0.93	1.74	0.99	1.16	2.15	1.14	1.68	2.05
B	12.29	12.06	11.95	12.79	11.80	11.50	11.84	11.50	11.68	12.57
C1	13.90	13.11	14.43	14.71	12.92	12.22	12.80	12.01	13.25	13.99
C2	16.71	16.85	17.34	17.03	14.71	13.87	14.86	13.09	15.51	16.48
C3	19.29	19.83	20.85	19.31	17.01	16.03	17.29	15.05	18.12	18.18
D	1.84	1.80	1.48	1.54	1.97	1.51	0.97	1.87	1.69	2.11
E1	4.77	3.99	7.14	5.96	5.41	1.48	4.23	1.24	2.89	6.04
E2	1.96	1.28	4.23	3.65	5.70	0.99	2.17	0.91	0.94	3.55
E3	0.64	1.93	1.01	1.42	1.44	2.57	0.50	2.40	1.98	1.86
F1	9.02	9.02	10.61	10.04	7.41	5.13	7.20	4.01	7.04	10.17
F2	10.63	10.86	13.09	11.97	8.53	5.85	8.17	4.52	8.61	11.59
F3	13.44	13.81	16.00	14.29	10.32	7.50	10.22	5.60	10.88	14.08
F4	16.02	16.78	19.51	16.57	12.62	9.66	12.65	7.56	13.48	15.77
CN1	12.18	11.95	12.39	13.14	11.78	11.36	11.58	11.53	11.84	12.47
CN2	14.99	14.90	15.31	15.46	13.57	13.01	13.64	12.61	14.11	14.96
CN3	17.56	17.88	18.81	17.74	15.87	15.17	16.07	14.57	16.71	11.81
FN1	8.91	8.90	11.05	10.40	7.39	4.99	6.95	4.04	7.20	10.07
FN2	11.72	11.85	13.97	12.72	9.18	6.64	9.00	5.12	9.47	12.55
FN3	14.29	14.83	17.47	15.00	11.48	8.80	11.43	7.08	12.07	14.25
EN1	6.49	5.95	9.18	7.53	6.55	2.09	5.45	1.45	4.30	7.56
EN2	3.68	3.00	6.27	5.22	4.77	0.83	3.39	0.89	2.03	5.08
EN3	1.24	0.55	2.76	2.93	3.59	1.71	1.14	1.92	0.69	3.38

Anexo 11

Errores absolutos promedio para ácidos carboxílicos poli-funcionales con ciclos termodinámicos y PCM.

	B3LYP	BLYP	BHandHLYP	BMK	M052X	PBE	PBE0	PW91	TPSS
A	2.11	1.54	4.62	2.17	1.86	1.36	2.91	1.37	1.87
B	5.10	6.06	5.91	5.15	6.45	2.36	4.71	2.43	3.18
C1	10.24	11.99	10.38	10.31	10.24	7.04	9.10	6.78	8.50
C2	11.71	14.67	11.87	12.60	11.47	8.55	10.55	7.72	10.41
C3	16.01	19.16	15.13	17.24	14.41	18.28	14.12	10.97	14.79
D	1.58	1.29	1.91	1.10	1.07	1.15	1.22	1.13	1.12
E1	2.20	2.63	1.23	0.94	1.06	4.08	2.34	5.00	3.05
E2	3.67	5.31	2.71	2.17	0.88	5.59	3.79	5.93	4.96
E3	7.97	9.80	5.98	6.81	3.76	9.23	7.36	9.19	9.34
F1	7.50	8.06	9.10	8.80	5.78	3.68	5.88	2.37	6.15
F2	12.63	13.99	13.57	13.96	9.57	8.35	10.27	6.72	11.46
F3	14.11	16.67	15.06	16.26	10.80	9.87	11.73	7.67	13.38
F4	18.41	21.16	18.33	20.90	13.74	13.51	15.30	10.92	17.76
CN1	5.72	6.84	6.50	6.06	7.10	3.31	5.13	3.46	4.19
CN2	7.19	9.52	7.99	8.36	8.33	4.83	6.59	4.41	6.11
CN3	11.48	14.01	11.25	13.00	11.26	8.46	10.15	7.65	10.48
FN1	8.11	8.84	9.70	9.72	6.43	4.62	6.30	3.41	7.16
FN2	9.58	11.52	11.18	12.01	7.66	6.14	7.75	4.35	9.07
FN3	13.88	16.00	14.45	16.65	10.60	9.78	11.32	7.60	13.45
EN1	2.32	2.53	2.65	4.37	3.55	1.09	1.64	1.68	1.46
EN2	1.87	1.33	1.57	2.07	2.32	1.87	0.65	2.62	1.04
EN3	3.44	4.64	2.10	2.57	0.61	5.50	3.38	5.86	5.03

Anexo 12

Errores absolutos promedio para ácidos carboxílicos poli-funcionales con SMD.

	B3LYP	BLYP	BHandHLYP	BMK	M052X	PBE	PBE0	PW91	TPSS
A	1.47	1.27	2.35	0.98	0.98	1.15	2.28	1.17	2.04
B	12.86	12.72	13.40	12.38	12.36	11.96	12.31	12.01	12.19
C1	14.52	14.12	15.26	14.82	13.58	12.71	13.38	12.48	13.88
C2	17.23	17.57	17.65	17.75	15.22	14.44	15.37	13.56	16.02
C3	19.89	20.63	19.96	21.11	17.48	16.48	17.73	15.46	18.65
D	1.34	1.39	1.22	1.18	1.54	1.90	0.92	1.39	1.29
E1	4.14	3.27	5.42	6.75	4.75	1.06	3.65	1.05	2.38
E2	1.44	1.17	3.02	3.83	4.44	1.35	1.67	1.21	0.88
E3	1.80	2.74	1.12	0.84	1.07	3.02	0.85	2.81	2.51
F1	9.59	9.67	10.66	11.04	7.97	5.59	7.68	4.52	7.55
F2	11.26	11.58	12.52	13.48	9.19	6.34	8.74	5.00	9.24
F3	13.96	14.52	14.91	16.41	10.83	8.07	10.73	6.07	11.38
F4	16.63	17.59	17.22	19.77	13.09	10.11	13.09	7.98	14.01
CN1	12.71	12.52	13.43	12.70	12.40	11.75	12.15	11.89	12.36
CN2	15.41	15.46	15.83	15.63	14.05	13.47	14.13	12.97	14.49
CN3	18.08	18.52	18.14	18.99	16.31	15.51	16.49	14.87	17.12
FN1	9.44	9.48	10.69	11.36	8.01	5.38	7.51	4.41	7.72
FN2	12.15	12.42	13.09	14.29	9.66	7.10	9.50	5.48	9.85
FN3	14.81	15.48	15.40	17.65	11.92	9.14	11.86	7.38	12.48
EN1	5.96	5.37	7.24	8.87	5.93	1.71	4.88	1.08	3.78
EN2	3.25	2.43	4.84	5.94	4.28	0.84	2.89	0.92	1.65
EN3	0.93	0.99	2.53	2.58	3.93	2.05	0.87	2.22	1.06

Anexo 13

Errores absolutos promedio para el conjunto completo de ácidos carboxílicos con ciclos termodinámicos y PCM.

	B3LYP	BLYP	BHandHLYP	BMK	M052X	PBE	PBE0	PW91	TPSS
A	1.32	2.40	2.44	1.24	1.03	1.83	1.64	2.14	1.26
B	3.25	4.37	3.67	3.08	4.67	1.45	2.87	1.78	1.85
C1	9.67	11.00	9.12	9.19	9.21	6.51	8.27	5.91	7.71
C2	11.91	14.51	11.42	12.22	12.77	8.35	10.47	8.23	10.43
C3	17.37	20.44	16.46	17.62	15.76	13.56	15.34	12.44	15.95
D	1.65	1.48	1.53	1.28	1.36	1.41	1.40	1.64	1.58
E1	2.81	2.84	3.53	4.45	4.60	1.95	2.81	2.63	2.46
E2	1.61	2.44	2.26	2.43	1.39	2.81	1.62	3.04	2.07
E3	6.20	8.10	4.34	4.22	2.20	7.55	5.60	7.43	7.52
F1	8.46	9.35	9.81	9.79	6.98	5.24	6.74	16.15	7.53
F2	14.87	15.97	15.28	15.82	11.52	10.80	12.42	18.21	13.65
F3	17.12	19.48	17.59	18.85	15.08	12.64	14.62	11.74	16.37
F4	22.48	25.41	22.55	24.17	18.09	17.85	19.49	15.70	22.06
CN1	4.34	5.02	4.66	4.35	5.11	1.94	4.13	2.09	2.56
CN2	6.59	8.53	6.96	7.25	8.67	3.90	5.98	3.84	5.28
CN3	12.05	14.46	12.01	12.66	11.65	8.98	10.64	8.25	10.79
FN1	9.55	9.99	10.83	10.85	7.41	6.23	7.72	5.56	8.50
FN2	11.80	13.50	13.13	13.88	10.98	8.07	9.92	7.35	11.22
FN3	17.26	19.43	18.17	19.29	13.96	13.28	14.79	11.77	16.74
EN1	6.83	7.32	7.47	9.25	8.51	4.41	5.50	4.27	5.93
EN2	4.88	4.24	5.29	6.15	4.95	3.22	4.10	3.01	3.64
EN3	1.77	2.29	1.37	2.23	2.32	3.05	2.10	3.56	2.41

Anexo 14

Errores absolutos promedio para el conjunto completo de ácidos carboxílicos con SMD.

	B3LYP	BLYP	BHandHLYP	BMK	M052X	PBE	PBE0	PW91	TPSS
A	1.47	1.27	2.35	0.98	0.98	1.15	2.28	1.17	2.04
B	12.86	12.72	13.40	12.38	12.36	11.96	12.31	12.01	12.19
C1	14.52	14.12	15.26	14.82	13.58	12.71	13.38	12.48	13.88
C2	17.23	17.57	17.65	17.75	15.22	14.44	15.37	13.56	16.02
C3	19.89	20.63	19.96	21.11	17.48	16.48	17.73	15.46	18.65
D	1.34	1.39	1.22	1.18	1.54	1.90	0.92	1.39	1.29
E1	4.14	3.27	5.42	6.75	4.75	1.06	3.65	1.05	2.38
E2	1.44	1.17	3.02	3.83	4.44	1.35	1.67	1.21	0.88
E3	1.80	2.74	1.12	0.84	1.07	3.02	0.85	2.81	2.51
F1	9.59	9.67	10.66	11.04	7.97	5.59	7.68	4.52	7.55
F2	11.26	11.58	12.52	13.48	9.19	6.34	8.74	5.00	9.24
F3	13.96	14.52	14.91	16.41	10.83	8.07	10.73	6.07	11.38
F4	16.63	17.59	17.22	19.77	13.09	10.11	13.09	7.98	14.01
CN1	12.71	12.52	13.43	12.70	12.40	11.75	12.15	11.89	12.36
CN2	15.41	15.46	15.83	15.63	14.05	13.47	14.13	12.97	14.49
CN3	18.08	18.52	18.14	18.99	16.31	15.51	16.49	14.87	17.12
FN1	9.44	9.48	10.69	11.36	8.01	5.38	7.51	4.41	7.72
FN2	12.15	12.42	13.09	14.29	9.66	7.10	9.50	5.48	9.85
FN3	14.81	15.48	15.40	17.65	11.92	9.14	11.86	7.38	12.48
EN1	5.96	5.37	7.24	8.87	5.93	1.71	4.88	1.08	3.78
EN2	3.25	2.43	4.84	5.94	4.28	0.84	2.89	0.92	1.65
EN3	0.93	0.99	2.53	2.58	3.93	2.05	0.87	2.22	1.06

Anexo 15

Errores absolutos promedio para aminas mono-funcionales con ciclos termodinámicos y PCM.

	BHandHLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2
A	1.45	2.41	3.43	2.21	2.53	3.29	1.94	3.23	1.36	2.03
B	1.24	1.28	3.01	1.39	2.25	1.15	0.97	1.07	1.12	1.07
C1	1.17	1.19	0.65	1.20	0.61	1.94	1.15	1.41	1.74	0.92
C2	2.84	3.33	2.18	4.12	1.24	3.57	2.68	2.68	3.92	2.37
C3	4.29	4.85	3.66	5.97	2.60	4.29	3.73	3.31	5.02	3.31
D	1.00	1.14	1.33	0.83	0.96	1.14	1.01	1.14	1.22	0.42
F1	3.38	3.04	5.01	4.46	1.84	1.00	1.36	1.10	3.29	3.88
F2	2.03	1.21	2.57	2.45	0.54	0.63	0.18	1.47	1.23	3.12
F3	0.57	1.09	0.46	0.93	1.91	2.25	1.51	2.73	0.95	1.65
F4	1.10	2.45	1.66	2.32	3.27	2.97	2.55	3.36	2.05	1.07
CN1	0.81	0.80	2.68	1.42	1.78	1.05	0.56	0.88	1.38	0.64
CN2	1.53	1.93	0.60	1.96	0.59	3.01	1.85	2.45	3.47	1.43
CN3	3.19	3.68	2.27	3.90	1.36	3.95	3.10	3.28	4.25	2.44
FN1	5.05	4.85	7.12	6.99	2.90	1.80	2.52	1.39	5.64	4.90
FN2	3.71	3.03	4.68	4.99	1.13	0.48	1.23	0.93	3.59	4.14
FN3	1.66	0.52	1.44	1.78	0.70	1.69	0.79	2.50	1.79	2.59
FN4	0.74	1.28	0.37	1.18	2.02	2.63	1.93	3.33	3.26	1.58

Anexo 16

Errores absolutos promedio para aminas mono-funcionales con SMD.

	BHandHLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2
A	0.92	0.56	1.24	0.65	1.11	1.49	0.44	1.37	0.86	0.43
B	11.91	11.57	11.50	11.50	11.39	10.17	10.49	10.32	11.28	10.53
C1	11.29	10.50	10.84	10.01	10.80	9.81	10.28	10.38	10.38	10.61
C2	9.42	8.94	11.41	8.05	8.99	9.29	9.15	9.95	9.49	9.34
C3	25.11	24.40	23.00	23.64	25.36	9.54	9.69	26.64	24.90	10.61
D	1.09	1.06	1.04	1.25	0.87	1.26	1.01	1.17	0.95	0.80
F1	5.00	4.14	4.29	5.99	2.84	0.55	1.70	1.33	2.48	3.96
F2	4.39	3.07	3.63	4.51	2.25	0.88	1.48	1.29	1.58	4.04
F3	2.51	1.50	4.20	2.54	0.64	1.24	0.58	1.70	0.69	2.77
F4	18.21	17.13	16.36	18.72	17.61	1.00	1.45	16.46	16.73	4.05
CN1	12.66	11.97	19.74	11.60	11.12	9.92	10.98	10.20	11.16	11.10
CN2	10.37	10.01	23.53	9.26	9.16	8.96	9.50	9.38	9.87	9.76
CN3	25.61	25.09	23.63	24.54	25.73	9.01	9.85	25.77	24.93	10.64
FN1	6.36	5.60	13.19	7.57	3.16	0.71	2.39	1.52	3.25	4.45
FN2	5.75	4.54	12.53	6.10	2.57	0.85	2.18	1.53	2.36	4.54
FN3	3.47	2.58	16.32	3.76	0.61	1.57	0.80	2.27	1.07	3.20
FN4	18.70	17.66	16.45	19.03	17.36	1.52	1.40	16.81	16.59	4.07

Anexo 17

Errores absolutos promedio para aminas poli-funcionales con ciclos termodinámicos y PCM.

	PBE0	BHANDHLYP	M05-2X
A	1.04	0.90	2.15
B	0.88	0.49	2.44
C1	0.41	0.63	1.14
D	3.41	3.20	3.23
F2	1.38	2.77	0.51
CN1	1.78	1.72	2.69

Anexo 18

Errores absolutos promedio para aminas poli-funcionales con SMD.

	PBE0	PBE	PW91
A	0.88	0.38	0.34
D	1.64	1.20	1.19
F1	2.73	0.96	0.33
F2	1.99	0.89	0.51
F3	0.90	0.34	0.53
FN1	3.36	0.76	0.69
FN2	2.62	0.76	0.89
FN3	1.53	0.66	1.07

Anexo 19

Errores absolutos promedio para el conjunto completo de aminas con ciclos termodinámicos y PCM.

	BHandHLYP	PBE0	M05-2X
A	1.18	1.49	2.34
B	0.87	0.93	2.34
C1	0.90	0.78	0.88
D	0.74	0.79	0.90
F2	2.40	0.78	0.52
CN1	1.26	1.17	2.23

Anexo 20

Errores absolutos promedio para el conjunto completo de aminas con SMD.

	PBE0	PBE	PW91
A	0.66	0.93	0.86
D	1.33	1.23	1.18
F1	2.21	0.75	0.83
F2	1.73	0.88	0.90
F3	0.74	0.79	1.12
FN1	2.87	0.73	1.11
FN2	2.40	0.80	1.21
FN3	1.16	1.11	1.67

Anexo 21

Energías libre de Gibbs (kcal/mol) obtenidas con el funcional BHandHLYP para el acomplejamiento de Cu(II) con dopamina.

	H₃D⁺	H₂D	HD⁻	D²⁻
<i>EFS0</i>				
sitio N		-10.70	-18.86	
sitio Om	30.95	26.96	-10.60	
sitio Op	0.66	27.20	-10.20	
sitio OO	-6.09	-6.08	-27.99	-40.87
<i>EFS4</i>				
sitio N		-1.51	-2.26	
sitio Om	9.09	8.16	-8.58	
sitio Op	7.41	5.67	0.63	
sitio OO	6.51	5.38	-13.32	-35.71
<i>EFS6</i>				
sitio N		-3.40	39.84	
sitio Om	8.29	5.39	-7.13	
sitio Op	7.55	5.46	-0.72	
sitio OO	5.35	4.58	-11.31	11.89
<i>EFS8</i>				
sitio N		0.22	1.24	
sitio Om	15.34	10.68	-5.66	
sitio Op	15.44	13.82	3.42	
sitio OO	7.69	7.09	-7.33	-28.78

Anexo 22

Energías libre de Gibbs (kcal/mol) obtenidas con el funcional BMK para el acomplejamiento de Cu(II) con dopamina.

	H₃D⁺	H₂D	HD⁻	D²⁻
<i>EFS0</i>				
sitio N		-15.87	-25.41	
sitio Om	-2.44	-3.84	-20.68	
sitio Op	-1.45	-2.97	-13.99	
sitio OO	-9.05	-10.09	-32.74	-48.44
<i>EFS4</i>				
sitio N		-3.41	-6.72	
sitio Om	9.24	8.35	-11.77	
sitio Op	7.47	5.64	-0.78	
sitio OO	7.64	6.59	-11.64	-36.00
<i>EFS6</i>				
sitio N		-5.06	-5.23	
sitio Om	8.81	5.93	-6.92	
sitio Op	8.12	6.14	-0.28	
sitio OO	6.31	5.63	-10.29	-33.10
<i>EFS8</i>				
sitio N		-0.85	0.37	
sitio Om	19.14	12.68	-5.20	
sitio Op	17.51	16.03	3.58	
sitio OO	10.30	9.14	-5.63	-27.56

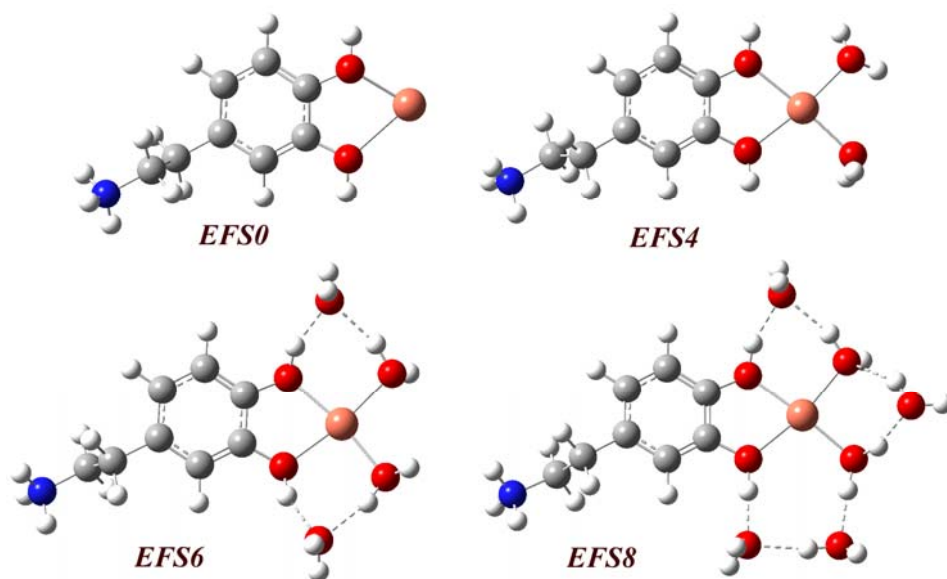
Anexo 23

Energías libre de Gibbs (kcal/mol) obtenidas con el funcional B3LYP para el acomplejamiento de Cu(II) con dopamina.

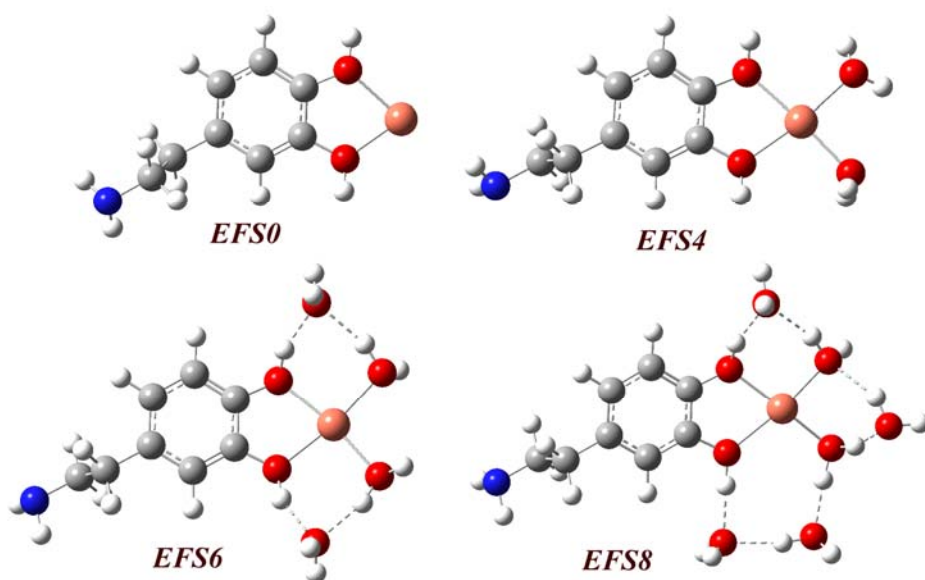
	H₃D⁺	H₂D	HD⁻	D²⁻
<i>EFS0</i>				
sitio N		-20.26	-42.34	
sitio Om	-8.43	-9.97	-36.18	
sitio Op	-8.12	-10.61	-30.78	
sitio OO	-11.33	-12.40	-37.22	-64.38
<i>EFS4</i>				
sitio N		-4.80	-7.63	
sitio Om	8.24	7.27	-14.31	
sitio Op	6.12	4.35	-2.74	
sitio OO	6.52	5.50	-13.56	-39.28
<i>EFS6</i>				
sitio N		-6.14	-5.74	
sitio Om	7.72	4.82	-8.50	
sitio Op	7.08	5.49	-0.90	
sitio OO	4.74	4.06	-12.34	-35.75
<i>EFS8</i>				
sitio N		-1.94	-0.06	
sitio Om	15.90	10.65	-5.34	
sitio Op	15.73	14.73	3.53	
sitio OO	9.16	7.07	-7.89	-31.36

Anexo 24

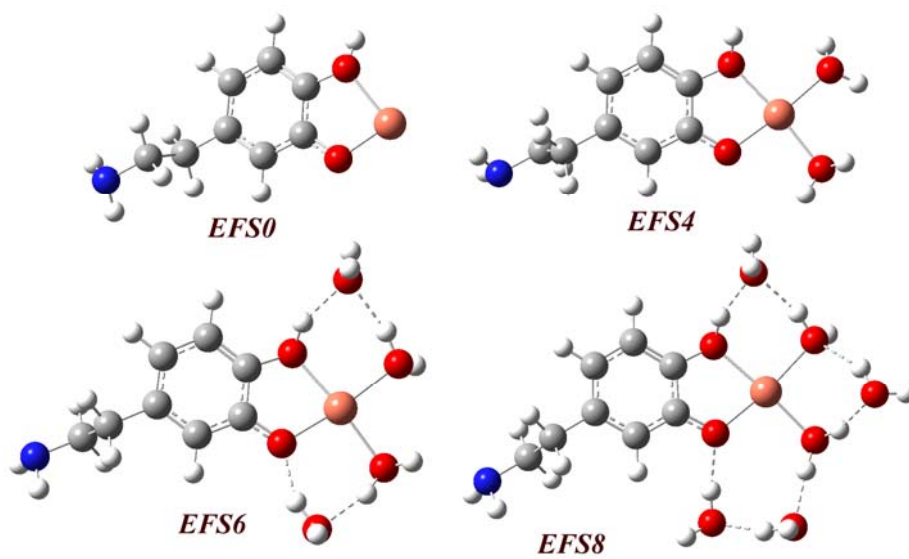
Estructuras de los complejos de Cu(II) con dopamina protonada.



Estructuras de los complejos de Cu(II) con dopamina neutra.



Estructuras de los complejos de Cu(II) con dopamina mono-aniónica.



Publicaciones

First Principles Calculations of pK_a Values of Amines in Aqueous Solution: Application to Neurotransmitters

Aida Mariana Rebollar-Zepeda and Annia Galano*

A systematic search for computational strategies able to accurately predict pK_a values of protonated amines (HA^+) has been performed. Thermodynamic cycles have been used in conjunction with the cluster-continuum model, including up to eight explicit water molecules per system. The calculations have been performed with nine different methods within the density functional theory (DFT), and with MP2. It was found that including one explicit water molecule in the vicinity of the protonated site is highly important for obtaining accurate results. We recommend the **F2** reaction scheme [$HA^+(H_2O) + 3H_2O = A + H_3O^+(3H_2O)$], combined with M05-2X/6-311++G(d,p) calculations. This particular combination produces mean unsigned errors (MUE) equal to 0.54 for all the tested

amines, and 0.51 pK_a units for the studied neurotransmitters. Moreover it was found to be the only one that systematically leads to errors lower than, or very close to, 1 unit of pK_a for every one of the studied species, with the maximum error equal to 1.09 pK_a units. This scheme has the additional advantages of being computationally feasible for amines of relative large size, independent of experimental values, and free of fitting-based corrections. In addition the high influence of the reference acid of choice, in the proton exchange scheme (D), on the quality of the pK_a values of amines is discussed. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/qua.24048

Introduction

Acid-base equilibria are very important phenomena, with implications not only for chemistry but also for other science fields such as pharmacology, medicine, biology, food science, etc. A huge amount of chemical compounds behave as Brønsted acids (or bases) in aqueous solution, and their acidity is commonly expressed as pK_a . Even though there are several experimental techniques that have been successfully applied to accurately obtain pK_a 's, sometimes this becomes a challenging, or very expensive, task^[1,2] Therefore it is not surprising that a large amount of effort has been devoted in the last decades to obtain reliable pK_a values using theoretical methods.^[3–27]

The most common strategies and current trends in pK_a calculations have been recently and thoroughly reviewed by Ho and Coote.^[28] Considering all the information gathered so far it becomes evident that estimating accurate pK_a 's using computational methodologies remains a very challenging task. In fact it is currently accepted that mean absolute deviations smaller than 2 units of pK_a are reasonably accurate.^[28] The main errors seem to arise from modeling solvation. In particular dielectric continuum models (DCM), which are currently the most widely used for calculating pK_a 's in aqueous solution, are frequently not good enough for systems where the short-range solute–solvent interactions are important. One way to overcome this issue, without making the calculations computationally prohibited, is to include a few solvent molecules in close proximity to the solute in addition of using a DCM. This hybrid explicit-implicit solvation model is known as the discrete-continuum,^[29] the supermolecule-reaction-field,^[30] and the cluster-continuum model.^[31] In addition using reaction schemes that maximize systematic error cancelations has been

shown to significantly improve the accuracy of calculated pK_a 's.^[19,32] Among all possible schemes, those that are experiment-independent are particularly appealing from a theoretical point of view. Another desirable characteristic of a computational strategy designed for predicting accurate pK_a 's is that it should be feasible at reasonable computational costs for relative large-sized systems. Otherwise it would be unpractical for most of the systems of chemical and biological interest.

In this work, we have tested up to 29 reaction schemes for the pK_a calculation of amines. This family of compounds has been chosen because they are ubiquitous in nature, where their behavior is frequently related to the pH of the media and their pK_a 's.^[33–37] We have used thermodynamic cycles, including up to four explicit water molecules for solvating the ionic species, and up to 8 water molecules per scheme. In addition we have tested 10 different levels of theory. They are used in combination with the polarizable continuum model (PCM), which is currently one of the most widely used for calculating pK_a 's in aqueous solution. The reliability of the tested approaches has been investigated using a set of 10 amines. A few strategies were identified as reliable and are recommended for accurate calculations of aqueous pK_a 's of relatively large-sized molecules with protonated amino sites, at reasonable computational cost. They have been used to calculate the pK_a 's of 5 neurotransmitters, and their efficiency was further tested this way.

A. M. Rebollar-Zepeda, A. Galano
Departamento de Química, División de Ciencias Básicas e Ingeniería,
Universidad Autónoma Metropolitana-Iztapalapa, Av San Rafael Atlixco No.186,
Col.Vicentina C.P.09340, México D.F
E-mail: agalano@xanum.uam.mx

© 2012 Wiley Periodicals, Inc.

Computational Details

Full geometry optimizations and frequency calculations were performed with the package of programs Gaussian 03^[38] without any symmetry constrain. Different levels of theory have been used: BHandHLYP, BLYP, B3LYP, PBE, PBE0, PW91, BMK, TPSS, and M05-2X density functional theory (DFT) methods; and the MP2 wave-function. All of them in conjunction with the 6-311++G(d,p) basis set. Local minima were identified by the absence of imaginary frequencies. The stationary points were first modeled in gas phase (vacuum), and solvent effects were included *a posteriori* by single point calculations using the polarizable continuum model, specifically the integral-equation-formalism (IEF-PCM).^[39] In PCM calculations the choice of the solute cavity is important because the computed energies and properties strongly depend on the cavity size. In this study, the cavity has been built using the united atom model for Hartree-Fock (UAHF) method,^[40] at HF/6-31+g(d), which is the recommended approach for predicting free energies of solvation according to the Gaussian 03 User's Reference.^[41]

Relative Gibbs free energies in solution (ΔG_s) have been computed using the Hess law and thermodynamic cycles, as the sum of the corresponding gas-phase free energy (ΔG_g) and the free energy of solvation (ΔG_{solv}). They include standard thermal corrections at 298.15 K. In all the cases the used reference state is 1M. Thermodynamic cycles allow obtaining ΔG_s from Gibbs free energies in gas phase (ΔG_g) and Gibbs free energies of solvation (ΔG_{solv}). This seems to be an efficient strategy since continuum solvation models are parameterized to produce accurate ΔG_{solv} , but the levels of theory at which they are usually implemented are not sufficient to accurately reproduce ΔG_s .²⁸

The aqueous Gibbs free energies for the deprotonation reactions are in turn used to compute the acid equilibrium constant (K_a), according to:

$$K_a = e^{-\Delta G_s/RT} \quad (1)$$

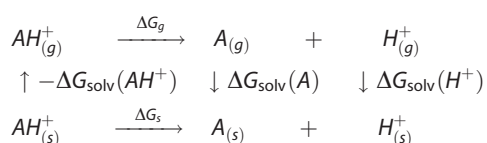
and the $\text{p}K_a$ values are obtained using its definition:

$$\text{p}K_a = -\log(K_a) \quad (2)$$

which can also be calculated directly from ΔG_s , using expressions that depend on the particular scheme of reaction used to model de deprotonation process.

Reaction schemes

The reaction schemes tested in this work are reported in Table 1, together with the expressions corresponding to direct calculations of $\text{p}K_a$ from ΔG_s . Details on these reaction schemes can be find elsewhere.^[32] The ΔG_s values have been calculated using the strategy based on thermodynamic cycles described above. For example for the reaction scheme **A**, the cycle is:



and the Gibbs free energy of reaction in solution (ΔG_s) is obtained as the sum of the Gibbs free energy of reaction in vacuum (ΔG_g) and the difference in solvation free energies ($\Delta\Delta G_{\text{solv}}$):

$$\Delta G_s = \Delta G_g + \Delta\Delta G_{\text{solv}} \quad (3)$$

where $\Delta\Delta G_{\text{solv}}$ and ΔG_g are calculated as:

$$\Delta\Delta G_{\text{solv}} = \Delta G_{\text{solv}}(A) + \Delta G_{\text{solv}}(H^+) - \Delta G_{\text{solv}}(AH^+) \quad (4)$$

$$\Delta G_g = G_g(A) + G_g(H^+) - \Delta G_g(AH^+) \quad (5)$$

with ΔG_{solv} representing the free energies of solvation of each species.

There are some details regarding the used reaction schemes for amines that should be discussed before analyzing the results. For scheme **A** we have used $\Delta G_{\text{solv}}(H^+) = -265.89$ kcal/mol, based on the recommendation of Camaioni and Schwerdtfeger.^[42] Regarding the cluster-continuum model, the number of water molecules explicitly included in the vicinity of the protonated site varies depending on the chemical nature of the studied amines. For NH_4^+ up to four molecules can be included, each of them H bonded to one of the H atoms in the protonated site. On the other hand, for primary (P), secondary (S), and tertiary (T) amines the maximum number of water molecules in the cluster-continuum model is 3, 2 and 1; respectively, i.e., equal to the number of H in the protonated N site. This means that even when for ammonium all the reaction schemes shown in Table 1 have been used, for primary amines schemes **C4**, **CN4**, **E4**, **EN4**, **F5**, and **FN5** are not included. For secondary amines the tested schemes are **A**, **B**, **C1**, **C2**, **CN1**, **CN2**, **E1**, **E2**, **EN1**, **EN2**, **F1**, **F2**, **F3**, **FN1**, **FN2**, and **FN3**; and for tertiary amines they are **A**, **B**, **C1**, **CN1**, **E1**, **EN1**, **F1**, **F2**, **FN1**, and **FN2**.

Even though reaction schemes **E** and **EN** are not expected to be among the best ones for amines, they have been used for comparison purposes with the results previously obtained for phenols.^[32] This first comparison has been made using only the BHandHLYP functional, since it has been previously demonstrated that the quality of the $\text{p}K_a$ predictions strongly depend on the chosen scheme, while is only moderately influenced by the method of calculation. The signed errors, with respect to the experimental values are provided in Supporting Information Table 1S, while the mean unsigned errors (MUE) are shown in Figure 1 to facilitate comparisons. As this figure shows **E** and **EN** schemes are those leading to the largest errors for all the tested amines. This is a logical finding since these schemes were specifically designed to compensate the errors arising from $\text{p}K_a$ calculations of deprotonation processes that do not conserve the number of charged species on both sides of the equilibrium; i.e., the reactants are all neutral species while the products are one anion and one cation. This is the case of phenols, and also of carboxylic acids, but is not the case of protonated amines. For the latter the deprotonation process does conserve the number and charge of the species at both sides of the equilibrium. Therefore using schemes **E** and **EN** causes and artificial increase in the errors associated

Table 1. Reaction schemes and the corresponding expressions to directly calculate pK_a values from ΔG_s for NH_4^+ .

Scheme	Equilibrium	$pK_a =$	Used for
A	$NH_4^+ \rightleftharpoons H^+ + NH_3$	$\frac{\Delta G_s}{RT \ln(10)}$	NH_4^+, HP^+, HS^+, HT^+
B	$NH_4^+ + H_2O \rightleftharpoons H_3O^+ + NH_3$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$	NH_4^+, HP^+, HS^+, HT^+
C1	$NH_4^+ (H_2O) \rightleftharpoons H_3O^+ + NH_3$	$\frac{\Delta G_s}{RT \ln(10)}$	NH_4^+, HP^+, HS^+, HT^+
C2	$NH_4^+ (2H_2O) \rightleftharpoons H_3O^+ + NH_3 + H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + \log[H_2O]$	NH_4^+, HP^+, HS^+
C3	$NH_4^+ (3H_2O) \rightleftharpoons H_3O^+ + NH_3 + 2H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 2 \log[H_2O]$	NH_4^+, HP^+
C4	$NH_4^+ (4H_2O) \rightleftharpoons H_3O^+ + NH_3 + 3H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 3 \log[H_2O]$	NH_4^+
CN1	$NH_4^+ (H_2O) + H_2O \leftarrow H_3O^+ + NH_3 (H_2O)$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$	NH_4^+, HP^+, HS^+, HT^+
CN2	$NH_4^+ (2H_2O) \rightleftharpoons H_3O^+ + NH_3 (H_2O)$	$\frac{\Delta G_s}{RT \ln(10)}$	NH_4^+, HP^+, HS^+
CN3	$NH_4^+ (3H_2O) \rightleftharpoons H_3O^+ + NH_3 (H_2O) + H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + \log[H_2O]$	NH_4^+, HP^+
CN4	$NH_4^+ (4H_2O) \rightleftharpoons H_3O^+ + NH_3 (H_2O) + 2H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 2 \log[H_2O]$	NH_4^+
D	$NH_4^+ + Ref \leftarrow NH_3 + HRef$	$\frac{\Delta G_s}{RT \ln(10)} + pKa(HRef)$	NH_4^+, HP^+, HS^+, HT^+
E1	$NH_4^+ (H_2O) + OH^- (3H_2O) \rightleftharpoons NH_3 + 5H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 5 \log[H_2O]$	NH_4^+, HP^+, HS^+, HT^+
E2	$NH_4^+ (2H_2O) + OH^- (3H_2O) \rightleftharpoons NH_3 + 6H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 6 \log[H_2O]$	NH_4^+, HP^+, HS^+
E3	$NH_4^+ (3H_2O) + OH^- (3H_2O) \rightleftharpoons NH_3 + 7H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 7 \log[H_2O]$	NH_4^+, HP^+
E4	$NH_4^+ (4H_2O) + OH^- (3H_2O) \rightleftharpoons NH_3 + 8H_2O$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 8 \log[H_2O]$	NH_4^+

(Continued)

Table 1 (Continued)			
Scheme	Equilibrium	$pK_a =$	Used for
EN1	$\text{NH}_4^+ (\text{H}_2\text{O}) + \text{OH}^- (3\text{H}_2\text{O}) \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + 4\text{H}_2\text{O}$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 4 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+ , HT^+
EN2	$\text{NH}_4^+ (2\text{H}_2\text{O}) + \text{OH}^- (3\text{H}_2\text{O}) \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + 5\text{H}_2\text{O}$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 5 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+
EN3	$\text{NH}_4^+ (3\text{H}_2\text{O}) + \text{OH}^- (3\text{H}_2\text{O}) \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + 6\text{H}_2\text{O}$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 6 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+
EN4	$\text{NH}_4^+ (\text{H}_2\text{O}) + \text{OH}^- (4\text{H}_2\text{O}) \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + 7\text{H}_2\text{O}$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 7 \log[\text{H}_2\text{O}]$	NH_4^+
F1	$\text{NH}_4^+ + 4\text{H}_2\text{O} \rightleftharpoons \text{NH}_3 + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 4 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+ , HT^+
F2	$\text{NH}_4^+ (\text{H}_2\text{O}) + 3\text{H}_2\text{O} \rightleftharpoons \text{NH}_3 + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 3 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+ , HT^+
F3	$\text{NH}_4^+ (2\text{H}_2\text{O}) + 2\text{H}_2\text{O} \rightleftharpoons \text{NH}_3 + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 2 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+
F4	$\text{NH}_4^+ (3\text{H}_2\text{O}) + \text{H}_2\text{O} \rightleftharpoons \text{NH}_3 + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+
F5	$\text{NH}_4^+ (4\text{H}_2\text{O}) \rightleftharpoons \text{NH}_3 + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)}$	NH_4^+
FN1	$\text{NH}_4^+ + 5\text{H}_2\text{O} \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 5 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+ , HT^+
FN2	$\text{NH}_4^+ (\text{H}_2\text{O}) + 4\text{H}_2\text{O} \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 4 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+ , HT^+
FN3	$\text{NH}_4^+ (2\text{H}_2\text{O}) + 3\text{H}_2\text{O} \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 3 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+ , HS^+
FN4	$\text{NH}_4^+ (3\text{H}_2\text{O}) + 2\text{H}_2\text{O} \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 2 \log[\text{H}_2\text{O}]$	NH_4^+ , HP^+
FN5	$\text{NH}_4^+ (4\text{H}_2\text{O}) + \text{H}_2\text{O} \rightleftharpoons \text{NH}_3 (\text{H}_2\text{O}) + \text{H}_3\text{O}^+ (3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - \log[\text{H}_2\text{O}]$	NH_4^+

HP^+ , protonated primary amines; HS^+ , protonated secondary amines; HT^+ , protonated tertiary amines.

with their pK_a calculations. Accordingly these schemes are not longer considered in this work, despite of the fact that scheme **E1** is recommended for phenols, for which it produces a MUE = 0.82 when used in combination with the BHandHLYP func-

tional, and even lower with PBE0.^[32] At this point it seems worthwhile to call attention to the fact that the different charges of the chemical species involved in acid/base equilibria, depending on the kind of compound involved, is an

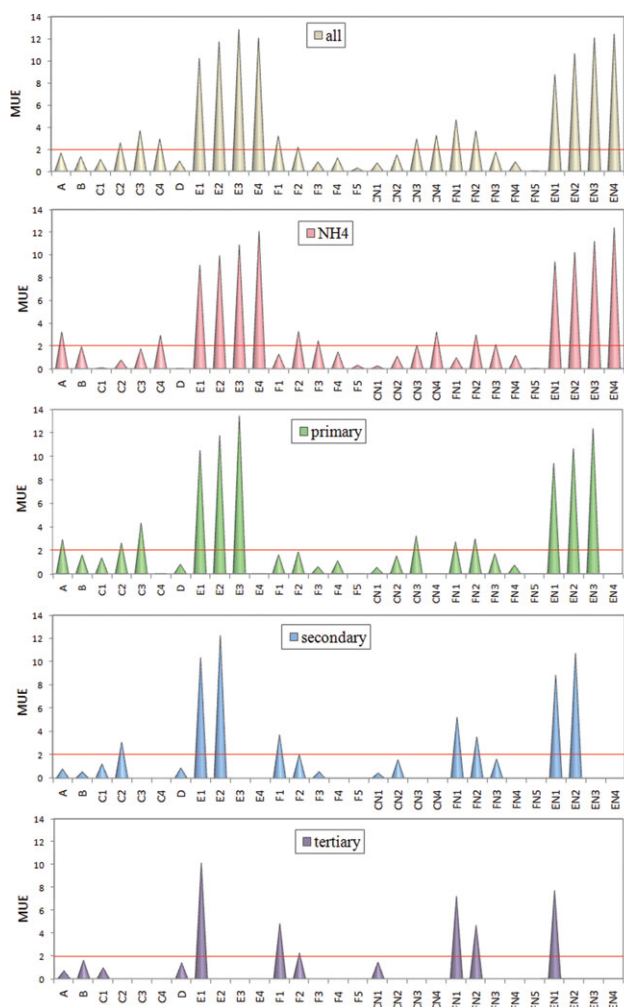


Figure 1. Mean unsigned errors (MUE) for the pK_a values calculated with BHandHLYP. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://www.interscience.wiley.com).]

additional problem for finding a universal strategy aiming to accurately reproduce pK_a values. Accordingly a practical approach could be to identify optimal strategies for each family of compounds, which is what we intend to do.

Geometry considerations

When using thermodynamic cycles the relaxation of the species, due to the presence of the solvent, is ignored in the calculation of the Gibbs free energies of solvation. Moreover the differences between geometries optimized in gas phase and those optimized in aqueous solution is expected to be larger for ionic species. Therefore we have tested how large these differences can be for the studied systems. To that purpose the geometries of the protonated forms of ammonia (NH_4^+), methyl-amine ($HP1^+$), dimethyl-amine ($HS1^+$), and trimethyl-amine ($HT1^+$) have also been optimized in aqueous solution with BHandHLYP, M05-2X, PBE0, and MP2 methods. The analysis of the differences has been performed based on the bond distances of the protonated N center, which are expected to be the geometrical parameters more sensitive to the presence of the solvent.

In general, and as expected, the geometries obtained in gas phase become closer to those that were optimized in aqueous solution as the number of explicit water molecules in the vicinity of the protonated site increases. It was found that the average differences are $\leq 0.014 \text{ \AA}$ for all the tested methods (Table 2). The maximum differences were found for $HT1^+$,

Table 2. Unsigned differences (\AA) with respect to geometries optimized in aqueous solution.

	BHandHLYP	M05-2X	PBE0	MP2
NH_4^+	d(N-H)	d(N-H)	d(N-H)	d(N-H)
0 H_2O	0.011	0.012	0.012	0.012
1 H_2O	0.016	0.017	0.019	0.016
2 H_2O	0.012	0.012	0.015	0.012
3 H_2O	0.005	0.005	0.007	0.005
4 H_2O	0.004	0.005	0.002	0.006
$HP1^+$	d(N-H)	d(N-H)	d(N-H)	d(C-H)
0 H_2O	0.015	0.016	0.016	0.016
1 H_2O	0.013	0.013	0.015	0.013
2 H_2O	0.006	0.007	0.008	0.006
3 H_2O	0.003	0.005	0.002	0.007
$HS1^+$	d(N-H)	d(N-H)	d(C-H)	d(C-H)
0 H_2O	0.020	0.021	0.021	0.021
1 H_2O	0.013	0.014	0.012	0.014
2 H_2O	0.008	0.010	0.007	0.010
$HT1^+$	d(N-H)	d(C-H)	d(C-H)	d(C-H)
0 H_2O	0.022	0.027	0.029	0.033
1 H_2O	0.015	0.020	0.019	0.025
Average	0.012	0.013	0.013	0.014

without any explicit water molecule, and equal to 0.022, 0.027, 0.029, and 0.033 \AA , for BHandHLYP, M05-2X, PBE0, and MP2, respectively. Accordingly, it can be stated that the differences are rather small, and the effect of ignoring relaxation in the calculation of the solvation energies is expected to be only minor.

Results and Discussion

The set of amines, in their base form, used to test the different computational strategies studied in this work are shown in Figure 2. They include ammonia, three primary amines (P1

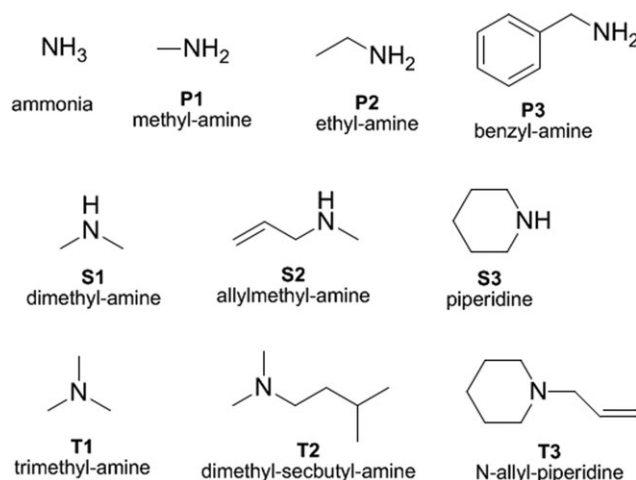


Figure 2. Studied amines in their base form.

to P3), three secondary amines (S1 to S3), and three tertiary amines (T1 to T3). As explained before (and shown in Table 1) the maximum number of explicit water molecules included in the cluster-continuum model changes depending on the kind of amine that is studied. This is why the statistic analyses have been performed for the whole set, but also independently for the different subsets.

Due to the large amount of calculations involved in this work, it was unfeasible to perform exhaustive conformational analyses for the solute-water clusters. We have used chemical intuition, and previous experience, instead to construct the starting geometries in each case. To provide information on the optimized structures of the modeled clusters, the geometries corresponding to the $\text{NH}_4^+/\text{NH}_3$, $\text{HP1}^+/\text{P1}$, $\text{HS1}^+/\text{S1}$, and $\text{HT1}^+/\text{T1}$ systems are shown in Figures 3–6. Equivalent configu-

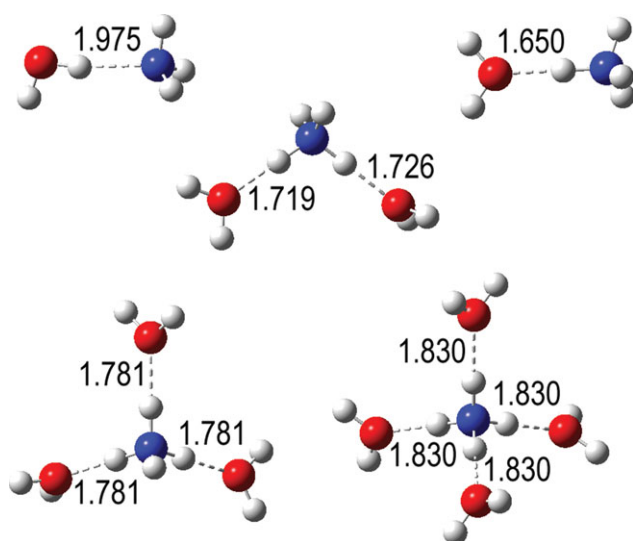


Figure 3. Geometries of the solute-water clusters, for the $\text{NH}_4^+/\text{NH}_3$ system, optimized at MP2/6-311++G(d,p) level of theory. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

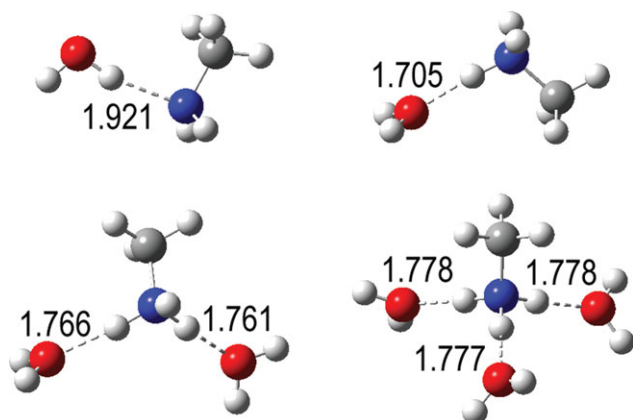


Figure 4. Geometries of the solute-water clusters, for the $\text{HP1}^+/\text{P1}$ system, optimized at MP2/6-311++G(d,p) level of theory. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

rations were located for all the studied amines, depending on their chemical nature (primary, secondary or tertiary). Only small variations were found from changing the calculation

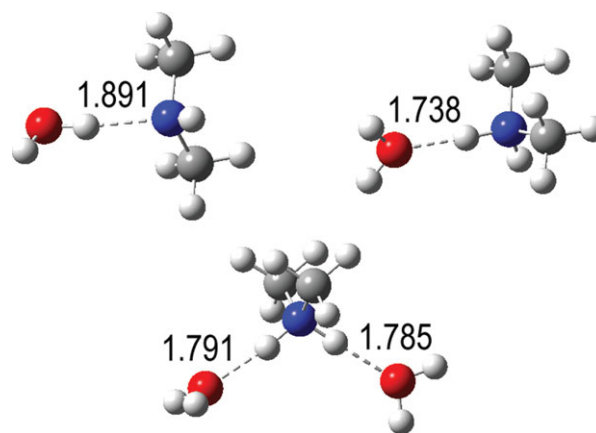


Figure 5. Geometries of the solute-water clusters, for the $\text{HS1}^+/\text{S1}$ system, optimized at MP2/6-311++G(d,p) level of theory. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

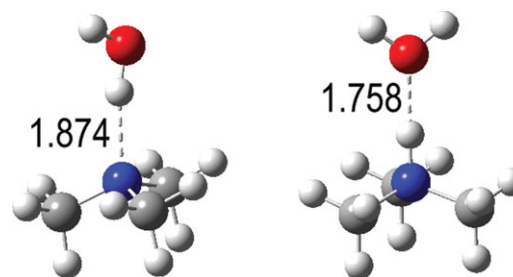


Figure 6. Geometries of the solute-water clusters, for the $\text{HT1}^+/\text{T1}$ system, optimized at MP2/6-311++G(d,p) level of theory. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

method (Fig. 7). Therefore we have chosen MP2 to show the geometrical distribution of the clusters.

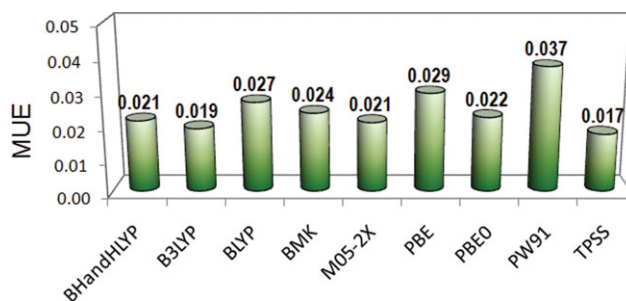


Figure 7. MUE for the H bond distances in the solute-water clusters (Å) optimized with the different DFT methods, using the MP2 values as reference. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The signed errors of the calculated $\text{p}K_a$'s obtained with the tested methods of calculation, and schemes of reaction, are provided in Tables 1S to 10S (Supporting Information). The corresponding mean unsigned errors are reported in Tables 3–7. In addition since the reported data is quite abundant, the values in these tables have also been plotted in Figures 1S to 5S (Supporting Information) to facilitate comparison among them. In these figures the error range associated

Table 3. Unsigned errors (UE) for the calculated pK_a values of NH_4^+ .

	BH&HLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2
A	3.17	4.14	5.24	3.75	4.56	4.22	2.90	4.22	2.55	3.07
B	1.88	1.14	1.20	0.77	0.02	1.67	1.10	1.58	1.24	1.23
C1	0.11	1.25	2.24	0.51	1.20	0.73	1.06	1.05	0.77	0.26
C2	0.72	0.54	1.04	0.76	0.11	0.35	0.38	0.54	0.42	0.59
C3	1.69	1.96	0.52	2.27	0.47	0.92	0.28	0.01	2.10	1.65
C4	2.86	3.37	2.52	4.26	1.76	1.80	1.59	0.66	3.32	2.34
D	0.02	0.30	0.32	0.51	0.01	0.92	0.76	0.86	0.51	0.54
F1	1.31	1.26	3.20	2.89	0.64	0.35	0.07	1.63	1.73	2.79
F2	3.31	3.64	4.24	4.17	0.54	2.05	2.23	1.00	3.74	4.28
F3	2.48	1.86	3.04	2.89	0.56	1.67	1.55	0.48	2.55	3.42
F4	1.50	0.44	1.48	1.39	1.13	0.40	0.90	0.06	0.87	2.37
F5	0.33	0.98	0.53	0.61	2.43	0.49	0.42	0.72	0.36	1.67
CN1	0.20	1.02	2.21	0.41	1.36	0.27	0.33	0.18	0.54	0.11
CN2	1.04	0.78	1.01	0.88	0.26	0.66	0.36	0.70	0.66	0.75
CN3	2.00	2.19	0.55	2.38	0.31	1.92	1.01	1.23	2.32	1.79
CN4	3.18	3.60	2.56	4.37	1.61	2.81	2.33	1.89	3.56	2.49
FN1	1.00	1.02	3.16	2.78	0.49	1.36	0.66	2.87	1.50	2.64
FN2	3.00	3.41	4.21	4.07	0.70	1.05	1.50	0.23	3.51	4.13
FN3	2.16	1.62	3.01	2.78	0.41	0.66	0.82	0.76	2.31	3.27
FN4	1.19	0.21	1.45	1.28	0.98	0.60	0.17	1.29	0.64	2.22
FN5	0.01	1.21	0.56	0.72	2.28	1.50	1.15	1.95	0.59	1.53

Table 4. Mean unsigned errors (MUE) for the calculated pK_a values of the primary amines.

	BH&HLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2
A	2.87	3.94	4.96	3.86	4.42	4.50	3.17	4.47	2.49	3.65
B	1.58	0.95	1.48	0.88	0.40	1.95	1.37	1.82	1.18	1.81
C1	1.34	1.08	0.71	1.69	0.33	1.58	1.06	1.11	1.65	1.45
C2	2.59	2.87	1.74	4.41	1.32	3.25	2.30	2.36	3.44	2.76
C3	4.29	4.85	3.66	5.97	2.60	4.29	3.73	3.31	5.02	3.31
D	0.78	0.82	0.95	0.55	0.55	0.59	0.50	0.60	0.77	0.45
F1	1.61	1.45	3.48	2.78	0.50	0.63	0.38	1.88	1.79	2.20
F2	1.86	1.32	2.71	1.97	0.99	0.28	0.16	1.16	1.32	2.57
F3	0.60	0.78	0.30	0.96	1.99	1.93	1.13	2.41	0.48	1.26
F4	1.10	2.45	1.66	2.32	3.27	2.97	2.55	3.36	2.05	1.07
CN1	0.53	0.62	2.11	0.61	0.92	1.24	0.60	1.08	3.17	0.57
CN2	1.50	1.70	0.44	2.35	0.74	2.92	1.69	2.33	3.84	1.89
CN3	3.19	3.68	2.27	3.90	1.36	3.95	3.10	3.28	4.25	2.44
FN1	2.71	2.62	4.86	4.84	0.79	0.69	0.86	1.86	5.22	3.07
FN2	2.96	2.49	4.11	4.04	0.32	0.50	0.74	1.13	4.76	3.45
FN3	1.70	0.79	1.64	1.45	0.75	1.60	0.75	2.39	3.20	2.13
FN4	0.74	1.28	0.37	1.18	2.02	2.63	1.93	3.33	3.26	1.58

Table 5. Mean unsigned errors (MUE) for the calculated pK_a values of the secondary amines.

	BH&HLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2
A	0.78	2.05	3.13	1.85	2.19	3.01	1.52	2.97	0.73	1.63
B	0.53	0.95	3.31	1.12	2.40	0.58	0.47	0.53	0.62	0.21
C1	1.20	1.21	0.52	1.14	0.57	1.97	1.18	1.46	1.65	0.82
C2	3.08	3.79	2.62	3.82	1.16	3.89	3.07	2.99	4.40	1.98
D	0.86	1.04	1.24	0.48	0.91	1.03	0.91	1.08	1.14	0.20
F1	3.71	3.34	5.31	4.78	1.73	0.85	1.46	0.55	3.55	4.22
F2	1.99	1.19	2.52	2.52	0.26	0.65	0.23	1.51	1.32	3.19
F3	0.54	1.39	0.62	0.90	1.83	2.57	1.89	3.05	1.43	2.04
CN1	0.44	0.44	2.39	1.13	1.76	1.19	0.36	1.01	0.50	0.60
CN2	1.57	2.15	0.76	1.56	0.45	3.11	2.01	2.56	3.10	0.97
FN1	5.22	4.98	7.17	7.05	2.91	1.63	2.52	0.65	4.85	5.23
FN2	3.51	2.83	4.39	4.79	1.10	0.32	1.06	1.07	2.62	4.20
FN3	1.63	0.25	1.24	2.10	0.64	1.79	0.83	2.61	0.38	3.05

with each pair scheme-calculation method is easily identifiable by color. Moreover, those pairs corresponding to the lowest errors are also easier to find this way.

The mean unsigned errors obtained for the pK_a values of NH_4^+ , using different methods of calculation are presented in Table 3. The values obtained with scheme **A**, which is one of the most frequently used, are among those with the largest errors with all the tested calculation methods, including MP2. All of them are larger than 2 units of pK_a . This supports the necessity of finding alternatives schemes of reaction to represent acid/base equilibria. For this particular system most of the alternative schemes produce errors lower than 1 unit of pK_a , and several lead to errors lower than 0.5 units of pK_a , at least with some of the tested methods. The only exception is scheme **CN4**. In general the lowest errors for the pK_a of NH_4^+ are obtained with schemes **D**, **C2**, **CN1**, **CN2**, **F5**, and **C1**. However schemes **FN4**, **F4**, **FN5**, **B**, and **C3** also produce reasonable results for this system. Among the tested methods M05-2X, PBE0, PBE and PW91 were found to be those that most frequently lead to low errors. In addition BHandHLYP, B3LYP, and TPSS calculations also lead to low errors in several cases. Particularly good values were obtained with M05-2X and schemes **D**, **B**, and **C2**; with PBE0 and schemes **F1**, and **FN4**; with BHandHLYP and schemes **D**, **FN5**, and **C1**; with PW91 and schemes **C3**, **F4**, and **CN1**; and with MP2 and scheme CN1. Accordingly, it seems that the pK_a of ammonium can be reproduced at acceptable levels of accuracy with a wide variety of computational strategies.

For primary amines (Table 4, Supporting Information Fig. 2S) it was found that the largest errors systematically arise from using schemes **A**, **C2**, **C3**, **F4**, and **CN3**, which lead to mean unsigned errors larger than 1 unit of pK_a with all the calculation methods, and most of the times larger than 2 or 3 pK_a units. It is interesting to notice that while schemes C2, C3, and F4 lead to a reasonable accuracy for the pK_a of NH_4^+ , they fail for primary amines. This suggests that in the particular case of amines even when a particular scheme of reaction may work well for a subset of them, this not necessarily means that it would work similarly well for the rest of them. Accordingly in this work we will be searching for computational strategies that systematically lead to low errors in the pK_a calculations of all kind of amines

Table 6. Mean unsigned errors (MUE) for the calculated pK_a values of the tertiary amines.

	BH&HLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2
A	0.69	1.23	2.18	0.91	0.99	2.35	1.15	2.27	0.87	0.80
B	1.61	1.94	4.26	2.15	3.96	0.93	1.07	0.86	1.56	1.20
C1	0.96	1.28	0.70	0.78	0.93	2.26	1.22	1.67	1.90	0.48
D	1.38	1.56	1.80	1.46	1.41	1.80	1.62	1.75	1.75	0.60
F1	4.81	4.34	6.26	5.81	3.30	1.52	2.25	0.88	4.53	5.22
F2	2.24	1.12	2.46	2.88	0.36	0.94	0.16	1.73	1.07	3.58
CN1	1.45	1.35	3.53	2.50	2.64	0.74	0.73	0.56	0.47	0.76
FN1	7.21	6.96	9.32	9.09	5.00	3.07	4.18	1.67	6.84	6.41
FN2	4.65	3.75	5.53	6.16	1.98	0.62	1.90	0.57	3.38	4.77

Table 7. Mean unsigned errors (MUE) for the calculated pK_a values of the whole set of tested amines.

	BH&HLYP	B3LYP	BLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS	MP2
A	1.62	2.58	3.61	2.36	2.74	3.38	2.04	3.33	1.48	2.13
B	1.31	1.26	2.83	1.32	2.03	1.20	0.98	1.12	1.13	1.09
C1	1.06	1.20	0.81	1.14	0.67	1.82	1.14	1.38	1.64	0.85
C2	2.54	2.93	2.02	3.64	1.08	3.11	2.36	2.37	3.42	2.11
C3	3.64	4.13	2.88	5.05	2.07	3.45	2.86	2.48	4.29	2.90
C4	2.86	3.37	2.52	4.26	1.76	1.80	1.59	0.66	3.32	2.34
D	0.91	1.06	1.23	0.80	0.86	1.12	0.98	1.11	1.15	0.43
F1	3.17	2.86	4.83	4.30	1.72	0.94	1.23	1.15	3.13	3.77
F2	2.16	1.45	2.73	2.62	0.54	0.77	0.39	1.42	1.48	3.23
F3	0.84	1.20	0.83	1.21	1.71	2.17	1.52	2.41	1.18	1.90
F4	1.20	1.95	1.62	2.08	2.73	2.33	2.14	2.54	1.76	1.39
F5	0.33	0.98	0.53	0.61	2.43	0.49	0.42	0.72	0.36	1.67
CN1	0.74	0.82	2.63	1.31	1.73	0.98	0.54	0.81	1.30	0.59
CN2	1.46	1.76	0.66	1.80	0.55	2.68	1.64	2.20	3.07	1.33
CN3	2.89	3.30	1.84	3.52	1.10	3.44	2.58	2.77	3.77	2.28
CN4	3.18	3.60	2.56	4.37	1.61	2.81	2.33	1.89	3.56	2.49
FN1	4.64	4.47	6.72	6.57	2.66	1.75	2.33	1.54	5.22	4.68
FN2	3.64	3.06	4.63	4.90	1.09	0.54	1.26	0.86	3.58	4.14
FN3	1.73	0.68	1.67	1.92	0.65	1.55	0.79	2.25	1.86	2.69
FN4	0.85	1.01	0.64	1.20	1.76	2.13	1.49	2.82	2.61	1.74
FN5	0.01	1.21	0.56	0.72	2.28	1.50	1.15	1.95	0.59	1.53

simultaneously. In the particular case of primary amines, the schemes that in general lead to better values of pK_a 's are **D**, **CN1**, **F3**, **C1**, and **B**. On the other hand calculations with the M05-2X and PBE0 functionals were identified as those associated with the lowest errors.

In general for secondary amines (Table 5, Supporting Information Fig. 3S) the reaction schemes that lead to the lowest errors with respect to the experimental values are **D**, **CN1**, **B**, **C1**, **F2** and **FN3**. The methods with the best performance, when using the tested schemes, are PBE0 and M05-2X. For tertiary amines (Table 6, Supporting Information Fig. 4S) reaction schemes **C1**, **A**, **CN1**, **D**, and **F2** lead to the lowest errors, while functionals PW91, PBE0, and PBE lead in general to values closest to the experimental ones. In addition schemes **D**, **C1**, **CN1** have been identified as those that are systematically included among the best ones for the different kind of amines.

At this point is important to consider the data for the whole set of tested amines altogether (Table 7, Supporting Information Fig. 5S). However, before doing that it seems relevant to remember that not all the schemes are applicable to all the studied amines. For example **F5** and **FN5** have only been used

for NH_4^+ . The schemes that are common to all the studied amines are **A**, **B**, **C1**, **CN1**, **E1**, **EN1**, **F1**, **F2**, **FN1**, and **FN2**. Therefore we will focus on them since we are trying to propose a reliable strategy for any kind of amine. Among them, schemes **FN1**, **FN2**, **F1**, and **A** poorly reproduce the experimental values. Accordingly they are not recommended for calculating the pK_a 's of amines. As a result schemes **FN1**, **FN2**, and **F1** will no longer be analyzed in this work. We will keep scheme **A** within our analysis since, together with schemes **B** and **D**, it is among the most widely used because of its relative simplicity. Comparing these three schemes (Fig. 8) it becomes evident that scheme **A** is the one with the largest errors. The only mean unsigned errors (MUE) that are below 2 units of pK_a are those arising from TPSS and BHandHLYP calculations. Nevertheless, even with these two methods the MUE values are much higher than the desired accuracy. In addition scheme **A** has another disadvantage, that not only concerns amines but any kind of acid/base equilibria. It involves the proton, and it is known that computational methods poorly reproduce the solvation energies of this particular species. This constitutes a major disadvantage since the reported experimental values of the

solvation free energy of the proton range from -259 to -264 kcal/mol.^[18] Such variation is an important source of error in the calculation of pK_a 's, i.e., it alone represents about 3 pK_a units.

A simple strategy to avoid using experimental values of ΔG_{solv} is to include a water molecule as a coreactant (scheme **B**), i.e., to use the H_3O^+ species instead of the naked proton. It certainly improves the results for amines, compared to those obtained with scheme **A** (Fig. 8). Most of the MUE values are lower than 2 pK_a units, but still higher than 1 pK_a unit for all the tested methods but PBE0. Therefore, even though these values are below the currently accepted accuracy for pK_a calculations, they are higher than what we intend to accomplish.

Scheme **D**, on the other hand, usually lead to more reliable values of pK_a 's. Within this scheme half of the tested methods give MUE values lower than 1 unit of pK_a . They are BHandHLYP, BMK, M05-2X, PBE0, and MP2, with the lowest MUE arising from the latter. The main problem with this approach is that its outcome strongly depends on the choice of the reference acid (*HRef*). Therefore, the quality of the results would be determined by the structural similarity between *HRef* and the

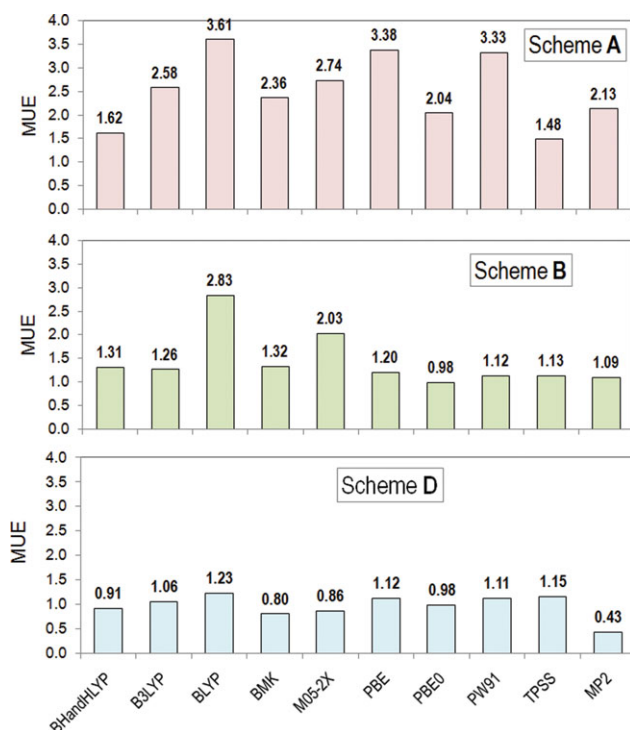


Figure 8. Mean unsigned errors (MUE) for the calculated pK_a values of the whole set of amines, using the reaction schemes **A**, **B**, and **D**. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

studied acid, and also by the accuracy of the experimental value of pK_a for *HRef*. From a theoretical point of view, schemes **D** has another drawback, it directly depends on experimental data. Such experiment-dependency weakens the predictive character of any computational methodology aiming for predicting pK_a 's, or any other chemical property. On the contrary schemes **C1**, **CN1**, and **F2** (Fig. 9) have the advantage of being experiment-independent.

When using scheme **C1**, all the tested methods lead to MUE < 2 pK_a units (Fig. 9). In addition three of them (BLYP, M05-2X, and MP2) yield mean unsigned errors lower than 1 unit of pK_a . For this scheme the lowest MUE value is obtained when the calculations are performed with the M05-2X functional. For scheme **CN1** all but one calculation method (BLYP) produce MUE below the 2 pK_a units limit. Moreover six of the ten tested methods give MUE < 1 unit of pK_a . They are BHandHLYP, B3LYP, PBE, PBE0, PW91, and MP2. In this case the lowest MUE value was obtained for PBE0 calculations. When using the **F2** scheme the results were found to be highly dependent on the method of choice. While calculations with BHandHLYP, BLYP, BMK, and MP2 lead to large errors (MUE > 2 pK_a units), those performed with M05-2X, PBE and PBE0 lead to low values of MUE.

Analyzing all possible combinations reaction scheme—calculation method some clarifications can be made on the gathered data. It was found that scheme **F2**—PBE0 functional was the one producing the lowest MUE (0.39) among all the tested ones. This combination only fails for NH_4^+ , leading to an unsigned error (UE) equal to 2.19 pK_a units (Supporting Information Table 7S). In addition the combination **CN1**—PBE0 also

gives excellent results, with MUE = 0.54 pK_a units. It produces UEs larger than 1 only for T3 (1.31) and P1 (1.05), and the latter is very close to this value. The combination **D**—MP2 also works very well, with MUE = 0.43, and the maximum UE = 1.69 pK_a units for T3, which is the only amine for which the error is larger than 1 (Supporting Information Table 10S). However, this particular combination has two disadvantages, the experiment-dependence mentioned above, and the computational cost of the MP2 methods as the size of the systems increases. MP2 calculations also give very good results when used in combination with the **CN1** scheme. In this case the MUE is 0.59, and the largest UEs are those corresponding to T3 (1.46) and P1 (1.01). The combination **F2**—M05-2X was found to be the only one that systematically lead to UEs lower than, or very close to, 1 unit of pK_a for all the studied amines (Supporting Information Table 5S). Moreover the maximum unsigned error was found to be 1.09, and the MUE = 0.54 pK_a units. The combination **C1**—M05-2X also works satisfactorily, the only amine for which the UE value is not below, or close to, 1 unit of pK_a is T3, and even in this case the UE value is 1.49 pK_a units, i.e., within the currently accepted accuracy. The MUE in this case was found to be 0.67 pK_a units.

It is interesting to notice that while the presence of one explicit water molecule solvating the protonated species seems to be important for properly modeling the acid/base equilibria of the studied amines, including more of them do not necessarily improve the results. We have not a definite explanation for this. However, nowadays the best results for pK_a calculations are based on obtaining a maximum

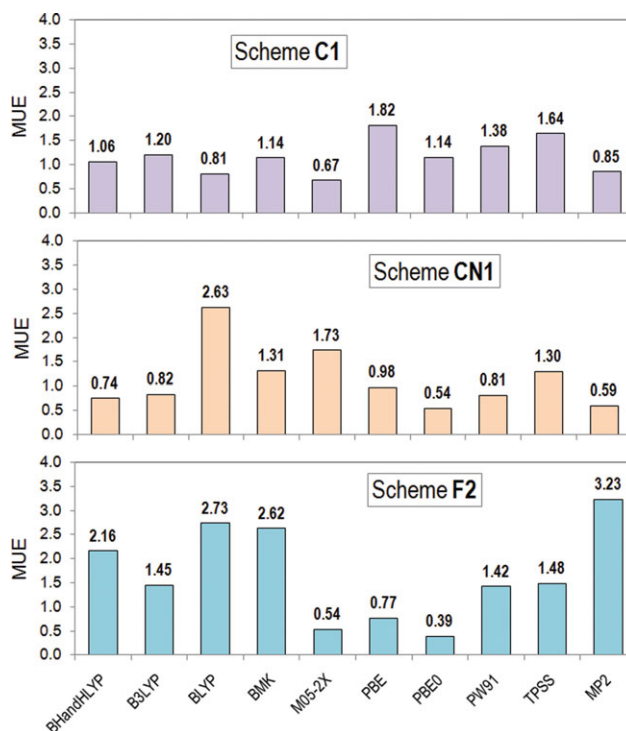


Figure 9. Mean unsigned errors (MUE) for the calculated pK_a values of the whole set of amines, using the reaction schemes **C1**, **CN1**, and **F2**. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

cancellation of errors,^[28] which unfortunately are abundant and of diverse nature for this particular problem. Therefore, apparently the cause of this behavior is related to error cancellation, which seems to be maximum when including one explicit water molecule in the vicinity of the protonated N site.

Based on the results for the tested amines we have chosen two of the best performing methods (PBE0 and M05-2X), together with one for which the performance changes depending on the used scheme of reaction (BHandHLYP) for applying the six selected schemes (**A**, **B**, **C1**, **D**, **CN1** and **F2**) to five neurotransmitters (Fig. 10). Schemes **C1**, **D**, **CN1** and **F2** have been selected because of their good performance, while

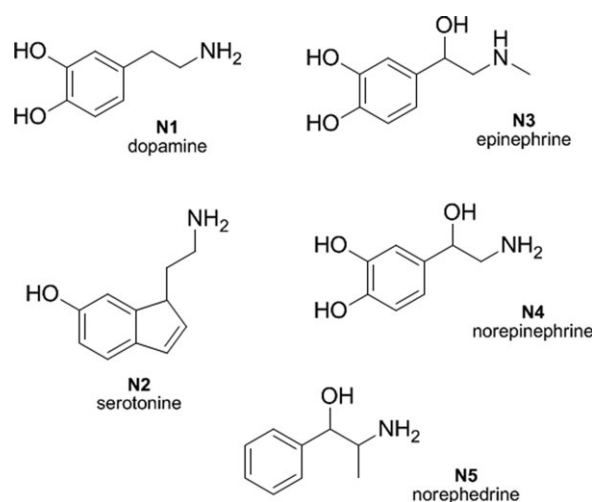


Figure 10. Neurotransmitters, in their base form.

schemes **A** and **B** have been chosen for being among the most currently used. There are several experimental values reported for the pK_a of these species with differences up to 0.18, 0.89, 1.07, 0.22 and 0.06 pK_a units for N1, N2, N3, N4 and N5, respectively (Supporting Information Table 11S). Therefore we have used the average of the values reported for each of the neurotransmitter as reference. The signed errors of the calculated pK_a 's are reported in Supporting Information Table 12S, and the corresponding MUE are presented in Figure 11.

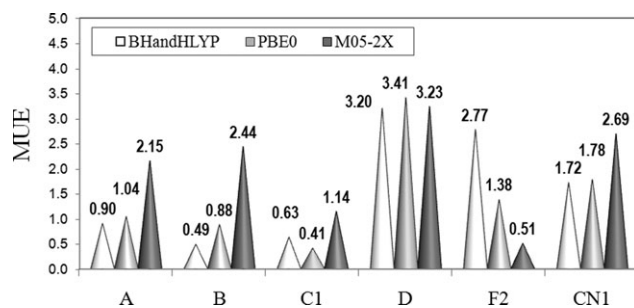


Figure 11. Mean unsigned errors (MUE) for the calculated pK_a values of the studied neurotransmitters.

The approaches that best reproduce the experimental values for the studied neurotransmitters are **C1**—PBE0 (MUE = 0.41), **B**—BHandHLYP (MUE = 0.49), and **F2**—M05-2X (MUE = 0.51).

All of them have MUE values lower than, or very close to 0.5 pK_a units. However, among them the combination **F2**—M05-2X is the only one that also has excellent results for the whole set of tested amines. Accordingly this particular approach is recommended for accurate pK_a calculations of amines in general. In Figure 11 the largest errors arise from using scheme **D**. The data in this figure were obtained using $HP1^+$ as *HRef* since all the studied neurotransmitters, but N3, are primary amines. This is a surprising result since this scheme usually works adequately. It also indicates that amines are a particularly challenging case for pK_a calculations.

To investigate how important is the selection of the reference acid when using the **D** scheme, different amines have been used to that purpose. Accordingly the pK_a calculations of the studied neurotransmitters have been performed using 5 different species as *HRef*. They are NH_4^+ , $HP1^+$, $HS1^+$, $HT1^+$, and $HN1^+$ (Fig. 12). The results in this figure clearly show the

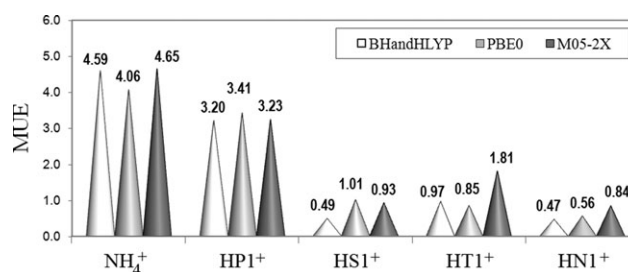


Figure 12. Mean unsigned errors (MUE) for the pK_a values of the studied neurotransmitters calculated with scheme **D** and different *HRef*.

importance of the *HRef* choice. When using ammonium as *HRef* the MUE are higher than 4 pK_a units, with all the tested methods, even larger than those obtained when using $HP1^+$, which are >3 pK_a units. This indicates that ammonium and primary amines are not adequate as *HRef* for neurotransmitters, despite of the fact that most of them are actually primary amines. Using a secondary amine ($HS1^+$) as *HRef* significantly improves the results which become closer to the experimental values, with MUE lower than, or very close to, 1 pK_a unit. Surprisingly using a tertiary amine ($HT1^+$) as *HRef* leads to better results than using a primary amine, despite of the fact that none of the studied neurotransmitters is a tertiary amine. Logically the best results are obtained when using a neurotransmitter ($HN1^+$) as *HRef*. In this case the MUE obtained with all the tested methods of calculation are lower than 1 pK_a unit. These results demonstrate how strong is the influence of the chosen *HRef* on the quality of pK_a calculations when scheme **D** is used.

The computational strategy recommended in this work for calculating pK_a 's of all kind of amines, the **F2**—M05-2X combination, is not only experiment-independent but also has been proven to produce accurate values. The MUE values (0.54 for all the studied amines, and 0.51 for neurotransmitters) are similar, or even lower, than the most accurate reported so far for this kind of compounds.^[3,37,43] It has the advantages of being computationally feasible for systems of relative large size, independent of experimental values, and free of fitting-based

corrections. It should be kept in mind, however, that this recommendation is valid within the cluster-continuum model, using the PCM model in conjunction with the UAHF cavity, and the strategy based on thermodynamic cycles. Changes in any of these aspects of the used methodology might lead to different outcomes.

Conclusions

Twenty-nine reaction schemes have been tested for the pK_a calculation of amines. We have used thermodynamic cycles, including up to four explicit water molecules for solvating the ionic species, and up to 8 water molecules per scheme. In addition we have tested 10 different levels of theory. They were used in combination with the polarizable continuum model (PCM), which is currently one of the most widely used for calculating pK_a 's in aqueous solution. The reliability of the tested approaches has been investigated first using a set of 10 amines, including different kind of them (ammonia, primary, secondary, and tertiary amines). A few viable strategies were identified and used to calculate the pK_a 's of 5 neurotransmitters.

It was found that amines are particularly challenging for pK_a calculations since computational strategies that perform well for one kind of amine can fail for other. The subset of reaction scheme—calculation method combinations that properly describe the whole set of tested amines are: **F2**—PBE0 (MUE = 0.39), **CN1**—PBE0 (MUE = 0.54), **D**—MP2 (MUE = 0.43), **CN1**—MP2 (MUE = 0.59), **F2**—M05-2X (MUE = 0.54), **C1**—M05-2X (MUE = 0.67).

For the studied neurotransmitters the best combinations were found to be **C1**—PBE0 (MUE = 0.41), **B**—BHandHLYP (MUE = 0.49), and **F2**—M05-2X (MUE = 0.51). Therefore the combination **F2**—M05-2X is the only one that give excellent results for the whole set of tested compounds. Moreover the corresponding MUE values are similar, or even lower, than the most accurate reported so far for this kind of compounds. The individual errors, for every studied species, were found to be lower than, or very close to, 1 unit of pK_a , with the maximum error equal to 1.09 pK_a units. In addition it has the advantages of being computationally feasible for relative large-sized systems, independent of experimental values, and free of fitting-based corrections. Accordingly, this particular approach is recommended for accurate pK_a calculations of amines in general.


An analysis of the importance of the chosen reference acid, when using the **D** scheme, has also been performed. It was found that the quality of pK_a calculations is strongly influenced by the *HRef* of choice. Moreover to obtain good results for neurotransmitters independently of the calculation method of choice, one of them must be used as *HRef*. Using methylamine, for example, lead to MUE >3 pK_a units, despite of the fact that 4 of the 5 studied neurotransmitters are also primary amines.

Acknowledgments

The authors thank Laboratorio de Visualización y Cómputo Paralelo at UAM, Iztapalapa for the access to its computer facilities. A.M.R.-Z. acknowledges CONACyT for Doctoral fellowship.

Keywords: acid constant · deprotonation · primary amines · secondary amines · tertiary amines · ammonia

How to cite this article: A. M. Rebollar-Zepeda, A. Galano, *Int. J. Quantum Chem.* **2012**, *112*, 3449–3460. DOI: 10.1002/qua.24048

 Additional Supporting Information may be found in the online version of this article.

- [1] M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus, G. C. Shields, *J. Am. Chem. Soc.* **2002**, *124*, 6421–6427.
- [2] C. Liao, M. C. Nicklaus, *J. Chem. Inf. Model.* **2009**, *49*, 2801–2812.
- [3] B. Kallies, R. Mitzner, *J. Phys. Chem. B* **1997**, *101*, 2959–2967.
- [4] G. S. Schuurmann, M. Cossi, V. Barone, J. Tomasi, *J. Phys. Chem. A* **1998**, *102*, 6706–6712.
- [5] C. O. Silva, M. A. Silva, M. A. C. Nascimento, *J. Phys. Chem. A* **1999**, *103*, 11194–11199.
- [6] A. M. Toth, M. D. Liptak, D. L. Phillips, G. C. Shields, *J. Chem. Phys.* **2001**, *114*, 4595–4606.
- [7] M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314–7319.
- [8] J. R. Pliego, J. M. Riveros, *J. Phys. Chem. A* **2002**, *106*, 7434–7439.
- [9] D. M. Chipman, *J. Phys. Chem. A* **2002**, *106*, 7413–7422.
- [10] K. R. Adam, *J. Phys. Chem. A* **2002**, *106*, 11963–11972.
- [11] G. A. A. Saracino, R. Improta, V. Barone, *Chem. Phys. Lett.* **2003**, *373*, 411–415.
- [12] A. Klamt, F. Eckert, M. Diedenhofen, M. E. Beck, *J. Phys. Chem. A* **2003**, *107*, 9380–9386.
- [13] G. I. Almerindo, D. W. Tondo, J. R. Pliego, *J. Phys. Chem. A* **2004**, *108*, 166–171.
- [14] A. M. Magill, K. J. Cavell, B. F. Yates, *J. Am. Chem. Soc.* **2004**, *126*, 8717–8724.
- [15] R. Vianello, Z. B. Maksic, *Eur. J. Org. Chem.* **2005**, *16*, 3571–3580.
- [16] T. N. Brown, N. Mora-Diez, *J. Phys. Chem. B* **2006**, *110*, 9270–9279.
- [17] M. Namazian, M. Zakery, M. R. Noorbala, M. L. Coote, *Chem. Phys. Lett.* **2008**, *451*, 163–168.
- [18] M. Alvarado-González, E. Orrantia-Borunda, D. Glossman-Mitnik, *J. Mol. Struct. Theochem* **2008**, *869*, 105.
- [19] J. Ho, M. L. Coote, *J. Chem. Theory Comput.* **2009**, *5*, 295–306.
- [20] Y. Song, J. Mao, M. R. Gunner, *J. Comput. Chem.* **2009**, *30*, 2231–2247.
- [21] E. Delgado, *J. Chem. Phys. Lett.* **2009**, *471*, 133–135.
- [22] R. Casanovas, J. Frau, J. Ortega-Castro, A. Salva, J. Donoso, F. Muñoz, *J. Mol. Struct. Theochem* **2009**, *912*, 5–12.
- [23] D. P. Dissanayake, R. Senthilnithy, *J. Mol. Struct. Theochem* **2009**, *910*, 93.
- [24] M. Uudsemaa, T. Kanger, M. Lopp, T. Tamm, *Chem. Phys. Lett.* **2010**, *485*, 83–86.
- [25] S. Rayne, K. Forest, *J. Mol. Struct. Theochem* **2010**, *949*, 60–69.
- [26] R. Casanovas, D. Fernandez, J. Ortega-Castro, J. Frau, J. Donoso, F. Muñoz, *Theor. Chem. Acc.* **2011**, *130*, 1–13.
- [27] B. Ghalami-Chooabar, H. Dezhmanpanah, P. Nikparsa, A. Ghiami-Shomami, *Int. J. Quantum Chem.* (in press).
- [28] J. Ho, M. L. Coote, *Theor. Chem. Acc.* **2010**, *125*, 3–21.
- [29] P. Claverie, J. P. Daudey, J. Langlet, B. Pullman, D. Piazzola, M. J. Huron, *J. Phys. Chem.* **1978**, *82*, 405–418.
- [30] L. C. G. Freitas, R. L. Longo, A. M. Simas, *J. Chem. Soc. Faraday Trans.* **1992**, *88*, 189–193.
- [31] J. R. Pliego, J. M. Riveros, *J. Phys. Chem. A* **2001**, *105*, 7241–7247.
- [32] A. M. Rebollar-Zepeda, T. Campos-Hernández, M. T. Ramírez-Silva, A. Rojas-Hernández, A. Galano, *J. Chem. Theory Comput.* **2011**, *7*, 2528–2538.
- [33] P. A. Gale, *Coord. Chem. Rev.* **2000**, *199*, 181–233.
- [34] P. A. Gale, *Coord. Chem. Rev.* **2001**, *213*, 79–128.
- [35] R. J. Fitzmaurice, G. M. Kyne, D. Douheret, J. D. Kilburn, *J. Chem. Soc. Perkin Trans. 1* **2002**, 841–864.
- [36] X. Huang, F. Zheng, P. A. Crooks, L. P. Dwozkin, C.-G. Zhan, *J. Am. Chem. Soc.* **2005**, *127*, 14401–14414.
- [37] V. S. Bryantsev, M. S. Diallo, W. A. Goddard, III, *J. Phys. Chem. A* **2007**, *111*, 4422–4430.

- [38] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, Revision E.01, Gaussian, Inc., Wallingford CT, **2004**.
- [39] (a) M. T. Cancès, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032–3041; (b) B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *106*, 5151–5158; (c) B. Mennucci, E. Cancès, J. Tomasi, *J. Phys. Chem. B* **1997**, *101*, 10506–10517; (d) J. Tomasi, B. Mennucci, E. Cancès, *J. Mol. Struct. Theochem.* **1999**, *464*, 211–226.
- [40] V. Barone, M. Cossi, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3210–3221.
- [41] A. E. Frisch, M. Frisch, G. W. Trucks, Gaussian 03 User's Reference. Gaussian Inc., Wallingford, **2003**.
- [42] D. M. Camaioni, C. A. Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, *109*, 10795–10797.
- [43] H. Lu, X. Chen, C. G. Zhan, *J. Phys. Chem. B* **2007**, *111*, 10599–10605.

Received: 26 December 2011
Revised: 18 January 2012
Accepted: 24 January 2012
Published online on 28 March 2012



Quantum Mechanical Based Approaches for Predicting pKa values of Carboxylic Acids: Evaluating the Performance of Different Strategies

Journal:	<i>RSC Advances</i>
Manuscript ID	Draft
Article Type:	Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Rebollar-Zepeda, Aida; Universidad Autonoma Metropolitana Iztapalapa Galano, Annia; Universidad Autonoma Metropolitana-Iztapalapa, Departamento de Quimica
Subject area & keyword:	Quantum & theoretical < Physical

Quantum Mechanical Based Approaches for Predicting pK_a values of Carboxylic Acids: Evaluating the Performance of Different Strategies

Aida Mariana Rebollar-Zepeda and Annia Galano*

Departamento de Química, División de Ciencias Básicas e Ingeniería. Universidad Autónoma Metropolitana-Iztapalapa, Av San Rafael Atlixco No.186, Col.Vicentina C.P.09340, México D.F.

Abstract

The performance of different computational protocols for predicting pK_a values of carboxylic acids, in aqueous solution, has been evaluated by comparison with experimental data. A set of 14 carboxylic acids was used to that purpose, including both aliphatic and aromatic species, as well as molecules of different chemical complexity. Two general strategies were explored, namely the reaction schemes design (RSD) and the fitting parameters (FP) methods. Within RSD, the reaction scheme chosen to represent the deprotonation process is the aspect influencing the most the agreement with the experiments. Other aspects, with modest but not negligible influence, are the way in which the solvation energies are calculated, and the used density functional theory (DFT) approach. The best results within the RSD strategy were found for the E3 and EN3 reaction schemes and the PBE0 functional, which led to mean unsigned errors (MUE) around 0.8 pK_a units. However, the FP method overperforms RSD, with MUE values below 0.3 pK_a units for most of the tested DFT approaches. Accordingly, the FP method is recommended for estimating pK_a values of carboxylic acids for which this information is still unknown, especially in combination with the PBE0 functional (MUE=0.26) and the SMD solvent model.

Keywords: acid constants; pK_a calculation; correlation; mean unsigned error; accuracy

* To whom correspondence should be addressed. E-mail: agal@xanum.uam.mx

Introduction

Acid dissociation constants (K_a) are important descriptors that quantify the tendency of chemical species to donate a proton. They are frequently expressed as pK_a , which are defined as $-\log(K_a)$ and rule the proportion of protonated and deprotonated forms of such species at a particular pH . Thus, accurate estimations of pK_a values are of crucial importance for numerous areas of knowledge including chemistry, biology, food science, medicine, and pharmacology. Despite of the fact that there several experimental techniques that have been proven to be successful for that purpose, it sometimes becomes a difficult task. Some paradigmatic examples are very weak or very strong acids, as well as short-living intermediates.[1] Accordingly, a huge amount of efforts have been devoted to improve the accuracy of pK_a values obtained from theoretical approaches,[1-21] albeit it still remains as a very challenging task. The difficulties, failures and successes of such strategies have been thoroughly reviewed by Ho and Coote,[22] who concluded that we are still far from achieving ideal results. In fact mean absolute deviations smaller than 2 pK_a units are currently accepted as reasonably accurate.[22]

Apparently, the largest contributions to the errors in pK_a calculations are caused by two aspects involved in the computational protocols. The first one is that acid-base equilibria involve charged species that usually are at the product side ($HA \rightleftharpoons H^+ + A^-$). Since errors arising from calculations, especially in solution, are typically larger for ionic than for non-charged species, the most significant errors do not cancel out when energies of reactions are calculated. Consequently the accuracy of such relative energies is no as predictable as for processes where the charged species are present at both sides of the equilibrium. The second main source of errors is the way in which the solvent is modeled. The most frequently used strategy for modeling solvation is using dielectric continuum models (DCM), which might be not good enough when short-range solute-solvent interactions are significant. In addition there is a wide variety of DCM that can be used for modeling solvation. Thus, the quality of the result may also depend on the DCM choice.

One way to deal with the first issue is by designing reaction schemes that maximize errors cancelation. For the second issue a viable solutions is to include some explicit solvent molecules in the vicinity of the solute, in conjunction with a DCM. The reaction schemes

design (RSD) approach is flexible on how the deprotonation process is formulated and allows simultaneously using the hybrid solvent model. Thus, it seems as an appealing option to produce accurate calculated pK_a values. It has been previously demonstrated to be successful for predicting pK_a s of chemical compounds belonging to the amine and phenol families.[23, 24] Within this strategy, the most attractive schemes are the experiment-independent ones. In addition, it usually requires only a few explicit water molecules for the hybrid solvent model, which guarantees that it can be carried out at reasonable computational costs for relatively large-sized systems. A drawback of the RSD strategy is that the best reaction scheme may change from one family of chemical compound to another. Therefore it is a “divide and conquer” kind of strategy that needs to be tested separately for each chemical group. This is mainly due to the fact that for each of them the number and kind of charged species may be different, and not necessarily at the same side of the equilibrium. For example the best reaction scheme for the protonated amine family is different from the best reaction scheme for neutral phenols.[23, 24]

Another strategy that circumvents the main error sources in pK_a calculations is the fitting parameters (FP) method.[25-34] It is based on using a standard linear regression fit for the calculated energy difference (between each acid and its conjugated base) and the corresponding experimental pK_a values. The two fitted parameters obtained from this regression are, in turn, used to calculate the pK_a of the molecules of interest. Within this strategy most of the errors in energy calculations are absorbed into the fitting parameters. In addition, it was demonstrated that in the particular case of the FP method, including explicit water molecules leads to larger errors than using only DCM.[29] This method presents the same disadvantage that RSD, i.e., each family of compound needs to be treated separately which means that a particular fit is necessary in each case.

In the present work the RSD and FP strategies are tested when used in conjunction with two different DCM approaches for a set carboxylic acids. In addition, two different ways for obtaining solvation energies have been tested. The influence of the electronic structure method chosen for the calculations has also been investigated, mainly within the density functional theory (DFT) framework, to assure that they are computationally feasible for relatively large molecules. The main purpose of this investigation is to identify the computational strategies most suitable for accurately predicting pK_a values of carboxylic

acids, at reasonable computational costs. By comparisons with the available experimental data, the strategies leading to deviations ≤ 1 pK_a unit are identified and recommended for obtaining reliable pK_a values from theoretical calculations.

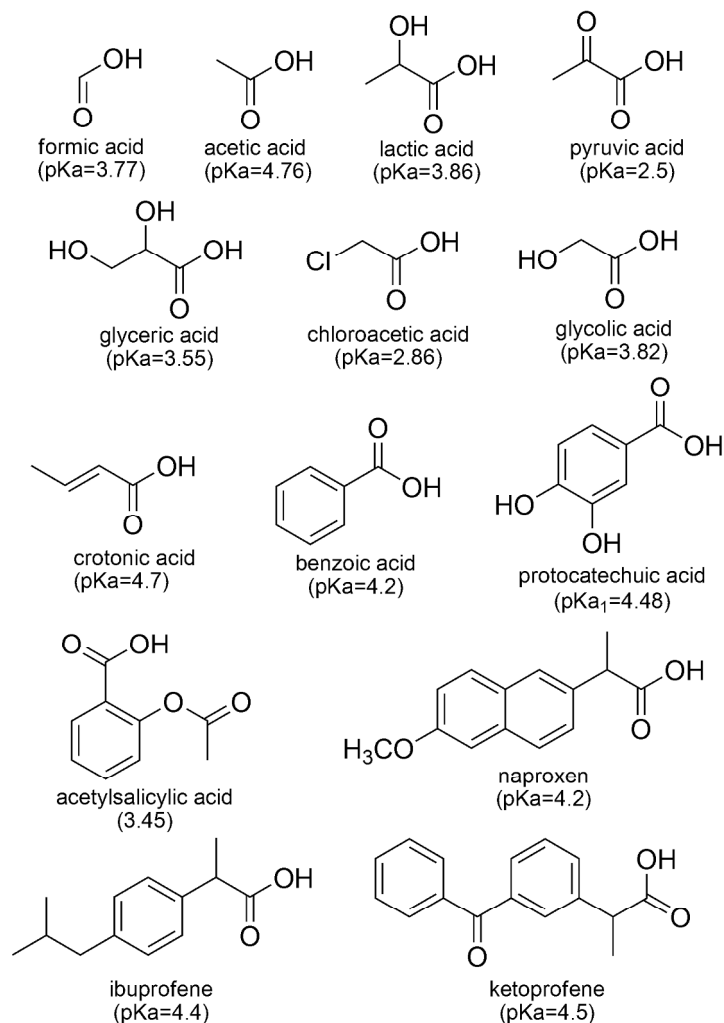
Computational Details

Full geometry optimizations and frequency calculations were performed for all the species involved in the studied acid-base equilibria without imposing any symmetry constraints. The electronic structure methods tested in this work are the DFT approximations BLYP, B3LYP, BHandHLYP, PBE, PBE0, PW91, BMK, TPSS, and M05-2X. All of them were used in conjunction with the 6-311++G(d,p) basis set, and the local minima were identified by the absence of imaginary frequencies. All the electronic calculations were carried out with the packages of programs Gaussian 03 [35] and Gaussian 09 [36].

Two RSD strategies were tested. They differ in the DCM of choice and in how the solvation energy is obtained. For RSD1 the stationary points were first modeled (geometry optimizations and frequency calculations) in gas phase, and solvent effects were included *a posteriori* by single point calculations using the polarizable continuum model, specifically the integral-equation-formalism (IEF-PCM).[37-40] Relative Gibbs free energies in solution (ΔG_s) are then computed using the Hess law and thermodynamic cycles, as the sum of the corresponding gas-phase free energy (ΔG_g) and the free energy of solvation (ΔG_{solv}). More details on this procedure can be found elsewhere.[23, 24] For the RSD2 strategy, on the other hand, all the geometry optimizations and frequency calculations were performed in solution using the SMD continuum model,[41] thus the ΔG_s values, for each species, are directly obtained.

The set of carboxylic acids studied in this work includes both simple, mono-functional, as well as more complex, multi-functional, species (Scheme 1). The reaction schemes used in both RSD strategies are shown in Table 1. In the particular case of scheme A, it involves the proton and it is known that computational methods poorly reproduce the solvation energies of this particular species. Therefore the $\Delta G_g(H^+)$ and $\Delta G_{solv}(H^+)$ values have been derived from experiments. In this work we have used $\Delta G_g(H^+) = -4.39$ kcal/mol, and $\Delta G_{solv}(H^+) = -265.89$ kcal/mol, based on the recommendation of Camaioni and

Schwerdtfeger,[42] which in aqueous solution leads to $\Delta G_s(\text{H}^+) = -270.28$ kcal/mol. These values, as well as all the data reported here, correspond to 1M standard state. Schemes A and D are the most widely used nowadays. The reaction scheme labeled here as D, is commonly referred to as the isodesmic method, the proton exchange method, or the relative method; while scheme A is usually referred to as the direct method.

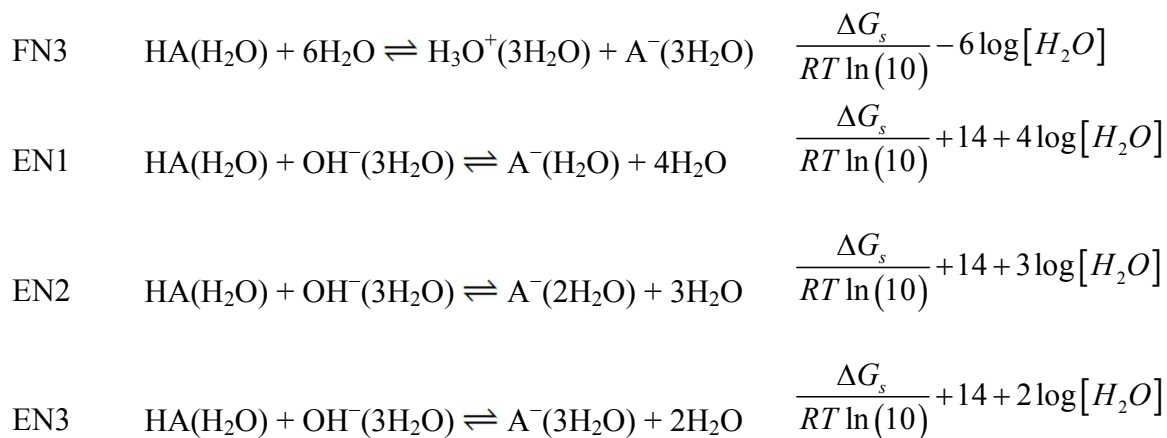


Scheme 1. Carboxylic acids studied in this work.

Table 1. Reaction schemes and the corresponding expressions to directly calculate pK_a values from ΔG_s .

Label	Reaction scheme	$pK_a =$
A	$\text{HA} \rightleftharpoons \text{H}^+ + \text{A}^-$	$\frac{\Delta G_s}{RT \ln(10)}$

B	$\text{HA} + \text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^-$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$
C1	$\text{HA} + 2\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^-(\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 2\log[H_2O]$
C2	$\text{HA} + 3\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^-(2\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 3\log[H_2O]$
C3	$\text{HA} + 4\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^-(3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 4\log[H_2O]$
D	$\text{HA} + \text{Ref}^- \rightleftharpoons \text{A}^- + \text{HRef}$	$\frac{\Delta G_s}{RT \ln(10)} + pKa(\text{HRef})$
E1	$\text{HA} + \text{OH}^-(3\text{H}_2\text{O}) \rightleftharpoons \text{A}^-(\text{H}_2\text{O}) + 3\text{H}_2\text{O}$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 3\log[H_2O]$
E2	$\text{HA} + \text{OH}^-(3\text{H}_2\text{O}) \rightleftharpoons \text{A}^-(2\text{H}_2\text{O}) + 2\text{H}_2\text{O}$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + 2\log[H_2O]$
E3	$\text{HA} + \text{OH}^-(3\text{H}_2\text{O}) \rightleftharpoons \text{A}^-(3\text{H}_2\text{O}) + \text{H}_2\text{O}$	$\frac{\Delta G_s}{RT \ln(10)} + 14 + \log[H_2O]$
F1	$\text{HA} + 4\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+(3\text{H}_2\text{O}) + \text{A}^-$	$\frac{\Delta G_s}{RT \ln(10)} - 4\log[H_2O]$
F2	$\text{HA} + 5\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+(3\text{H}_2\text{O}) + \text{A}^-(\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 5\log[H_2O]$
F3	$\text{HA} + 6\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+(3\text{H}_2\text{O}) + \text{A}^-(2\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 6\log[H_2O]$
F4	$\text{HA} + 7\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+(3\text{H}_2\text{O}) + \text{A}^-(3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 7\log[H_2O]$
CN1	$\text{HA}(\text{H}_2\text{O}) + \text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^-(\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - \log[H_2O]$
CN2	$\text{HA}(\text{H}_2\text{O}) + 2\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^-(2\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 2\log[H_2O]$
CN3	$\text{HA}(\text{H}_2\text{O}) + 3\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+ + \text{A}^-(3\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 3\log[H_2O]$
FN1	$\text{HA}(\text{H}_2\text{O}) + 4\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+(3\text{H}_2\text{O}) + \text{A}^-(\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 4\log[H_2O]$
FN2	$\text{HA}(\text{H}_2\text{O}) + 5\text{H}_2\text{O} \rightleftharpoons \text{H}_3\text{O}^+(3\text{H}_2\text{O}) + \text{A}^-(2\text{H}_2\text{O})$	$\frac{\Delta G_s}{RT \ln(10)} - 5\log[H_2O]$



For the FP strategies the mono-functional molecules were used as the training set (TS) to obtain the empirical parameters m and C_0 , from a standard linear regression fit, according to:

$$pK_a^{\text{exp}} = m \Delta G_{s(\text{BA})} + C_0 \quad (1),$$

where $\Delta G_{s(\text{BA})}$ represents the difference in Gibbs free energy, in aqueous solution, between each mono-functional acid (A) and its corresponding conjugated base (B):

$$\Delta G_{s(\text{BA})} = G_{s(\text{B})} - G_{s(\text{A})} \quad (2),$$

It was calculated at 298.15 K and expressed in kcal/mol. The two parameters (m and C_0) were used to calculate the pK_a values of the multi-functional carboxylic acids (test set) as:

$$pK_a^{\text{calc}} = m \Delta G_{s(\text{BA})}^{\text{TestSet}} + C_0 \quad (3)$$

This strategy was carried out using SMD to mimic the solvent, without including any explicit water molecule.

The accuracy of both approaches, RSD and FP, was estimated by comparisons with the experimental pK_a values of the studied carboxylic acids. The deviation from experiments was expressed as mean unsigned errors (MUE).

Results and Discussion

The MUE obtained for the investigated set of carboxylic acids using the RSD1 and RSD2 approaches are all summarily shown in Figure 1. In addition, the values of those combinations (reaction scheme / DFT approximation) with at least one MUE lower than 2 pK_a units are reported in Table 2. Figure 1 clearly shows that the reaction scheme used for

representing the acid-base equilibria influences the agreement with the experiments to a larger extent than the DFT of choice, regardless of the way in which the solvation energies are calculated. However, the three calculation aspects explored here (reaction scheme, strategy for calculating the solvation energy, DFT approach) influence the quality of the pK_a calculations for carboxylic acids.

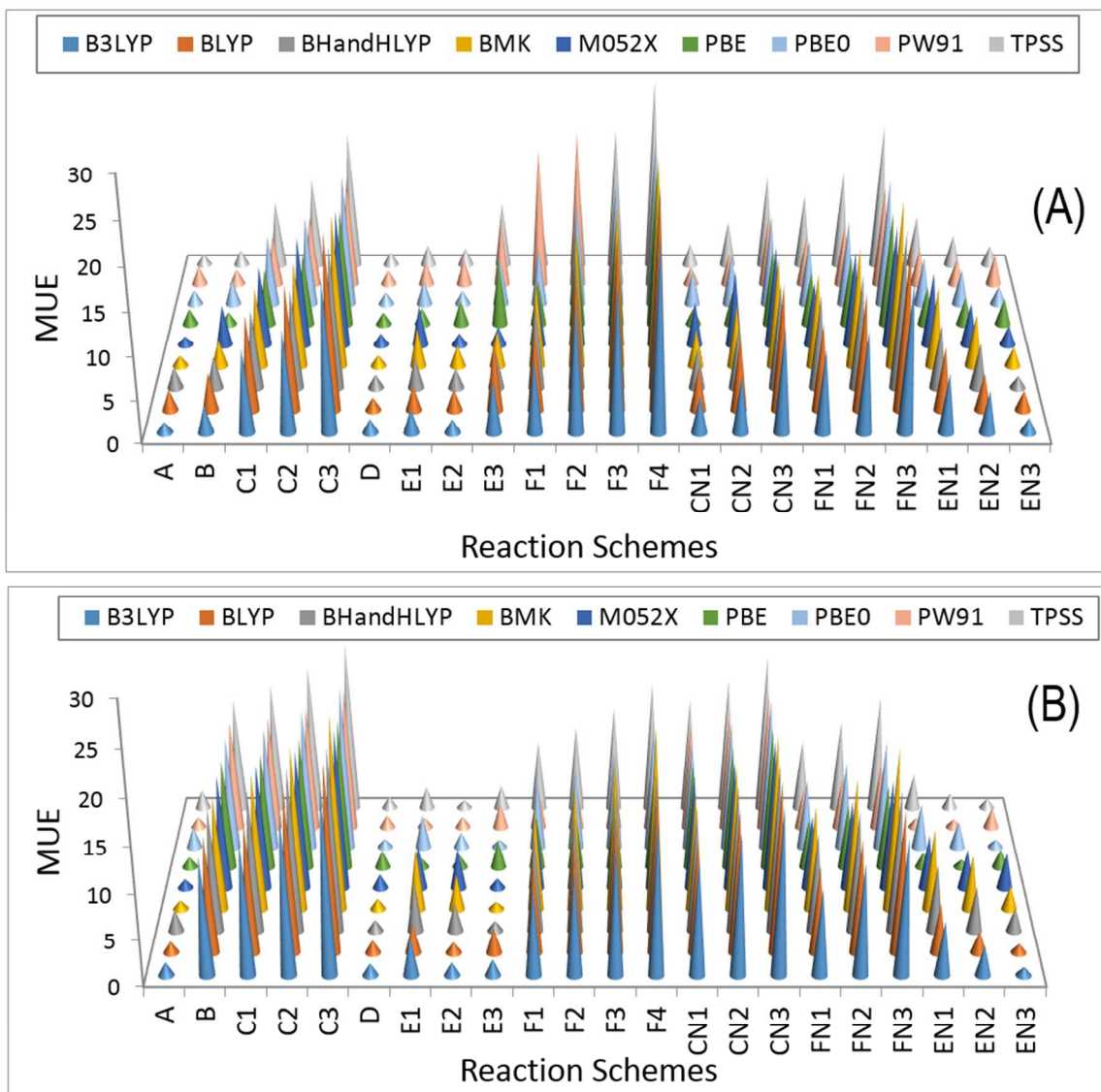


Figure 1. Mean unsigned errors (MUE) obtained for pK_a values calculated with the RSD strategy: (A) RSD1 and (B) RSD2.

Table 2. Combinations reaction scheme / DFT approximation with at least one MUE value lower than $2 pK_a$ units.

	B3LYP	BLYP	BH&HLYP	BMK	M05-2X	PBE	PBE0	PW91	TPSS
RSD1									
A	1.32	2.40	2.44	1.24	1.03	1.83	1.64	2.14	1.26
B	3.25	4.37	3.67	3.08	4.67	1.45	2.87	1.78	1.85
D	1.65	1.48	1.53	1.28	1.36	1.41	1.40	1.64	1.58
E1	2.81	2.84	3.53	4.45	4.60	1.95	2.81	2.63	2.46
E2	1.61	2.44	2.26	2.43	1.39	2.81	1.62	3.04	2.07
CN1	4.34	5.02	4.66	4.35	5.11	1.94	4.13	2.09	2.56
RSD2									
A	1.47	1.27	2.35	0.98	0.98	1.15	2.28	1.17	2.04
D	1.34	1.39	1.22	1.18	1.54	1.90	0.92	1.39	1.29
E1	4.14	3.27	5.42	6.75	4.75	1.06	3.65	1.05	2.38
E2	1.44	1.17	3.02	3.83	4.44	1.35	1.67	1.21	0.88
E3	1.80	2.74	1.12	0.84	1.07	3.02	0.85	2.81	2.51
EN1	5.96	5.37	7.24	8.87	5.93	1.71	4.88	1.08	3.78
EN2	3.25	2.43	4.84	5.94	4.28	0.84	2.89	0.92	1.65
EN3	0.93	0.99	2.53	2.58	3.93	2.05	0.87	2.22	1.06

When using the RSD1 approach, the reaction schemes that –in general– lead to the best agreement with the experimental data are A, D and E2, albeit none of the obtained MUE values are lower than $1 pK_a$ unit. For reaction scheme A, six DFT approaches produce $MUE < 2$, with the lowest values corresponding to the functionals M05-2X (1.03), BMK (1.24) and TPSS (1.26). For reaction scheme E2 only three DFT approaches lead to $MUE < 2$, i.e, M05-2X ($MUE=1.39$), B3LYP ($MUE=1.61$) and PBE0 ($MUE=1.62$). On the other hand, when using the reaction scheme D, all the obtained MUE values are below the $2 pK_a$ units threshold. The best results in this case were obtained with BMK (1.28), PBE0 (1.40) and PBE (1.41).

The agreement with the experimental data improves for the RSD2 approach, compared to RSD1. This is reflected in both the MUE values and the number of combinations (reaction scheme / DFT approximation) with $MUE < 2$. Contrary to what was already discussed for RSD1, several MUE values lower than $1 pK_a$ unit were obtained with RSD2 approach. In

this case the best agreement with the experimental data arise from using the reaction schemes A, D, E3 and EN3. Six DFT approaches lead to $MUE < 2$, when using scheme A, with the lowest values corresponding to BMK and M05-2X (0.98 in both cases), which are also the only two MUE values lower than 1 pK_a unit for this scheme. For the reaction scheme E3, five of the tested DFT approaches led to $MUE < 2$, while only two produce $MUE < 1$. They are BMK (MUE=0.84) and PBE0 (MUE=0.85). For the reaction scheme EN3, MUE values lower than 2 and 1 pK_a units were found for 4 and 3 DFT approaches, respectively. The best agreement with the experimental data was obtained for PBE0 (MUE=0.87) and B3LYP (MUE=0.93). Similarly to what was found for the RSD1 approach, when using the RSD2 approach all the MUE values obtained with the reaction scheme D are below the 2 pK_a units threshold. However, the MUE obtained with RSD2 and PBE0 is lower than 1 pK_a unit (0.92).

The results obtained with the FP method are presented in Table 3. For all the tested DFT approaches the standard linear regression fit correlation coefficients (R^2) were found to be around 0.8. However, the agreement of the calculated pK_a values with the experimental data is very good, with $MUE < 0.45$ in all the cases. Moreover, six DFT approaches produce $MUE < 0.3$. They are PBE (0.23), PBE0 (0.26), BLYP (0.26), PW91 (0.28), TPSS (0.29), and BMK (0.29). It is also interesting to note that the values of the m parameter are very similar with all the tested DFT approaches, ranging from 0.223 to 0.273. On the contrary, there is a significant variation in the values of C_0 (from -57.488 to -71.478).

Table 3. Parameters m y C_0 , linear correlation coefficient (R^2), and mean unsigned error (MUE) for the pK_a values calculated with the FP method.

	m	C_0	R^2	MUE
B3LYP	0.231	-59.831	0.840	0.32
BLYP	0.223	-57.501	0.864	0.26
BHandHLYP	0.260	-68.336	0.870	0.41
BMK	0.273	-71.478	0.841	0.29
M05-2X	0.257	-66.694	0.762	0.31
PBE	0.227	-58.702	0.792	0.23
PBE0	0.240	-62.765	0.788	0.26
PW91	0.223	-57.488	0.768	0.28

TPSS	0.224	-58.266	0.792	0.29
------	-------	---------	-------	------

To better understand the dissimilar dependence of pK_a values, calculated using the FP method, with the DFT approach a more detailed analysis is provided next. For the conventional deprotonation reaction ($HA \rightleftharpoons H^+ + A^-$), the expression to directly calculate pK_a values from ΔG_s is:

$$pKa = \frac{\Delta G_s}{RT \ln(10)} \quad (4),$$

where ΔG_s can be expressed as:

$$\Delta G_s = \Delta G_{s(BA)} + \Delta G_s(H^+) \quad (5)$$

In addition, the linear regression fit used to obtain the m and C_0 parameters is equivalent to including a scaling factor (${}^s f$) that corrects for the systematic error of any computational method (for similar deprotonation reactions, i.e., for each particular family of chemical compounds). Therefore, equation (1) can be rewritten as:

$$pKa^{\text{exp}} = \frac{{}^s f \Delta G_{s(BA)}}{RT \ln(10)} + \frac{{}^s f \Delta G_s(H^+)}{RT \ln(10)} \quad (6)$$

This means that:

$$m = \frac{{}^s f}{RT \ln(10)} \quad (7),$$

and

$$C_0 = \frac{{}^s f \Delta G_s(H^+)}{RT \ln(10)} \quad (8)$$

Using equations (7) and (8) is now possible to estimate the scaling factor and the value of $\Delta G_s(H^+)$ for each DFT approach used in this work, according to:

$${}^s f = mRT \ln(10) \quad (9)$$

$$\Delta G_s(H^+) = \frac{C_0 RT \ln(10)}{{}^s f} = \frac{C_0}{m} \quad (10)$$

Their values are reported in Table 4. The ${}^s f$ values are similar and about 0.3, which supports the idea that ${}^s f$ is mainly ruled by the chemical nature of the family of compounds for which

the acid-base equilibria is investigated. On the other hand, $\Delta G_s(H^+)$ not only depend on the chemical family but also on the electronic structure method of choice and, in principle can be used to identify which of them leads to a better agreement with the experiment.

Table 4. Scaling factors ($^s f$) and Gibbs free energies of the proton in aqueous solution ($\Delta G_s(H^+)$, kcal/mol) obtained for the deprotonation reactions of carboxylic acids with different DFT methods.

	$^s f$	$\Delta G_s(H^+)$
B3LYP	0.315	-259.0
BLYP	0.304	-257.9
BHandHLYP	0.355	-262.8
BMK	0.372	-261.8
M05-2X	0.351	-259.5
PBE	0.310	-258.6
PBE0	0.327	-261.5
PW91	0.304	-257.8
TPSS	0.306	-260.1

However, the measured values of $\Delta G_{sol}(H^+)$ significantly vary, depending on the conditions of the particular experiment carried out. The variations are rather significant with the reported $\Delta G_{sol}(H^+)$ values ranging from -254 to -265.89 kcal/mol.[22, 31] Consequently, the $\Delta G_s(H^+)$ values derived from experiments range from -258.4 to -270.3 kcal/mol. The values obtained here range from -257.8 to -262.8 kcal/mol. All the DFT approaches, except BLYP, produce $\Delta G_s(H^+)$ values within the experimental range, with BHandHLYP, PBE0 and BMK leading to the best agreement with the value recommended by Camaioni and Schwerdtfeger.[42] Among these three functionals PBE0 is the one leading to the lowest MUE, thus it is recommended for being used in combination with the FP method to obtain accurate pK_a values for carboxylic acids.

Analyzing the gathered data altogether it can be stated that the MUE values found when using the FP method are significantly lower than any of those obtained with the RSD strategies. This indicates that the FP method is a better choice than RSD to estimate pK_a values of carboxylic acid for which this information is not available. However, it has two

main drawbacks: (i) it strongly depend on experimental data, and (ii) it requires to perform extra calculations for a training set of molecules, using the same calculation methods (DFT approach, basis set, and solvent model) that are used for the molecules of interest. An alternative to overcome these two disadvantages, for carboxylic acids, is using the RSD2 strategy in conjunction with the E3 or EN3 reaction schemes and the PBE0 functional. Albeit it was found that this strategy leads to higher MUE values than the FP method, the discrepancies with the experiments are still below 1 pK_a unit (~ 0.8).

Conclusions

It was found that the agreement with the experimental data is better with RSD2 than with RSD1 for pK_a calculations of carboxylic acids. Therefore, at least for this chemical family it is better to calculate solvation energies directly (with SMD) than using thermodynamic cycles and PCM. In particular, we recommend using RSD2 in conjunction with the E3 or EN3 reaction schemes and the PBE0 functional. This combination leads to MUE values around 0.8 pK_a units.

The performance of the FP method was better than that of the RSD strategy. The agreement of the calculated pK_a values with the experimental data in this case is very good, with MUE < 0.45 in all the cases. Moreover, six DFT approaches produce MUE < 0.3 . In addition, according to the $\Delta G_s(H^+)$ values arising from the FP method, the best performing functionals are BHandHLYP, PBE0 and BMK. Among these three functionals PBE0 is the one leading to the lowest MUE (0.26), thus it is recommended for being used in combination with the FP method to obtain accurate pK_a values for carboxylic acids.

According to the gathered data, we recommend using the FP method for estimating pK_a values of carboxylic acids for which this information is still unknown, especially in combination with the PBE0 functional and the SMD solvent model.

Acknowledgements

The authors thank Laboratorio de Visualización y Cómputo Paralelo at UAM - Iztapalapa for the access to its computer facilities. A. M. R.-Z. acknowledges CONACyT for Doctoral fellowship.

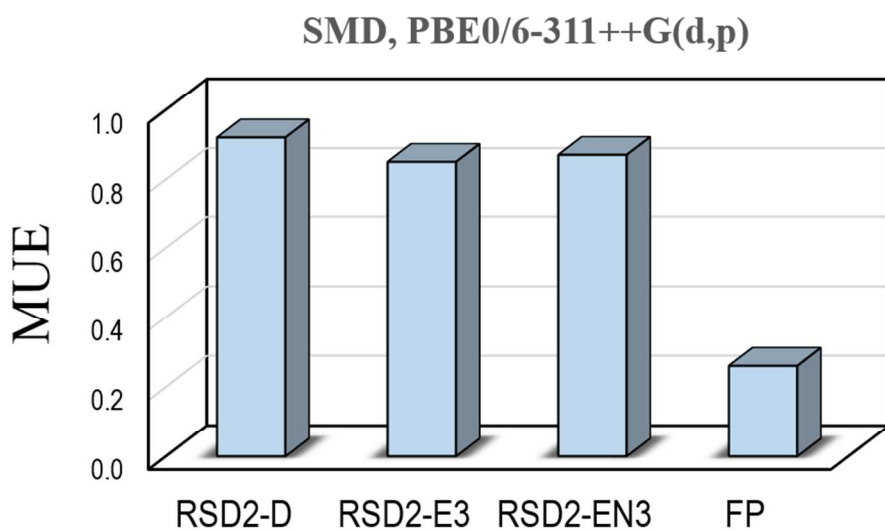
References

- [1] J.R. Pliego Jr, J.M. Riveros, Theoretical calculation of pKa using the cluster-continuum model, *J Phys Chem A*, 106 (2002) 7434-7439.
- [2] G. Schüürmann, M. Cossi, V. Barone, J. Tomasi, Prediction of the pKa of carboxylic acids using the ab initio continuum-solvation model PCM-UAHF, *J Phys Chem A*, 102 (1998) 6706-6712.
- [3] C.O. da Silva, E.C. da Silva, M.A.C. Nascimento, Ab Initio Calculations of Absolute pKa Values in Aqueous Solution I. Carboxylic Acids, *J Phys Chem A*, 103 (1999) 11194-11199.
- [4] A.M. Toth, M.D. Liptak, D.L. Phillips, G.C. Shields, Accurate relative pKa calculations for carboxylic acids using complete basis set and Gaussian-n models combined with continuum solvation methods, *J Chem Phys*, 114 (2001) 4595-4606.
- [5] M.D. Liptak, G.C. Shields, Accurate pKa Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined with CPCM Continuum Solvation Methods, *J Am Chem Soc*, 123 (2001) 7314-7319.
- [6] M.D. Liptak, K.C. Gross, P.G. Seybold, S. Feldgus, G.C. Shields, Absolute pKa Determinations for Substituted Phenols, *J Am Chem Soc*, 124 (2002) 6421-6427.
- [7] D.M. Chipman, Computation of pKa from Dielectric Continuum Theory, *J Phys Chem A*, 106 (2002) 7413-7422.
- [8] K.R. Adam, New Density Functional and Atoms in Molecules Method of Computing Relative pKa Values in Solution, *The Journal of Physical Chemistry A*, 106 (2002) 11963-11972.
- [9] G.A.A. Saracino, R. Improta, V. Barone, Absolute pKa determination for carboxylic acids using density functional theory and the polarizable continuum model, *Chem Phys Lett*, 373 (2003) 411-415.
- [10] A. Klamt, F. Eckert, M. Diedenhofen, M.E. Beck, First Principles Calculations of Aqueous pKa Values for Organic and Inorganic Acids Using COSMO-RS Reveal an Inconsistency in the Slope of the pKa Scale, *J Phys Chem A*, 107 (2003) 9380-9386.
- [11] G.I. Almerindo, D.W. Tondo, J.R. Pliego, Ionization of Organic Acids in Dimethyl Sulfoxide Solution: A Theoretical Ab Initio Calculation of the pKa Using a New Parametrization of the Polarizable Continuum Model, *The Journal of Physical Chemistry A*, 108 (2004) 166-171.
- [12] A.M. Magill, K.J. Cavell, B.F. Yates, Basicity of Nucleophilic Carbenes in Aqueous and Nonaqueous Solvents Theoretical Predictions, *J Am Chem Soc*, 126 (2004) 8717-8724.
- [13] R. Vianello, Z.B. Maksić, Strong acidity of some polycyclic aromatic compounds annulated to a cyclopentadiene moiety and their cyano derivatives - A density functional B3LYP study, *Eur J Org Chem*, (2005) 3571-3580.
- [14] M. Namazian, M. Zakery, M.R. Noorbala, M.L. Coote, Accurate calculation of the pKa of trifluoroacetic acid using high-level ab initio calculations, *Chem Phys Lett*, 451 (2008) 163-168.
- [15] J. Ho, M.L. Coote, pKa Calculation of Some Biologically Important Carbon Acids - An Assessment of Contemporary Theoretical Procedures, *J Chem Theory Comput*, 5 (2009) 295-306.

- [16] Y. Song, J. Mao, M.R. Gunner, MCCE2: Improving protein pKa calculations with extensive side chain rotamer sampling, *J. Comput. Chem.*, 30 (2009) 2231-2247.
- [17] E.J. Delgado, DFT calculation of pKa's for dimethoxypyrimidinylsalicylic based herbicides, *Chem Phys Lett*, 471 (2009) 133-135.
- [18] R. Casanovas, J. Frau, J. Ortega-Castro, A. Salvà, J. Donoso, F. Muñoz, Absolute and relative pKa calculations of mono and diprotic pyridines by quantum methods, *J Mol Struct: THEOCHEM*, 912 (2009) 5-12.
- [19] D.P. Dissanayake, R. Senthilnithy, Thermodynamic cycle for the calculation of ab initio pKa values for hydroxamic acids, *J Mol Struct: THEOCHEM*, 910 (2009) 93-98.
- [20] M. Uudsemaa, T. Kanger, M. Lopp, T. Tamm, pKa calculation for monoprotinated bipiperidine, bimorpholine and their derivatives in H₂O and MeCN, *Chem Phys Lett*, 485 (2010) 83-86.
- [21] S. Rayne, K. Forest, Theoretical studies on the pKa values of perfluoroalkyl carboxylic acids, *J Mol Struct: THEOCHEM*, 949 (2010) 60-69.
- [22] J. Ho, M.L. Coote, A universal approach for continuum solvent pKa calculations: Are we there yet?, *Theor Chem Acc*, 125 (2009) 3-21.
- [23] A.M. Rebollar-Zepeda, A. Galano, First principles calculations of pK a values of amines in aqueous solution: Application to neurotransmitters, *Int J Quant Chem*, 112 (2012) 3449-3460.
- [24] A.M. Rebollar-Zepeda, T. Campos-Hernández, M.T. Ramírez-Silva, A. Rojas-Hernández, A. Galano, Searching for computational strategies to accurately predict pK as of large phenolic derivatives, *J Chem Theor Comput*, 7 (2011) 2528-2538.
- [25] J.J. Klicić, R.A. Friesner, S.-Y. Liu, W.C. Guida, Accurate Prediction of Acidity Constants in Aqueous Solution via Density Functional Theory and Self-Consistent Reaction Field Methods, *The Journal of Physical Chemistry A*, 106 (2002) 1327-1335.
- [26] F. Eckert, A. Klamt, Accurate prediction of basicity in aqueous solution with COSMO-RS, *Journal of Computational Chemistry*, 27 (2006) 11-19.
- [27] S. Zhang, J. Baker, P. Pulay, A reliable and efficient first principles-based method for predicting pKa Values. 1. Methodology, *J Phys Chem A*, 114 (2010) 425-431.
- [28] S. Zhang, J. Baker, P. Pulay, A reliable and efficient first principles-based method for predicting pKa values. 2. Organic acids, *J Phys Chem A*, 114 (2010) 432-442.
- [29] S. Zhang, A reliable and efficient first principles-based method for predicting pK a values. III. Adding explicit water molecules: Can the theoretical slope be reproduced and pK a values predicted more accurately?, *J. Comput. Chem.*, 33 (2012) 517-526.
- [30] S. Zhang, A reliable and efficient first principles-based method for predicting pKa values. 4. organic bases, *J. Comput. Chem.*, 33 (2012) 2469-2482.
- [31] T. Matsui, A. Oshiyama, Y. Shigeta, A Simple scheme for estimating the pKa values of 5-substituted uracils, *Chemical Physics Letters*, 502 (2011) 248-252.
- [32] T. Matsui, T. Baba, K. Kamiya, Y. Shigeta, An accurate density functional theory based estimation of pK a values of polar residues combined with experimental data: From amino acids to minimal proteins, *Phys. Chem. Chem. Phys.*, 14 (2012) 4181-4187.
- [33] R. Álvarez-Diduk, M.T. Ramírez-Silva, A. Galano, A. Merkoçi, Deprotonation mechanism and acidity constants in aqueous solution of flavonols: A combined experimental and theoretical study, *J Phys Chem B*, 117 (2013) 12347-12359.
- [34] T. Baba, T. Matsui, K. Kamiya, M. Nakano, Y. Shigeta, A density functional study on the pKa of small polyprotic molecules, *International Journal of Quantum Chemistry*, 114 (2014) 1128-1134.

- [35] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A.J. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 03, Gaussian, Inc., Wallingford, CT, USA, 2003.
- [36] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N.J. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Gaussian, Inc., Wallingford, CT, USA, 2009.
- [37] E. Cancès, B. Mennucci, J. Tomasi, A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics, *J Chem Phys*, 107 (1997) 3032-3041.
- [38] B. Mennucci, J. Tomasi, Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries, *J Chem Phys*, 106 (1997) 5151-5158.
- [39] B. Mennucci, E. Cancès, J. Tomasi, Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation, and Numerical Applications, *J Phys Chem B*, 101 (1997) 10506-10517.
- [40] J. Tomasi, B. Mennucci, E. Cancès, The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level, *J Mol Struct: THEOCHEM*, 464 (1999) 211-226.
- [41] A.V. Marenich, C.J. Cramer, D.G. Truhlar, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions, *J Phys Chem B*, 113 (2009) 6378-6396.
- [42] D.M. Camaioni, C.A. Schwerdtfeger, Comment on "accurate experimental values for the free energies of hydration of H⁺, OH⁻, and H₃O⁺", *J. Phys. Chem. A*, 109 (2005) 10795-10797.

Graphical Abstract



The FP method is recommended for estimating pK_a values of carboxylic acids for which this information is still unknown, especially in combination with the PBE0 functional (MUE=0.26) and the SMD solvent model.

México D.F., June 22nd, 2016.

Prof. Mike D Ward
Editor-in-chief, RSC Advances

Dear Professor Ward:

We are submitting the manuscript entitled: “Quantum Mechanical Based Approaches for Predicting pKa values of Carboxylic Acids: Evaluating the Performance of Different Strategies” by Aida Mariana Rebollar-Zepeda and Annia Galano, for your consideration for publication in RSC Advances.

This manuscript presents a systematic investigation on the performance of different computational protocols for predicting pKa values of carboxylic acids, in aqueous solution. The reliability of the calculated data has been evaluated by comparison with experimental values. A large diversity of computational protocols were tested, and the most promising ones were identified and recommended for future theoretical pKa estimations. Novel reaction schemes are proposed for the first time.

Potential reviewers:

- Joaquin Ortega-Castro. Universitat Illes Balears Departament de Química (Química-Física) Edifici Mateu Orfila i Rotger E-07071. Palma de Mallorca, e-mail: joaquin.castro@uib.es
- Nino Russo, Dipartimento di Chimica - University of Calabria. E-Mail: nrusso@unical.it
- Juan Frau Munar, Universitat Illes Balears Departament de Química (Química-Física) Edifici Mateu Orfila i Rotger E-07071. Palma de Mallorca, e-mail: juan.frau@uib.es

- Francisco Muñoz, Universitat Illes Balears Departament de Química (Química-Física) Edifici Mateu Orfila i Rotger E-07071. Palma de Mallorca, e-mail: dqufmi0@uib.es
- José P. Cerón-Carrasco, Department of Physical Chemistry, University of Murcia, Spain. E-Mail: jpceron@um.es

Sincerely yours,

Annia



Dr. Annia Galano
Departamento de Química,
División de Ciencias Básicas e Ingeniería.
Universidad Autónoma Metropolitana-Iztapalapa.
San Rafael Atlixco 186, Col. Vicentina. Iztapalapa.
C. P. 09340. México D. F. México.
Phone: (5255) 58044675
www.agalano.com
Researche ID: B-9809-2009
ORCID: 0000-0002-1470-3060

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Empirically Fitted Parameters for Calculating pKa Values with Small Deviations from Experiments Using a Simple Computational Strategy

Journal:	<i>Journal of Chemical Information and Modeling</i>
Manuscript ID	ci-2016-00310p.R1
Manuscript Type:	Article
Date Submitted by the Author:	15-Aug-2016
Complete List of Authors:	Galano, Annia; Universidad Autonoma Metropolitana, Departamento de Quimica Pérez-Gonzalez, Adriana; Universidad Autonoma Metropolitana, Quimica Castañeda-Arriaga, Romina ; Universidad Autonoma Metropolitana, Departamento de Quimica Muñoz-Rugeles, Leonardo; Universidad Nacional Autónoma de México, Departamento de Física y Química Teórica Mendoza-Sarmiento, Gabriela; Universidad Autonoma Metropolitana Iztapalapa, Quimica Romero-Silva, Antonio; Universidad Nacional Autonoma de Mexico, Facultad de Quimica Ibarra-Escutia, Agustin; Universidad Autonoma Metropolitana, Departamento de Quimica Rebollar-Zepeda, Aida; Universidad Autónoma Metropolitana -Iztapalapa , Departamento de Química León-Carmona, Jorge; Universidad Nacional Autónoma de México, Instituto de Investigaciones en Materiales Hernández-Olivares, Manuel ; Universidad Autónoma Metropolitana - Iztapalapa , Departamento de Química Alvarez-Idaboy, Juan Raul; Universidad Nacional Autonoma de Mexico, Facultad de Quimica

SCHOLARONE™
Manuscripts

Empirically Fitted Parameters for Calculating pK_a Values with Small Deviations from Experiments Using a Simple Computational Strategy

Annia Galano,^{1*} Adriana Pérez-González,² Romina Castañeda-Arriaga,¹ Leonardo Muñoz-Rugeles,³ Gabriela Mendoza-Sarmiento,¹ Antonio Romero-Silva,³ Agustín Ibarra-Escutia,¹ Aida Mariana Rebollar-Zepeda,¹ Jorge Rafael León-Carmona,³ Manuel Alejandro Hernández-Olivares,¹ Juan Raúl Álvarez-Idaboy^{3*}

¹ *Departamento de Química, División de Ciencias Básicas e Ingeniería. Universidad Autónoma Metropolitana-Iztapalapa, Av San Rafael Atlixco No.186, Col.Vicentina C.P.09340, México D.F.*

² *CONACYT - Departamento de Química, División de Ciencias Básicas e Ingeniería. Universidad Autónoma Metropolitana-Iztapalapa, Av San Rafael Atlixco No.186, Col.Vicentina C.P.09340, México D.F.*

³ *Facultad de Química, Departamento de Física y Química Teórica, Universidad Nacional Autónoma de México, México DF 04510, México.*

E-mail: agal@xanum.uam.mx, jidaboy@unam.mx

Abstract

Two empirically fitted parameters have been derived for 74 levels of theory. They allow fast and reliable pK_a calculations using only the Gibbs energy difference between an acid and its conjugated base in aqueous solution ($\Delta G_{s(BA)}$). The parameters were obtained by least-squares fits of $\Delta G_{s(BA)}$ vs. experimental pK_a values for phenols, carboxylic acids and amines using training sets of 20 molecules for each chemical family. Test sets of 10 molecules per family –completely independent from the training set– were used to verify the reliability of the fitting parameters method. It was found that, except for MP2, deviations from experiments are lower than 0.5 pK_a units. Moreover, mean unsigned errors lower than 0.35 pK_a units were found for the 98.6%, 98.6%, and 94.6% of the tested levels of theory for phenols, carboxylic acids and amines, respectively. The parameters estimated here are expected to facilitate computationally-based estimations of pK_a values of species for which this magnitude is still unknown, with uncertainties similar to the experimental ones. However, the present study deals only with molecules of modest complexity, thus the reliability of the FP method for more complex systems remains to be tested.

Introduction

Chemical compounds susceptible to be involved in acid-base equilibria are ubiquitous in nature and in daily human life. They are present not only within our body, but also in our diet and in many substances with pharmacological properties. The tendency of these compounds to deprotonate is quantified by their acid dissociation constants (K_a), which are frequently expressed as pK_a . The latter are defined as $-\log(K_a)$ and rule the proportion of protonated and deprotonated species at a particular pH . Thus, accurate estimations of pK_a values are of crucial importance for numerous areas of knowledge including chemistry, biology, food science, medicine, and pharmacology.

Despite of the fact that there are several experimental techniques that have been proven to be successful for estimating pK_a s, for some compounds it may become a difficult task. Some paradigmatic examples are very weak or very strong acids, as well as short-living intermediates.¹ Accordingly, a huge amount of efforts have been devoted to improve the accuracy of pK_a values obtained from theoretical approaches, based on electronic structure methods.¹⁻²¹ However, it still remains a very challenging task. The difficulties, failures and successes of such strategies have been thoroughly reviewed by Ho and Coote,²² who concluded that we are still far from achieving ideal results. In fact absolute deviations smaller than 2 units of pK_a are currently accepted as reasonably accurate.²²

Apparently, the largest contributions to the errors in pK_a calculations arise from two particular aspects of the computational protocols. The first one is that acid-base equilibria involve charged species that usually are at the product side ($HA \rightleftharpoons H^+ + A^-$). Since errors arising from calculations, especially in solution, are typically larger for ionic than for non-charged species, the most significant errors do not cancel out when energies of reactions are calculated. Consequently the accuracy of such relative energies is not as predictable as for reactions with species equally charged at both sides of the equilibrium. The second main source of errors is the way in which the solvent is modeled. The most frequently used strategy for this purpose is using a dielectric continuum model (DCM), which might not be good enough when short-range solute-solvent interactions are important, which is just the case of ions in aqueous solution. In addition, there is a wide variety of DCM approaches

1
2
3 that can be used for modeling solvation. Thus, the quality of the result may also depend on
4 the DCM of choice.

5
6
7 However, there is an strategy that circumvents the main error sources in pK_a calculations,²³⁻
8
9³² which is here referred to as the fitting parameters (FP) method. Generally speaking, it
10 consists of calculating an energy difference that involves the acid and its conjugated base
11 for a set of chemical compounds. Then this data is used to obtain a linear regression with
12 the corresponding experimental pK_a s. This procedure allows obtaining two empirically
13 fitted parameters that can –in turn– be used to estimate the pK_a values of the species of
14 interest, and should be carried out separately for each family of chemical compounds. In
15 general, the pK_a values calculated with the FP method are in good agreement with the
16 experimental data, the mean unsigned errors (MUE), root-mean square errors (RMSE) and
17 maximum absolute errors (MAE) are usually below 0.5, 0.5 and 1.0 pK_a units, respectively.

18
19
20 Therefore the FP method is a very appealing strategy that significantly improves the
21 accuracy of pK_a calculations, compared to other methodologies. However, this method
22 presents a major inconvenience, a very laborious task should be performed any time a pK_a
23 calculation is needed. Albeit, there are several reports of fitted parameters, they were
24 obtained with a particular combination of electronic structure method and basis sets.
25 Therefore, researchers interested in using a level of theory for which these parameters have
26 not been reported, face the necessity of performing numerous calculations –additional to
27 their own investigation– to use the FP method. This is a rather frequent case since there are
28 abundant chemical problems that are significantly influenced by acid-base equilibria. In
29 addition, a systematic assessment on the potential use of different levels of theory
30 (electronic structure method / basis set) to estimate pK_a values using the FP method has not
31 been performed yet.

32
33
34 Accordingly, the main goal of the present work is to perform such a systematic
35 investigation and provide a rather large list of parameters that can be later used to calculate
36 pK_a values. For that purpose 20 electronic structure methods and 4 basis sets were
37 considered, which are among the most frequently used ones. Their performance for the task
38 at hand was analyzed based on the corresponding MUE, RMSE and MAE values, and the
39 corresponding fitted parameters are reported for the first time. Two independent sets of
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 chemical compounds were used as training and test sets. Both include three families of
4 molecules: carboxylic acids, phenols and amines. The reported results are expected to
5 promote further use of the FP method in the estimation of pK_a values of species for which
6 this magnitude is still unknown.
7
8
9

10 11 12 **Computational Details**

13
14 The FP method has been previously used in different ways. For example, the linear
15 regressions used to obtain the fitted parameters have been performed using a variety of
16 calculated data, including pK_a ,²³ enthalpy,²⁵ and Gibbs energies.^{24, 30, 31} The solvent effects
17 have also been included using different models such as COSMO,^{24, 25} PCM³⁰ and SMD.³¹
18 Since changing any of these aspects may modify the values of the fitting parameters, and
19 the possible combinations are almost endless –when considering also the electronic
20 structure method and the basis set– it is important to explain in detail exactly how the FP
21 method was carried out here.
22
23
24
25
26
27

28
29 All the electronic calculations were carried out with the package of programs Gaussian
30 09.³³ Full geometry optimizations, without imposing any symmetry constrains, and
31 frequency calculations were performed for all the species involved in the studied acid-base
32 equilibria using the solvation model based on density (SMD).³⁴ Geometry optimizations
33 were performed in solution because optimized geometries can be significantly different
34 from those obtained in gas phase, particularly for phenolic compounds. At the same time, it
35 has been pointed out that the partition functions obtained when using dielectric continuum
36 solvent models are not reliable to calculate Gibbs energies in solution.³⁵ So far there is only
37 work demonstrating otherwise.³⁶ In that work it was demonstrated that optimizing
38 geometries with the SMD method, and using the corresponding partition functions, is a
39 correct approach for averaging over solute degrees of freedom when computing free
40 energies of solutes in solution. That is why the SMD method was chosen for the present
41 work.
42
43
44
45
46
47
48
49
50
51

52
53 Local minima were identified by the absence of imaginary frequencies. The electronic
54 structure methods tested in this work comprise 17 density functional theory (DFT)
55 approximations: PBE, PBE0, LC- ω PBE, TPSS, TPSSh, B3LYP, cam-B3LYP,
56
57
58
59
60

1
2
3 BHandHLYP, BMK, MPWB1K, M05-2X, M06-2X, M06, ω B97X ω B97X-D, B2PLYP,
4 and B2PLYP-D; one wave function method (MP2), one complete basis set method (CBS-
5 QB3), and one complex method for computing very accurate energies (G3). The DFT and
6 MP2 methods were used in conjunction with the basis sets 6-31+G(d), 6-311+G(d), 6-
7 311+G(d,p) and 6-311++G(d,p). CBS-QB3 and G3 were used as designed, i.e., with the
8 specific default basis sets originally conceived for each step of the composite calculation.
9

10
11
12
13
14 DFT approximations are particularly important since they are currently the most widely
15 used electronic structure methods. This is because they represent an optimum compromise
16 between accuracy and computational cost, especially for systems of medium to large size.
17
18 There is a huge number of DFT approximations available nowadays, thus it is not possible
19 to investigate them all. The subset used in the present study has been chosen to include
20 some of the most used ones and, at the same time, a diversity of features. The chosen DFT
21 methods belong to different rungs of the Jacob's ladder,³⁷ except the first one, and they
22 comprise both pure and hybrid functionals. Among the latter different variants have been
23 considered including global, range-separated and double-hybrids. The set also includes
24 DFT approximations with and without dispersion corrections, as well as non-parameterized
25 and empirically parameterized methods.
26
27
28
29
30
31
32
33

34 The standard linear regression fit necessary to obtain the empirical parameters (m and C_0)
35 for the FP method was obtained using the training set of molecules shown in Table 1,
36 according to:
37
38

$$pK_a^{\text{exp}} = m \Delta G_{s(\text{BA})} + C_0 \quad (1),$$

39
40 where $\Delta G_{s(\text{BA})}$ represents the difference in Gibbs free energy, in aqueous solution, between
41 each acid (A) and its corresponding conjugated base (B):
42
43
44

$$\Delta G_{s(\text{BA})} = G_{s(\text{B})} - G_{s(\text{A})} \quad (2),$$

45
46 $\Delta G_{s(\text{BA})}$ was calculated at 298.15 K and expressed in kcal/mol. The training set includes 20
47 molecules for each chemical family, namely phenols, carboxylic acids and amines. They
48 were chosen to include as many structural variations as possible. The phenols subset
49 includes functional groups with diverse electron donating and electron withdrawing effects,
50 at different ring sites. The carboxylic acids and amine subsets include both aliphatic and
51 aromatic species, and also different groups that modify their acidity. Their experimental
52
53
54
55
56
57
58
59
60

pK_a ranges are 6.33–10.31, -0.2–4.87, and 0.97–11.125 for phenols, carboxylic acids and amines, respectively.

Table 1. Training sets used to obtain the m and C_0 parameters and experimental pK_a values.

Phenol	pK_a^{exp}	Carboxylic acid	pK_a^{exp}	Amine	pK_a^{exp}
phenol	9.997 ^a	formic	3.744 ^a	methylamine	10.632 ^a
3-aminophenol	10.02 ^a	acetic	4.757 ^a	methoxyamine	4.62 ^a
4-aminophenol	10.30 ^b	chloroacetic	2.862 ^a	2-propen-1-amine	9.52 ^a
2-chlorophenol	8.531 ^a	oxalic	1.250 ^a	aminoacetonitrile	5.15 ^a
3-chlorophenol	9.125 ^a	methoxyacetic	3.570 ^a	dimethylamine	10.774 ^a
4-chlorophenol	9.426 ^a	prop-2-enoic	4.258 ^a	morpholine	8.492 ^a
o-cresol	10.31 ^a	trichloroacetic	0.660 ^b	piperidine	11.125 ^a
m-cresol	10.095 ^a	<i>trans</i> -2-butenoic	4.686 ^a	trimethylamine	9.799 ^a
p-cresol	10.269 ^a	3-butenoic	4.342 ^a	1-(2-propenyl)piperidine	9.68 ^c
2-nitrophenol	7.230 ^a	benzoic	4.202 ^a	phenilamine	4.601 ^a
3-nitrophenol	8.360 ^a	4-nitrobenzoic	3.442 ^a	1,4-benzenediamine	6.06 ^a
4-nitrophenol	7.150 ^a	2-methylbenzoic	3.906 ^a	2-chloroaniline	2.653 ^a
resorcinol	9.32 ^b	4-methylbenzoic	4.370 ^a	3-hydroxyaniline	4.27 ^a
hydroquinone	9.85 ^a	2-chlorobenzoic	2.925 ^a	4-chloroaniline	3.979 ^a
3-methoxyphenol	9.652 ^a	4-chlorobenzoic	3.986 ^a	pyridine	5.2 ^a
4-methoxyphenol	10.21 ^a	4-aminobenzoic, pKa2	4.870 ^a	pyrazine	0.97 ^a
4-hydroxy-benzonitrile	7.97 ^a	trifluoroacetic	-0.2 ^a	benzylamine	9.34 ^b
4-hydroxy-benzaldehyde	7.617 ^a	cyanoacetic	2.472 ^a	N-methylallylamine	10.11 ^d
2,3-dichlorophenol	7.696 ^a	nitroacetic	1.68 ^a	4-methoxypyridine	6.47 ^a
4-nitrosophenol	6.33 ^a	4-hydroxybenzoic	4.577 ^a	4-aminotoluene	5.08 ^a

^a reference ³⁸, ^b reference ³⁹, ^c reference ⁴⁰, ^d reference ⁴¹

The two parameters (m and C_0) were obtained separately for each chemical family, at each level of theory (electronic structure method / basis set). Then, they were used to test the performance of the FP method, in each case, using the test set of molecules shown in Table 2. To that purpose the calculated pK_a values were obtaining as:

$$pK_a^{\text{calc}} = m \Delta G_{s(\text{BA})}^{\text{TestSet}} + C_0 \quad (3)$$

The test set is completely independent from the training set, i.e., none of the molecules used to obtain the m and C_0 parameters were included in the test sets. The latter comprises 10 molecules per chemical family. In addition, a variety of structural features have been included. This selection has been made searching for increasing deviations from experimental values, thus the errors found here are as large as possible. Accordingly, it is expected that further use of the parameters proposed here lead to errors similar or lower than those arising from the test set. All the individual values of $\Delta G_{s(\text{BA})}$ calculated for both the training and the test sets can be downloaded from <http://www.agalano.com/deltaG.xlsx>.

Table 2. Test sets used to test the performance of the FP method, and experimental pKa values.

Phenol	pK_a^{exp}	Carboxylic acid	pK_a^{exp}	Amine	pK_a^{exp}
3,4-dihydroxybenzoic acid, pKa2	8.75 ^a	mercaptoacetic	3.68 ^b	glycine, methyl ester	7.66 ^a
2-fluorophenol	8.705 ^a	dichloroacetic	1.35 ^b	1-butanamine	10.640 ^a
2-chloro-4-nitrophenol	5.45 ^c	phenylacetic	4.31 ^a	O-methylhydroxylamine	4.75 ^c
2-(methylthio)phenol	9.53 ^b	fumaric, pKa1	3.02 ^a	dimethyl-N-propylamine	9.99 ^b
4-hydroxy-3-methoxybenzaldehyde	7.396 ^a	fumaric, pKa2	4.38 ^b	4-fluoroaniline	4.654 ^a
4-hydroxy-3-methoxybenzoic acid, pKa2	9.39 ^c	3,4-dihydroxybenzoic, pKa1	4.30 ^a	3-methoxy-benzenamine	4.204 ^a
1-(4-hydroxyphenyl)-ethanone	8.05 ^a	4-methoxybenzoic	4.475 ^a	imidazole	6.993 ^a
4-hydroxy-benzenemethanol	9.82 ^c	4-hydroxy-3-methoxybenzoic, pKa1	4.31 ^e	hydrazine	7.98 ^a
4-trifluoromethyl-phenol	8.675 ^a	dimethylpropanoic	5.032 ^a	ethylamine	10.673 ^a
3-fluorophenol	9.206 ^a	3-nitrobenzoic	3.449 ^a	N-methylaniline	4.832 ^a

^a reference ³⁸, ^b reference ³⁹, ^c reference ⁴⁰, ^e reference ⁴²

It is important to note that some of the molecules used here are polyprotic, i.e., they present more than one acid site. Six of them are diprotic molecules with two equivalent deprotonation sites (resorcinol, hydroquinone, oxalic acid, fumaric acid, 1,4-

1
2
3 benzenediamine, and hydrazine). Accordingly, each of them has two possible deprotonation
4 sites with identical acidity. This particular feature makes crucial to consider acid
5 dissociation micro-equilibria instead of the conventional ones.⁴³ The corresponding
6 expression of K_a for phenols and carboxylic acids ($H_2A \rightleftharpoons H^+ + HA^-$), as modeled in this
7 work, would be:
8
9
10
11

$$12 \quad K_a = 2e^{-\Delta G_a/RT} \quad (4),$$

13 while for amines ($H_2A^+ \rightleftharpoons H^+ + HA$):
14
15
16
17

$$18 \quad K_a = \frac{1}{2}e^{-\Delta G_a/RT} \quad (5),$$

19 where ΔG_a is the Gibbs free energy of the acid dissociation equilibrium, considering any of
20 the equivalent deprotonation sites. From equations (4) and (5) it becomes evident that:
21
22
23
24

$$25 \quad \log(K_a) = \pm \log(2) - \frac{\Delta G_a}{2.303RT} \quad (6)$$

26 This means that the correct calculated pK_a value differs from that obtained using the
27 conventional equilibrium constant by $\pm \log(2)$. However, in the present work the relevant
28 calculated magnitude is $\Delta G_{s(BA)}$, i.e., it is the one that needs to be corrected when the FP
29 method is used. Equation (6) can be easily modified to obtain the necessary correction to
30 ΔG_a :
31
32
33
34
35
36
37

$$38 \quad \Delta G_a^{corr} = \Delta G_a \pm 2.303RT \log(2) \quad (7),$$

39 which, at 298.15 K, represents a correction of ± 0.41 kcal/mol to the calculated ΔG_a . Since
40 in the present study $\Delta G_{s(BA)}$ is used instead of ΔG_a (and the Gibbs energy of the proton is
41 obtained from the linear fitting), this correction is made to the calculated $\Delta G_{s(BA)}$ values of
42 the above mentioned symmetric diprotic molecules. For phenols and carboxylic acids the
43 correction is negative, while for the amines it is positive. This is because for the
44 deprotonation of symmetric amines the starting molecule was modeled using the mono-
45 cation, i.e., the species with only one protonated group.
46
47
48
49
50
51
52
53

54 In addition, there are other molecules in the used sets with more than one pK_a . The more
55 similar the pK_a values, the higher the important of considering acid dissociation micro-
56
57
58
59
60

equilibria. In the present study the polyprotic molecule with the closest pK_a values is the 4-aminobenzoic acid (2.41 and 4.87). These values are 2.46 pKa units apart, which would represent a correction to $\Delta G_{s(BA)}$ equal to ± 0.002 kcal/mol, at 298.15 K, when using the ionization micro-equilibrium instead of the conventional one. This value is much lower than the chemical accuracy, thus it seems that the conventional approximation is adequate in this case, and probably also for similar cases with pK_a differences ≤ 2 pK_a units, or differences in calculated ΔG_a values ≤ 2.7 kcal/mol (they would need ΔG corrections $\leq \pm 0.006$ kcal/mol, at 298.15 K). The interested reader can check the micro-equilibrium treatment of polyprotic compounds in more detail elsewhere.⁴³

The main criteria for accuracy used here are mean unsigned errors (MUE), maximum absolute errors (MAE), root-mean square errors (RMSE), and the linear regression coefficients (R^2). Their target values are 0.5 pK_a units, 1.0 pK_a units, 0.5 pK_a units and 0.9 respectively. The threshold values for MUE, MAE and RMSE are significantly below the 2 pK_a units currently accepted as reasonably accurate for pK_a calculations,²² and close to deviations arising from using different experimental techniques.

In addition, the Gibbs energy of the proton $\Delta G_s(H^+)$ obtained from the FP method has also been estimated, and compared with the experimental data. For a conventional deprotonation reaction ($HA \rightleftharpoons H^+ + A^-$), the expression to directly calculate pK_a values is:

$$pK_a = \frac{\Delta G_s}{RT \ln(10)} \quad (8),$$

where ΔG_s is the Gibbs energy of the deprotonation reaction, in aqueous solution, and can be expressed as:

$$\Delta G_s = \Delta G_{s(BA)} + \Delta G_s(H^+) \quad (9)$$

In addition, the linear regression fit used to obtain the m and C_0 parameters is equivalent to including a scaling factor ($^s f$) that corrects for the systematic errors of any computational method (for similar deprotonation reactions, i.e., for each chemical family). Therefore, equation (1) can be rewritten as:

$$pK_a^{\text{exp}} = \frac{{}^s f \Delta G_{s(BA)}}{RT \ln(10)} + \frac{{}^s f \Delta G_s(H^+) }{RT \ln(10)} \quad (10),$$

which means that:

$$m = \frac{{}^s f}{RT \ln(10)} \quad (11)$$

$$C_0 = \frac{{}^s f \Delta G_s(H^+)}{RT \ln(10)} \quad (12)$$

Rearranging equations (11) and (12) it is possible to estimate the scaling factor and the value of $\Delta G_s(H^+)$, for each level of theory used in this work, according to:

$${}^s f = mRT \ln(10) \quad (13)$$

$$\Delta G_s(H^+) = \frac{C_0 RT \ln(10)}{{}^s f} = \frac{C_0}{m} \quad (14)$$

The $\Delta G_s(H^+)$ values obtained this way were compared with that derived from experiments. Such experimental values were calculated as:

$$\Delta G_s(H^+) = G_g(H^+) + \Delta G_{solv}(H^+) \quad (15),$$

where $G_g(H^+)$ and $\Delta G_{solv}(H^+)$ are the gas-phase free energy and the free energy of solvation of the proton.

However, while the $G_g(H^+)$ value is well established, measured values of $\Delta G_{solv}(H^+)$ significantly vary, depending on the conditions of the particular experiment carried out. The variations are rather large, with the reported $\Delta G_{solv}(H^+)$ values ranging from -259 to -265.89 kcal/mol. These values, as well as all the data reported here, correspond to 1M standard state. Because of the variation in $\Delta G_{solv}(H^+)$, the $\Delta G_s(H^+)$ values derived from experiments range from -263.4 to -270.28 kcal/mol, with the latter been the recommended by Camaioni and Schwerdtfeger.⁴⁴ Thus, this range was used as the experimental reference for the calculated $\Delta G_s(H^+)$ values.

Results and Discussion

The linear regression coefficients (R^2) for all the tested levels of theory are reported in Tables S1, S2 and S3 (Supporting Information) for the training sets of phenols, carboxylic acids and amines, respectively. All of them were found to be higher than 0.9, ranging from 0.92 to 0.99. Thus the linear relationship between the calculated $\Delta G_{s(BA)}$ values and the

experimental pK_a s was confirmed for all the investigated chemical families, and for all the tested levels of theory.

The MUE and RMSE values are reported in Tables S4 to S6 and S1 to S3, respectively (Supporting Information). They are also shown in Figures 1 and 2 to facilitate rapid analyses. The most important finding is that all the MUE are below the target value (0.5 pK_a units), except for amines with MP2/6-311+G(d), and that all the RMSE are below the same target value, except for amines with the same level of theory and with MP2/6-311+G(d). In these cases MUE is equal to 0.54 pK_a units, and RMSE values are 0.59 and 0.68. In general, the results for carboxylic acids are better than those obtained for the other two chemical families. It is interesting to note that most of the MUE and RMSE values obtained with the tested DFT methods are similar to those arising from using high theory levels (CBS-QB3 and G3). Therefore, pK_a estimations with equivalent accuracy can be achieved using less computationally expensive methods. Most of the tested functionals also produce better results than MP2, which has significantly higher computational costs. Thus, using this method for the task at hand is not recommended.

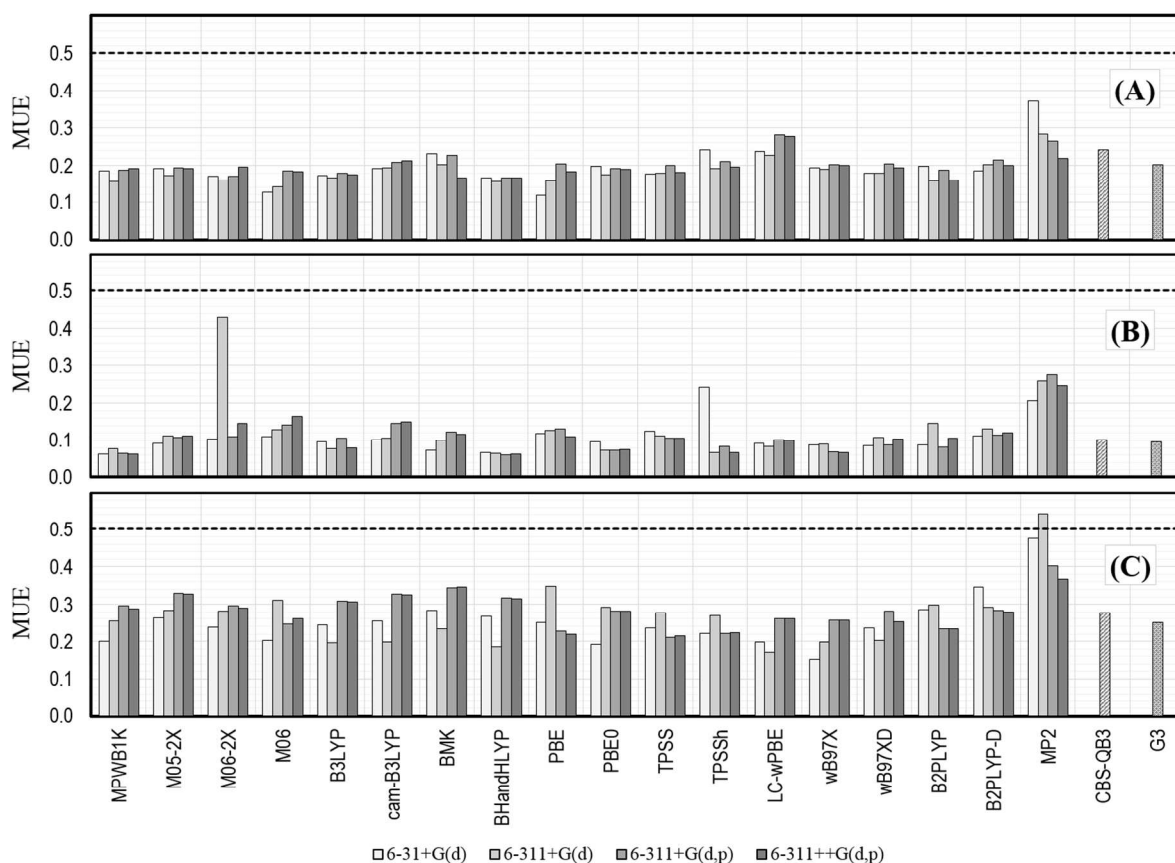


Figure 1. Mean unsigned errors (MUE), in pK_a units, obtained for the test sets of phenols (A), carboxylic acids (B), and amines (C). The dashed line represents the target value.

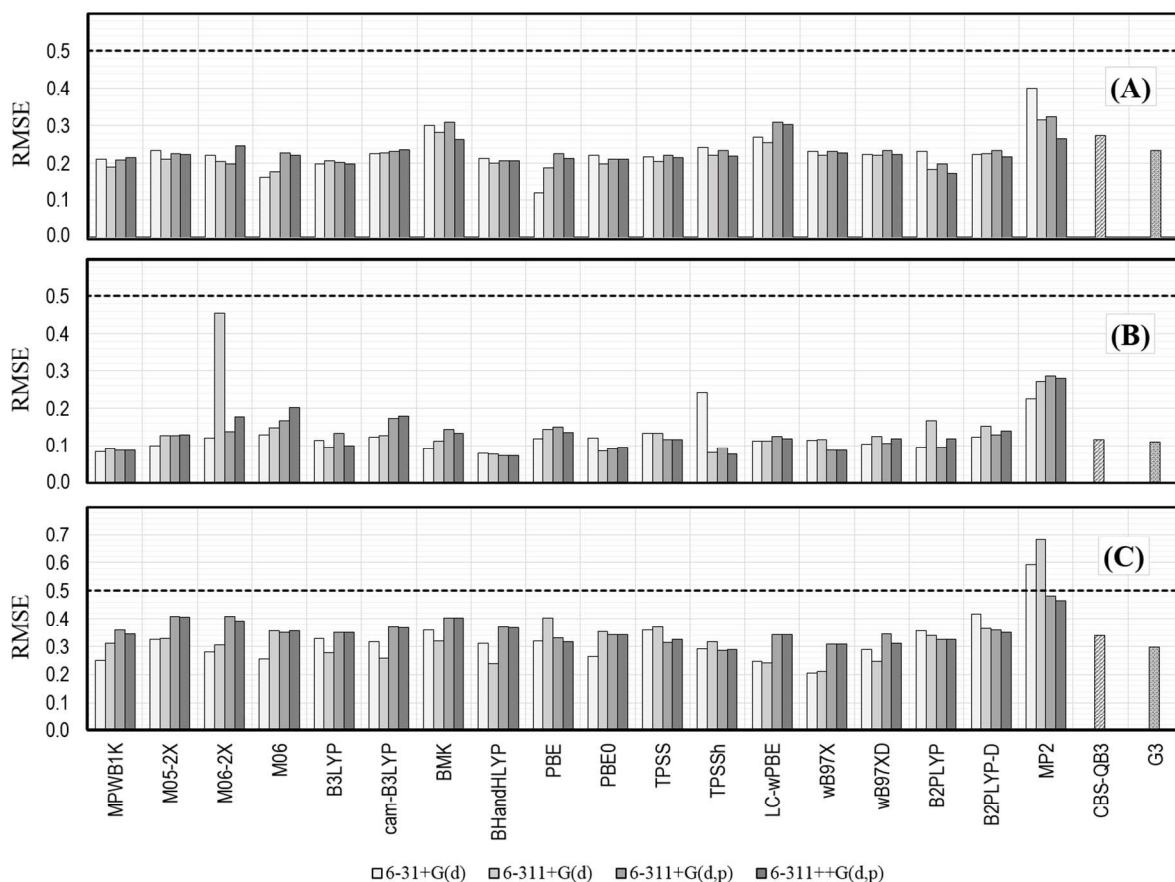


Figure 2. Root-mean square errors (RMSE), in pK_a units, obtained for the test sets of phenols (A), carboxylic acids (B), and amines (C). The dashed line represents the target value.

The MUE and RMSE values obtained with the G3 method are systematically lower than those obtained with CBS-QB3, regardless of the chemical family, i.e., phenols, carboxylic acids, or amines. Comparing the results obtained with ω B97X and B2PLYP, with their counterparts ω B97XD and B2PLYP-D, it can be concluded that including dispersion has no significant effect on the accuracy of pK_a calculations, when the FP method is used. Amines were identified as the chemical family leading to the largest MUE and RMSE values. This

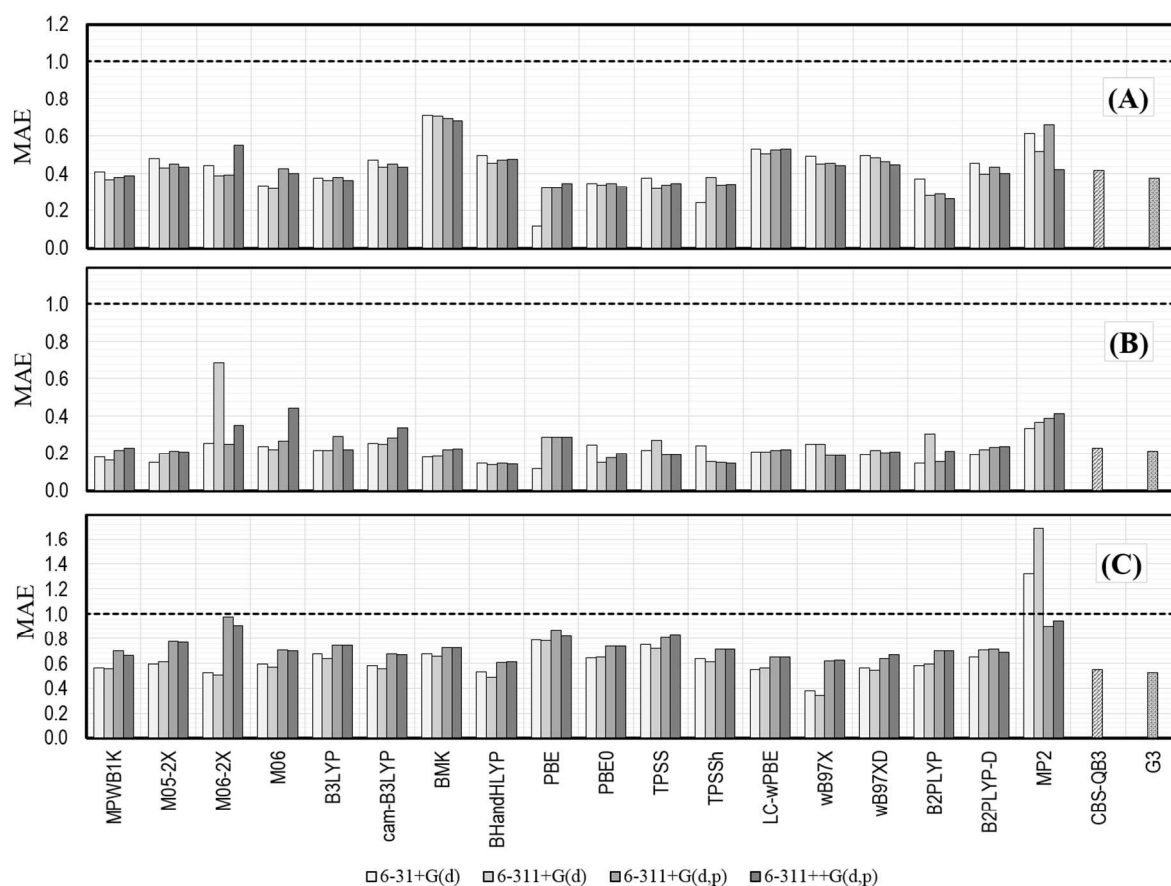
1
2
3 is rationalized based on the fact that this family is particularly diverse. It includes aromatic,
4 aliphatic and allylic species, heterocyclic compounds, and imidazoles. In addition the acid
5 site can correspond to primary, secondary or tertiary amines. However, despite of such wide
6 structural diversity, all the tested methods –except for MP2/6-311+G(d)– yield MUE and
7 RMSE values lower than the threshold. Moreover, 94% of the obtained MUE and RMSE
8 are lower than 0.35 and 0.45 pK_a units, respectively, for amines. For phenols and
9 carboxylic acids, on the other hand, this number is 98.6%. In general all the DFT
10 approaches are adequate to be used in conjunction with the FP method. However, it should
11 be noted that the computational costs of B2PLYP and B2PLYP-D are significantly higher
12 than those of the other tested functionals. Thus, as the size of the investigated system
13 increases they might become rather expensive.

14
15
16
17
18
19
20
21
22
23 Regarding the influence of the basis set, MP2 was the most sensitive electronic structure
24 method, while for the DFT approaches such influence is only small. However, it should be
25 noted that all the basis sets used here include diffuse functions. This feature is expected to
26 be crucial in pK_a calculations, where there are anions involved. This is because such
27 species have more spread out electron densities than neutral species, thus diffuse functions
28 must be added to treat them properly. Since, for the anionic species studied here, the
29 negative charge is mainly located on O or N atoms, including diffuse function on non-
30 hydrogen atoms is enough to properly describe the systems of interest. This is demonstrated
31 by the fact that the results obtained with the basis sets 6-311+G(d,p), and 6-311++G(d,p)
32 are almost identical. The only exception is MP2, which is known for having a strong
33 dependence with the basis set size.

34
35
36
37
38
39
40
41
42
43 On the other hand, including polarization functions on the H atom has almost negligible
44 effects on the MUE and RMSE values obtained for phenols and carboxylic acids, when
45 using DFT methods. On the contrary, a significant effect was found for amines with most
46 of the tested functionals. Curiously, for some DFT approximations the effect was contrary
47 to the expected one, i.e., for this chemical family the MUE and RMSE values obtained with
48 the 6-31+G(d) and 6-311+G(d) basis sets are lower than those obtained with basis sets 6-
49 311+G(d,p), and 6-311++G(d,p). This is rather surprising since adding p polarization
50 functions to hydrogen atoms, which are relevant to acid equilibria, was expected to
51 systematically improve the results. The methods showing the expected trend are MP2, and
52
53
54
55
56
57
58
59
60

the DFT approximations TPSS, TPSSh, B2PLYP and B2PLYP-D. Thus, apparently, for the DFT approaches that are up to the fourth rung of the Jacob's ladder non-hybrid and hybrid functionals with low amount of exact exchange are the ones with the expected behavior.

Albeit MUE and RMSE values are probably the most widely used, and –arguably– the most important descriptors of accuracy, MAE values are also relevant. In the context of the present work they indicate how divergent from the experimental data can be a particular calculated pK_a . The MAE values are reported in Tables S4 to S6 (Supporting Information), and are schematically shown in Figure 3. Logically, most of the trends for this descriptor are in line with those corresponding to MUE and RMSE values. Only two values were found to be above the MAE threshold (1 pK_a unit), and all correspond to MP2 calculations. In addition, a high proportion of the tested approaches produce MAE values lower than 0.7 pK_a units (97.3%, 100%, and 60.8% for phenols, carboxylic acids and amines, respectively).



1
2
3 **Figure 3.** Maximum absolute errors (MAE), in pK_a units, obtained for the test sets of
4 phenols (A), carboxylic acids (B), and amines (C). The dashed line represents the target
5 value.
6
7

8
9
10 The Gibbs energy of the proton in aqueous solution, $\Delta G_s(H^+)$, obtained from the FP method
11 are reported in Tables 7S to 9S (Supporting Information), and shown together with the
12 experimental range in Figure 4. The calculated $\Delta G_s(H^+)$ are strongly influenced by the
13 chemical family for which the FP method was applied. The largest deviation from the
14 values derived from experiments was found for phenols. For this family of compounds the
15 only values that are in the upper limit of the experimental range are those obtained with the
16 G3 and CBS-QB3 methods, and with MP2 and the two largest basis sets. On the contrary,
17 for carboxylic acids most of the calculated $\Delta G_s(H^+)$ are within the experimental range, and
18 some of them are close to the value recommended by Camaioni and Schwerdtfeger.⁴⁴
19 Amines show an intermediate behavior, with the agreement depending on the used
20 electronic structure method.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

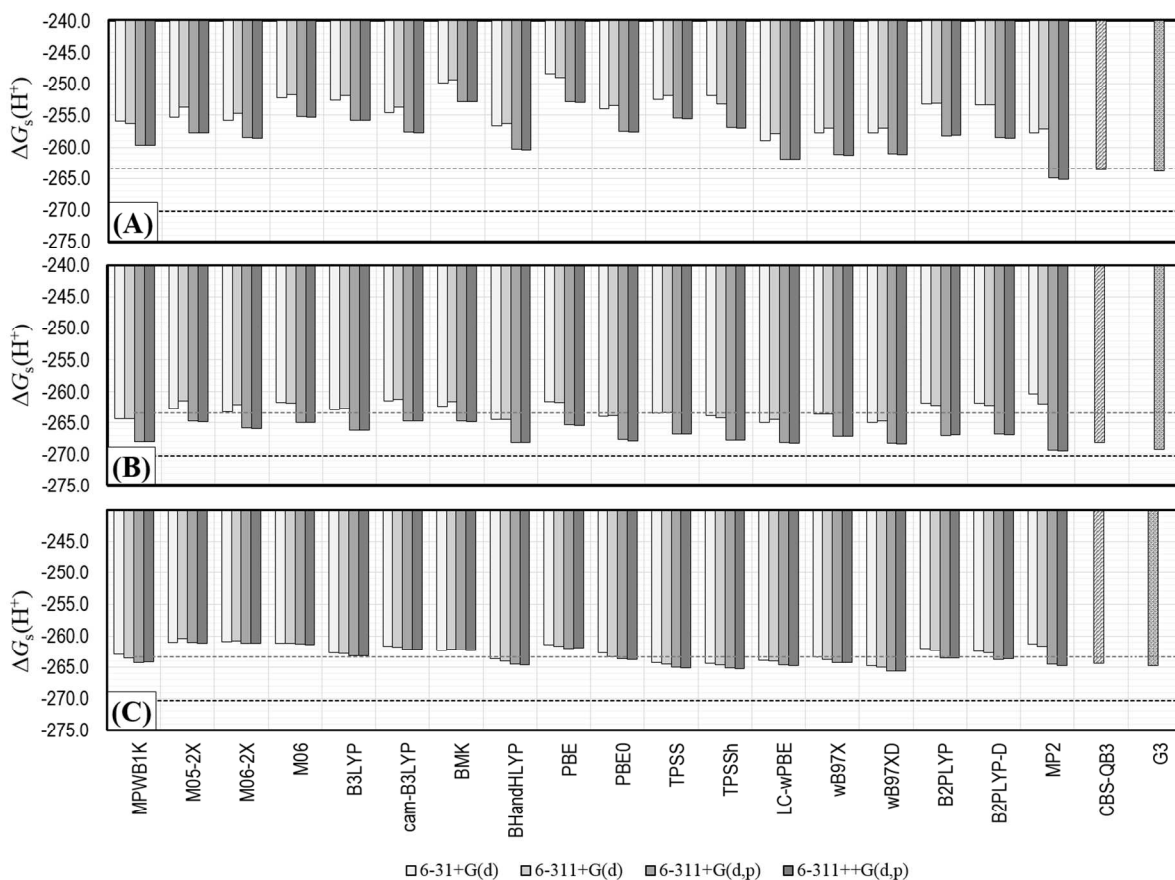


Figure 4. Calculated $\Delta G_s(H^+)$ values, in kcal/mol, obtained from the FP method for phenols (A), carboxylic acids (B), and amines (C). The dashed lines represent the experimental range, with the black one corresponding to value recommended by Camaioni and Schwerdtfeger.⁴⁴

It is interesting to note that the expected trend with the basis set size is fulfilled for the calculated $\Delta G_s(H^+)$ values. In general, as the basis set increases so does the agreement with the experimental $\Delta G_s(H^+)$. Such agreement can also be taken as an indication of how accurate may be a pK_a calculation using the direct method (equation 8). Therefore, it is anticipated that G3, CBS-QB3, MP2/6-311+G(d,p) and MP2/6-311++G(d,p) levels of theory, might produce reasonably accurate pK_a values for the three investigated families of compounds when using the direct method. This is provided that this method is used modeling the solvent with SMD. On the contrary, poor agreement with the experimental values is expected for phenolic compounds when their pK_a s are calculated with the direct method, regardless of the used level of theory.

Figure 4 also demonstrate that the optimal value of $\Delta G_s(H^+)$, required for accurately predicting pK_a values, is significantly different for each level of theory and for each chemical family. Accordingly, using the direct method with the experimental $\Delta G_s(H^+)$ value, for example -270.28 kcal/mol, would frequently lead to rather large errors. This is compensated by the FP method, which is why it is such an appealing computational protocol for pK_a calculations. Moreover, currently it is probably the most reliable way to computationally estimate pK_a values, within the electronic structure based strategies, at least for molecules of modest complexity.

From a pure theoretical point of view, arguments against empirically parameterizing the calculation of chemical properties might be provided. However, from a pragmatic point of view, it is important to have reliable estimation of chemical properties, especially when they can be hard to acquire from experiments. In addition, there are many chemical fields where computational tools are being used to designed new molecules with potential practical applications. Thus, the necessity of estimating pK_a values of molecules that have not even synthesized yet is increasingly frequent. The FP method is a very promising computational tool for that purpose, albeit further validations are still necessary. For example, the performance of this method still needs to be tested for other families of compounds, and for molecules of increasing complexity. The values obtained of the m and C_0 parameters proposed here are presented in Tables 3, 4 and 5 for phenols, carboxylic acids and amines, respectively.

Table 3. Empirically fitted parameters (m and C_0) for calculating pK_a values of phenols.

	6-31+G(d)		6-311+G(d)		6-311+G(d,p)		6-311++G(d,p)	
	m	C_0	m	C_0	m	C_0	m	C_0
MPWB1K	0.299	-76.603	0.305	-78.215	0.302	-78.493	0.303	-78.564
M05-2X	0.313	-80.020	0.313	-79.435	0.316	-81.497	0.317	-81.693
M06-2X	0.315	-80.680	0.314	-80.073	0.318	-82.209	0.318	-82.349
M06	0.285	-71.924	0.289	-72.746	0.290	-74.064	0.291	-74.411
B3LYP	0.284	-71.669	0.284	-71.465	0.285	-72.933	0.286	-73.092
cam-B3LYP	0.306	-77.994	0.306	-77.676	0.309	-79.509	0.309	-79.644
BMK	0.265	-66.137	0.269	-67.173	0.270	-68.221	0.270	-68.350

BHandHLYP	0.302	-77.406	0.301	-77.230	0.304	-79.183	0.305	-79.473
PBE	0.267	-66.340	0.278	-69.341	0.280	-70.833	0.281	-71.053
PBE0	0.292	-74.192	0.294	-74.605	0.296	-76.202	0.297	-76.472
TPSS	0.280	-70.720	0.280	-70.559	0.282	-71.926	0.282	-72.194
TPSSh	0.269	-67.830	0.285	-72.120	0.286	-73.542	0.287	-73.765
LC- ω PBE	0.328	-85.041	0.325	-83.818	0.327	-85.680	0.328	-85.839
ω B97X	0.328	-84.464	0.326	-83.859	0.332	-86.633	0.332	-86.819
ω B97XD	0.312	-80.511	0.313	-80.509	0.317	-82.746	0.318	-82.998
B2PLYP	0.301	-76.171	0.303	-76.676	0.305	-78.854	0.305	-78.867
B2PLYP-D	0.298	-75.527	0.302	-76.540	0.305	-78.702	0.306	-79.128
MP2	0.363	-93.670	0.352	-90.368	0.358	-94.844	0.363	-96.222
<hr/>								
CBS-QB3	0.350	-92.353						
G3	0.342	-90.268						

Table 4. Empirically fitted parameters (m and C_0) for calculating pK_a values of carboxylic acids.

	6-31+G(d)		6-311+G(d)		6-311+G(d,p)		6-311++G(d,p)	
	m	C_0	m	C_0	m	C_0	m	C_0
MPWB1K	0.345	-91.274	0.349	-92.266	0.347	-93.017	0.347	-93.123
M05-2X	0.360	-94.513	0.360	-94.101	0.356	-94.380	0.357	-94.622
M06-2X	0.361	-95.155	0.376	-98.630	0.357	-95.033	0.358	-95.120
M06	0.314	-82.101	0.328	-85.970	0.330	-87.368	0.326	-86.411
B3LYP	0.337	-88.582	0.341	-89.657	0.343	-91.387	0.343	-91.399
cam-B3LYP	0.341	-89.049	0.345	-90.158	0.345	-91.268	0.344	-91.194
BMK	0.344	-90.158	0.344	-89.932	0.340	-90.095	0.342	-90.599
BHandHLYP	0.348	-91.988	0.348	-91.988	0.350	-93.812	0.350	-93.858
PBE	0.323	-84.597	0.340	-89.087	0.346	-91.883	0.347	-92.136
PBE0	0.351	-92.737	0.357	-94.106	0.360	-96.221	0.365	-97.737
TPSS	0.319	-84.104	0.324	-85.339	0.326	-86.890	0.325	-86.709
TPSSh	0.322	-84.934	0.336	-88.881	0.340	-90.992	0.340	-91.075
LC- ω PBE	0.377	-99.898	0.377	-99.742	0.379	-101.521	0.379	-101.724
ω B97X	0.366	-96.619	0.371	-97.807	0.371	-99.160	0.371	-99.157
ω B97XD	0.361	-95.530	0.366	-96.990	0.366	-98.307	0.368	-98.617
B2PLYP	0.349	-91.420	0.359	-94.236	0.358	-95.463	0.359	-95.830
B2PLYP-D	0.345	-90.420	0.354	-92.894	0.347	-92.658	0.348	-92.775
MP2	0.376	-97.830	0.396	-103.808	0.401	-107.958	0.403	-108.537

CBS-QB3	0.390	-104.497
G3	0.392	-105.646

Table 5. Empirically fitted parameters (m and C_0) for calculating pK_a values of amines.

	6-31+G(d)		6-311+G(d)		6-311+G(d,p)		6-311++G(d,p)	
	m	C_0	m	C_0	m	C_0	m	C_0
MPWB1K	0.498	-130.930	0.505	-133.064	0.497	-131.441	0.494	-130.452
M05-2X	0.475	-124.072	0.475	-123.807	0.464	-121.000	0.464	-121.235
M06-2X	0.504	-131.574	0.499	-130.033	0.486	-126.867	0.485	-126.727
M06	0.513	-133.920	0.536	-139.968	0.518	-135.294	0.520	-135.911
B3LYP	0.497	-130.452	0.502	-131.801	0.484	-127.244	0.484	-127.509
cam-B3LYP	0.499	-130.528	0.503	-131.693	0.486	-127.438	0.487	-127.546
BMK	0.483	-126.785	0.492	-128.862	0.472	-123.746	0.473	-123.980
BHandHLYP	0.493	-129.951	0.498	-131.390	0.481	-127.277	0.483	-127.691
PBE	0.497	-130.042	0.507	-132.605	0.489	-127.969	0.485	-126.877
PBE0	0.494	-129.783	0.508	-133.801	0.490	-129.155	0.491	-129.461
TPSS	0.496	-131.141	0.505	-133.541	0.492	-130.308	0.495	-131.308
TPSSh	0.494	-130.509	0.503	-133.212	0.490	-129.870	0.491	-130.366
LC- ω PBE	0.497	-131.249	0.494	-130.465	0.491	-129.854	0.492	-130.184
ω B97X	0.506	-133.174	0.514	-135.622	0.502	-132.552	0.503	-132.935
ω B97XD	0.489	-129.398	0.500	-132.565	0.496	-131.715	0.491	-130.301
B2PLYP	0.491	-128.740	0.503	-132.033	0.497	-131.023	0.497	-130.989
B2PLYP-D	0.490	-128.701	0.495	-129.952	0.486	-128.235	0.483	-127.285
MP2	0.493	-128.841	0.495	-129.379	0.522	-138.182	0.528	-139.753
CBS-QB3	0.531	-140.475						
G3	0.517	-136.908						

These parameters, were obtained using a large enough training set and their reliability was assessed using the pK_a values calculated with them, for an independent test set of molecules. It seems worthwhile to insist on the fact that using these parameters –for molecules of modest complexity– is expected to produce calculated pK_a values in excellent

1
2
3 agreement with the experimental ones, with deviations lower than 0.5 pK_a units for the
4 three studied chemical families using all the tested methods except MP2.
5
6

7 8 **Conclusions**

9
10 The parameters m and C_0 that allow fast and reliable pK_a calculations, based on electronic
11 structure methods, using the FP method are provided for the first time for phenols,
12 carboxylic acids and amines, at 74 levels of theory. The coefficients were obtained by least-
13 squares fits of $\Delta G_{s(BA)}$ vs. experimental pK_a values using training sets of 20 molecules for
14 each chemical family. The reliability of the FP method was confirmed using test sets of 10
15 molecules per family, that are not included in the training sets.
16
17

18 It was found that, except for MP2, deviations from experiments are lower than 0.5 pK_a
19 units. 94% of the obtained MUE and RMSE are lower than 0.35 and 0.45 pK_a units,
20 respectively, for amines. For phenols and carboxylic acids, on the other hand, this number
21 is 98.6%. In addition, a high proportion of the tested approaches produce MAE values
22 lower than 0.7 pK_a units (97.3%, 100%, and 60.8% for phenols, carboxylic acids and
23 amines, respectively).
24
25

26 The parameters estimated here are expected to facilitate pK_a calculations, using electronic
27 structure based strategies, with uncertainties close to the experimental ones for phenols,
28 carboxylic acids, and amines. It should be noted, however, that the present study deals only
29 with molecules of modest complexity, thus the reliability of the FP method for more
30 complex systems remains to be tested.
31
32

33 **Acknowledgements**

34 We gratefully acknowledge the Laboratorio de Supercómputo y Visualización en Paralelo
35 at Universidad Autónoma Metropolitana-Iztapalapa and DGTIC, UNAM for computing
36 time. This work was partially supported by project DGAPA PAPIIT- IN220215. A. P.-G.
37 acknowledges the economic support of the Program of Cátedras - CONACYT from
38 CONACyT - UAM (2015-2025). R. C.-A. and J. R. L.-C. acknowledge the financial
39 support from CONACyT postdoctoral fellowship; and L. M.-R., G. M.-S., A. R.-S., A. I.-
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

E., A. M. R.-Z. and M. A. H.-O. acknowledge the financial support from CONACyT doctoral fellowship

Supporting Information. Linear regression coefficients (R²), root mean square errors (RMSE), mean unsigned errors (MUE), maximum absolute errors (MAE), and Gibbs energy of the proton in aqueous solution. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

1. Pliego Jr, J. R.; Riveros, J. M., Theoretical Calculation of Pka Using the Cluster-Continuum Model. *J Phys Chem A* **2002**, 106, 7434-7439.
2. Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J., Prediction of the Pka of Carboxylic Acids Using the Ab Initio Continuum-Solvation Model Pcm-Uahf. *J Phys Chem A* **1998**, 102, 6706-6712.
3. da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C., Ab Initio Calculations of Absolute Pka Values in Aqueous Solution I. Carboxylic Acids. *J. Phys. Chem. A* **1999**, 103, 11194-11199.
4. Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C., Accurate Relative Pka Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-N Models Combined with Continuum Solvation Methods. *J Chem Phys* **2001**, 114, 4595-4606.
5. Liptak, M. D.; Shields, G. C., Accurate Pka Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-N Models Combined with Cpcm Continuum Solvation Methods. *J Am Chem Soc* **2001**, 123, 7314-7319.
6. Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C., Absolute Pka Determinations for Substituted Phenols. *J Am Chem Soc* **2002**, 124, 6421-6427.
7. Chipman, D. M., Computation of Pka from Dielectric Continuum Theory. *J. Phys. Chem. A* **2002**, 106, 7413-7422.
8. Adam, K. R., New Density Functional and Atoms in Molecules Method of Computing Relative Pka Values in Solution. *J. Phys. Chem. A* **2002**, 106, 11963-11972.
9. Saracino, G. A. A.; Improta, R.; Barone, V., Absolute Pka Determination for Carboxylic Acids Using Density Functional Theory and the Polarizable Continuum Model. *Chem Phys Lett* **2003**, 373, 411-415.
10. Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E., First Principles Calculations of Aqueous Pka Values for Organic and Inorganic Acids Using Cosmo-Rs Reveal an Inconsistency in the Slope of the Pka Scale. *J. Phys. Chem. A* **2003**, 107, 9380-9386.
11. Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Ionization of Organic Acids in Dimethyl Sulfoxide Solution: A Theoretical Ab Initio Calculation of the Pka Using a New Parametrization of the Polarizable Continuum Model. *J. Phys. Chem. A* **2004**, 108, 166-171.
12. Magill, A. M.; Cavell, K. J.; Yates, B. F., Basicity of Nucleophilic Carbenes in Aqueous and Nonaqueous Solventstheoretical Predictions. *J Am Chem Soc* **2004**, 126, 8717-8724.

13. Vianello, R.; Maksić, Z. B., Strong Acidity of Some Polycyclic Aromatic Compounds Annulated to a Cyclopentadiene Moiety and Their Cyano Derivatives - a Density Functional B3lyp Study. *Eur J Org Chem* **2005**, 3571-3580.
14. Namazian, M.; Zakery, M.; Noorbala, M. R.; Coote, M. L., Accurate Calculation of the Pka of Trifluoroacetic Acid Using High-Level Ab Initio Calculations. *Chem Phys Lett* **2008**, 451, 163-168.
15. Ho, J.; Coote, M. L., Pka Calculation of Some Biologically Important Carbon Acids - an Assessment of Contemporary Theoretical Procedures. *J. Chem. Theory Comput.* **2009**, 5, 295-306.
16. Song, Y.; Mao, J.; Gunner, M. R., Mcce2: Improving Protein Pka Calculations with Extensive Side Chain Rotamer Sampling. *J. Comput. Chem.* **2009**, 30, 2231-2247.
17. Delgado, E. J., Dft Calculation of Pka's for Dimethoxypyrimidinylsalicylic Based Herbicides. *Chem. Phys. Lett.* **2009**, 471, 133-135.
18. Casanovas, R.; Frau, J.; Ortega-Castro, J.; Salvà, A.; Donoso, J.; Muñoz, F., Absolute and Relative Pka Calculations of Mono and Diprotic Pyridines by Quantum Methods. *J. Mol. Struct.: THEOCHEM* **2009**, 912, 5-12.
19. Dissanayake, D. P.; Senthilnithy, R., Thermodynamic Cycle for the Calculation of Ab Initio Pka Values for Hydroxamic Acids. *J. Mol. Struct.: THEOCHEM* **2009**, 910, 93-98.
20. Uudsemaa, M.; Kanger, T.; Lopp, M.; Tamm, T., Pka Calculation for Monoprotonated Bipiperidine, Bimorpholine and Their Derivatives in H₂O and MeCN. *Chem Phys Lett* **2010**, 485, 83-86.
21. Rayne, S.; Forest, K., Theoretical Studies on the Pka Values of Perfluoroalkyl Carboxylic Acids. *J Mol Struct: THEOCHEM* **2010**, 949, 60-69.
22. Ho, J.; Coote, M. L., A Universal Approach for Continuum Solvent Pka Calculations: Are We There Yet? *Theor. Chem. Acc.* **2009**, 125, 3-21.
23. Kličić, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C., Accurate Prediction of Acidity Constants in Aqueous Solution Via Density Functional Theory and Self-Consistent Reaction Field Methods. *J. Phys. Chem. A* **2002**, 106, 1327-1335.
24. Eckert, F.; Klamt, A., Accurate Prediction of Basicity in Aqueous Solution with Cosmo-Rs. *J. Comput. Chem.* **2006**, 27, 11-19.
25. Zhang, S.; Baker, J.; Pulay, P., A Reliable and Efficient First Principles-Based Method for Predicting Pka Values. 1. Methodology. *J Phys Chem A* **2010**, 114, 425-431.
26. Zhang, S.; Baker, J.; Pulay, P., A Reliable and Efficient First Principles-Based Method for Predicting Pka Values. 2. Organic Acids. *J Phys Chem A* **2010**, 114, 432-442.
27. Zhang, S., A Reliable and Efficient First Principles-Based Method for Predicting Pka Values. Iii. Adding Explicit Water Molecules: Can the Theoretical Slope Be Reproduced and Pka Values Predicted More Accurately? *J. Comput. Chem.* **2012**, 33, 517-526.
28. Zhang, S., A Reliable and Efficient First Principles-Based Method for Predicting Pka Values. 4. Organic Bases. *J. Comput. Chem.* **2012**, 33, 2469-2482.
29. Matsui, T.; Oshiyama, A.; Shigeta, Y., A Simple Scheme for Estimating the Pka Values of 5-Substituted Uracils. *Chem. Phys. Lett.* **2011**, 502, 248-252.
30. Matsui, T.; Baba, T.; Kamiya, K.; Shigeta, Y., An Accurate Density Functional Theory Based Estimation of Pka Values of Polar Residues Combined with Experimental Data: From Amino Acids to Minimal Proteins. *Phys. Chem. Chem. Phys.* **2012**, 14, 4181-4187.

- 1
2
3 31. Álvarez-Diduk, R.; Ramírez-Silva, M. T.; Galano, A.; Merkoçi, A., Deprotonation
4 Mechanism and Acidity Constants in Aqueous Solution of Flavonols: A Combined
5 Experimental and Theoretical Study. *J. Phys. Chem. B* **2013**, 117, 12347-12359.
- 6
7 32. Baba, T.; Matsui, T.; Kamiya, K.; Nakano, M.; Shigeta, Y., A Density Functional
8 Study on the Pka of Small Polyprotic Molecules. *Int. J. Quantum Chem.* **2014**, 114, 1128-
9 1134.
- 10
11 33. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.;
12 Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.;
13 Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg,
14 J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.;
15 Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro,
16 F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi,
17 R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.;
18 Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.;
19 Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.;
20 Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.;
21 Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.;
22 Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.:
23 Wallingford, CT, USA, 2009.
- 24
25 34. Cramer, C. J.; Truhlar, D. G., A Universal Approach to Solvation Modeling. *Acc.*
26 *Chem. Res.* **2008**, 41, 760-768.
- 27
28 35. Ho, J.; Klamt, A.; Coote, M. L., Comment on the Correct Use of Continuum
29 Solvent Models. *J Phys Chem A* **2010**, 114, 13442-13444.
- 30
31 36. Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Use of Solution-
32 Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J*
33 *Phys Chem B* **2011**, 115, 14556-14562.
- 34
35 37. Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka,
36 G. I., Prescription for the Design and Selection of Density Functional Approximations:
37 More Constraint Satisfaction with Fewer Fits. *J Chem Phys* **2005**, 123, 62201.
- 38
39 38. Smith, R. M.; Martell, A. E.; Motekaitis, R. J., In *NIST Standard Reference*
40 *Database 46*; National Institute of Standards and Technology (NIST): Texas A&M
41 University, 2004.
- 42
43 39. Dissociation Constants of Organic Acids and Bases. In *Handbook of Chemistry and*
44 *Physics*, Haynes, W. M.; Lide, D. R.; Bruno, T. J., Eds.; CRC Press: Boca Raton FL, 2014;
45 Chapter 5, p 94.
- 46
47 40. Ionization Constants of Acids and Bases. In *Handbook of Biochemistry and*
48 *Molecular Biology.*, Lundblad, R. L.; Macdonald, F., Eds.; CRC Press: Boca Raton FL,
49 2010, p 595.
- 50
51 41. Hall Jr, H. K., Potentiometric Determination of the Base Strengths of Amines in
52 Non-Protolytic Solvents. *J. Phys. Chem.* **1956**, 60, 63-70.
- 53
54 42. Erdemgil, F. Z.; Şanlı, S.; Şanlı, N.; Özkan, G.; Barbosa, J.; Guiteras, J.; Beltrán, J.
55 L., Determination of Pka Values of Some Hydroxylated Benzoic Acids in Methanol–Water
56 Binary Mixtures by Lc Methodology and Potentiometry. *Talanta* **2007**, 72, 489-496.
- 57
58 43. Szakacs, Z.; Noszal, B., Protonation Microequilibrium Treatment of Polybasic
59 Compounds with Any Possible Symmetry. *J. Math. Chem.* **1999**, 26, 139-155.
- 60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

44. Camaioni, D. M.; Schwerdtfeger, C. A., Comment on "Accurate Experimental Values for the Free Energies of Hydration of H⁺, OH⁻, and H₃O⁺". *J. Phys. Chem. A* **2005**, 109, 10795-10797.

TOC

