

### ¿Qué pasa con la Química Computacional y los equilibrios de acidez?

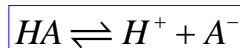
“With respect to acid/base equilibria, the prediction of an absolute  $pK_a$  by any first-principles theory remains extraordinarily hard, in part because of the difficulty in computing accurate gas-phase deprotonation free energies and also because of the large magnitude of solvation effects for charged species.”

“Typical accuracies in the computed quantities are rarely better than 2-5 kcal/mol, which results in absolute errors of possibly several log units in  $pK_a$ s a disappointing result given the accuracy possible from experimental measurement (when such measurements can be made).”

“Fortunately, trends in  $pK_a$ s across similar compounds tend to be much more accurately predicted”

Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, 99, 2161.

### ¿Cuáles son los problemas fundamentales?



- El número y tipo de especies cargadas no se conserva a ambos lados del equilibrio.
- No hay métodos que reproduzcan energías en solución comparables a Gn o CBS para fase gas.
- El tamaño de la mayor parte de los sistemas de interés.
- Modelos continuos de solvente: no interacciones específicas soluto-solvente (corto alcance).
- Modelos discretos: es necesario incluir muchas moléculas de solvente para reproducir interacciones de largo alcance ∴ computacionalmente inviables (con DFT or WFT).

Un error de 1.36 kcal/mol en  $\Delta G \rightarrow$   
error de 1 unidad de  $pK_a$

¿Dónde estamos?

Desviaciones de 2 unidades de pKa se consideran con exactitud aceptable.



**Búsqueda de estrategias que maximicen la cancelación de errores**

- Diferentes formas de modelar el solvente
- Diferente número de moléculas de agua explícitas
- Diferentes esquemas de reacción
- Diferentes niveles de cálculo (DFT, WFT, ...)
- Diferentes funciones de base
- Separación gas-solución (ciclos termodinámicos, Hess)

Número de variaciones  $\rightarrow \infty$



*Divide et impera*

¿Dónde estamos?

**Solvente**

- Modelos Continuos
- Modelos discretos
- Modelos mixtos

**Modelos Continuos y Mixtos:**

Optimización de Geometrías  
Cálculos de Frecuencias  
En fase gas  
( $\Delta G_{\text{gas}}$ )  
+  
Punto simple en solución  
( $\Delta G_{\text{solvatación}}$ )

Optimización de Geometrías  
Cálculos de Frecuencias  
En solución  
( $\Delta G_{\text{sol}}$ )

Ciclos termodinámicos

### Algunos Esquemas de reacción :

Scheme	Equilibrium	$pK_a =$
<i>A</i>	$HA \leftrightarrow H^+ + A^-$	$\frac{\Delta G_z}{RT \ln(10)}$
<i>B</i>	$HA + H_2O \leftrightarrow H_3O^+ + A^-$	$\frac{\Delta G_z}{RT \ln(10)} - \log[H_2O]$
<i>C1</i>	$HA + 2H_2O \leftrightarrow H_3O^+ + A^- (H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 2\log[H_2O]$
<i>C2</i>	$HA + 3H_2O \leftrightarrow H_3O^+ + A^- (2H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 3\log[H_2O]$
<i>C3</i>	$HA + 4H_2O \leftrightarrow H_3O^+ + A^- (3H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 4\log[H_2O]$
<i>D</i>	$HA + Ref^- \leftrightarrow A^- + HRef$	$\frac{\Delta G_z}{RT \ln(10)} + pK_a(HRef)$
<i>E1</i>	$HA + OH^- (3H_2O) \leftrightarrow A^- (H_2O) + 3H_2O$	$\frac{\Delta G_z}{RT \ln(10)} + 14 + 3\log[H_2O]$
<i>E2</i>	$HA + OH^- (3H_2O) \leftrightarrow A^- (2H_2O) + 2H_2O$	$\frac{\Delta G_z}{RT \ln(10)} + 14 + 2\log[H_2O]$
<i>E3</i>	$HA + OH^- (3H_2O) \leftrightarrow A^- (3H_2O) + H_2O$	$\frac{\Delta G_z}{RT \ln(10)} + 14 + \log[H_2O]$

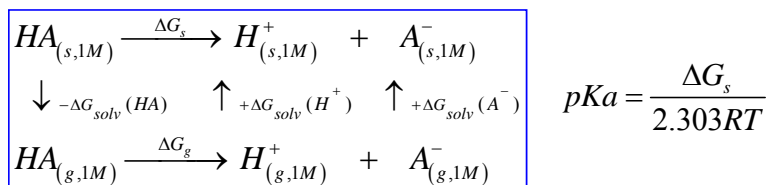
<i>F1</i>	$HA + 4H_2O \leftrightarrow H_3O^+ (3H_2O) + A^-$	$\frac{\Delta G_z}{RT \ln(10)} - 4\log[H_2O]$
<i>F2</i>	$HA + 5H_2O \leftrightarrow H_3O^+ (3H_2O) + A^- (H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 5\log[H_2O]$
<i>F3</i>	$HA + 6H_2O \leftrightarrow H_3O^+ (3H_2O) + A^- (2H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 6\log[H_2O]$
<i>F4</i>	$HA + 7H_2O \leftrightarrow H_3O^+ (3H_2O) + A^- (3H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 7\log[H_2O]$
<i>CN1</i>	$HA(H_2O) + H_2O \leftrightarrow H_3O^+ + A^- (H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - \log[H_2O]$
<i>CN2</i>	$HA(H_2O) + 2H_2O \leftrightarrow H_3O^+ + A^- (2H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 2\log[H_2O]$
<i>CN3</i>	$HA(H_2O) + 3H_2O \leftrightarrow H_3O^+ + A^- (3H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 3\log[H_2O]$
<i>FN1</i>	$HA(H_2O) + 4H_2O \leftrightarrow H_3O^+ (3H_2O) + A^- (H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 4\log[H_2O]$
<i>FN2</i>	$HA(H_2O) + 5H_2O \leftrightarrow H_3O^+ (3H_2O) + A^- (3H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 5\log[H_2O]$
<i>FN3</i>	$HA(H_2O) + 6H_2O \leftrightarrow H_3O^+ (3H_2O) + A^- (3H_2O)$	$\frac{\Delta G_z}{RT \ln(10)} - 6\log[H_2O]$
<i>EN1</i>	$HA(H_2O) + OH^- (3H_2O) \leftrightarrow A^- (H_2O) + 4H_2O$	$\frac{\Delta G_z}{RT \ln(10)} + 14 + 4\log[H_2O]$
<i>EN2</i>	$HA(H_2O) + OH^- (3H_2O) \leftrightarrow A^- (2H_2O) + 3H_2O$	$\frac{\Delta G_z}{RT \ln(10)} + 14 + 3\log[H_2O]$
<i>EN3</i>	$HA(H_2O) + OH^- (3H_2O) \leftrightarrow A^- (3H_2O) + 2H_2O$	$\frac{\Delta G_z}{RT \ln(10)} + 14 + 2\log[H_2O]$

¿y esos no son todos?!!!



**Los más usados:**

**Esquema A:**



$$\Delta G_{sol} = \Delta G_{gas} + \Delta \Delta G_s$$

$$\Delta \Delta G_s = \Delta G_s(A^-) + \Delta G_s(H^+) - \Delta G_s(HA)$$

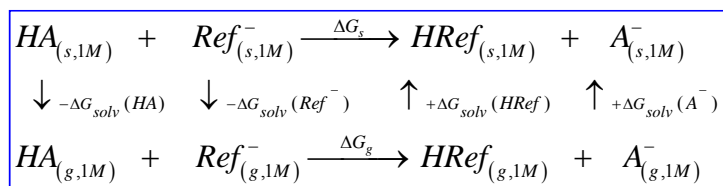
$$\text{exp} \left\{ \begin{array}{l}
 \Delta G_{gas}(H^+) = -6.28 \text{ kcal/mol} \quad (1 \text{ atm}) \\
 \Delta G_{gas}(H^+) = -4.39 \text{ kcal/mol} \quad (1 \text{ M}) \\
 \Delta G_s(H^+) = -264.61 \text{ kcal/mol}
 \end{array} \right. + RT \ln(V_M) = 1.89 \text{ kcal/mol}$$

-259 to -265 kcal/mol

**Los más usados:**

**Esquema D:**

Errores pequeños



$$pKa = \frac{\Delta G_s}{2.303RT} + pKa[HRef]$$

**Esquemas A y D:**

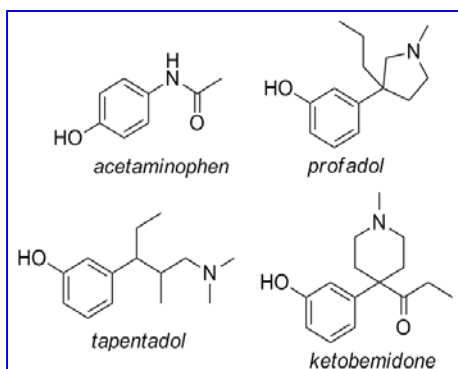
Dan resultados razonablemente buenos, especialmente **D**  
(máxima cancelación de errores)

pero...

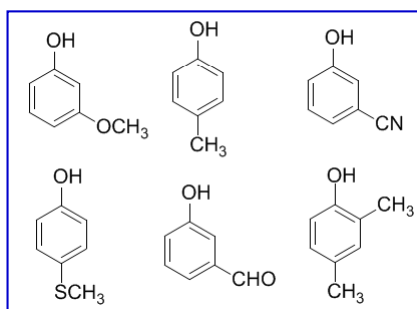
**Ambos dependen de datos experimentales**

### Caso particular: fenoles

Anti-inflamatorios no esteroidales



Otros fenoles



Conjunto de funciones de base: 6-311++G(d,p)

Solvente: PCM, radii=UAHF (HF/6-31+g(d))

Ciclos termodinámicos

22 esquemas de reacción

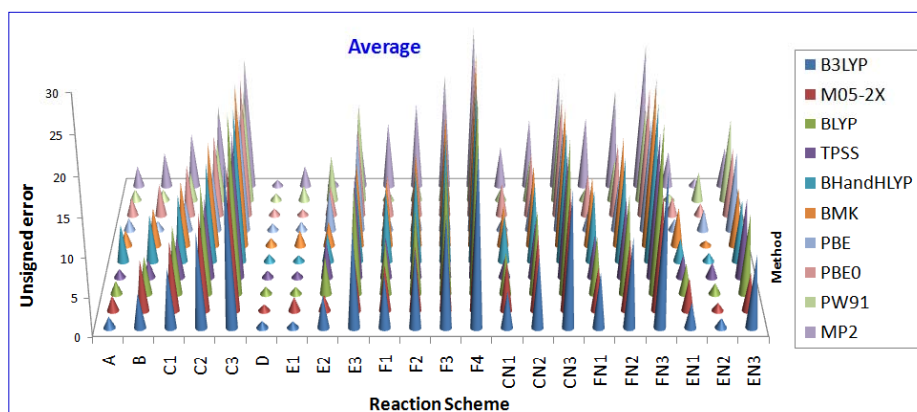
BLYP, B3LYP, BHandHLYP,

PBE, PBE0, PW91, BMK,

TPSS, M05-2x, MP2

*J. Chem. Theory Comput.* **2011**, *7*, 2528

### Caso particular: fenoles

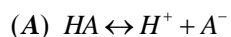


La calidad de los resultados depende más de la selección del esquema de reacción que del método de cálculo

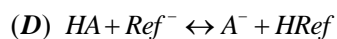
*J. Chem. Theory Comput.* **2011**, *7*, 2528

### Caso particular: fenoles

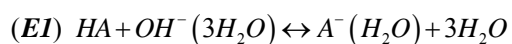
Los mejores resultados se obtuvieron con los esquemas:



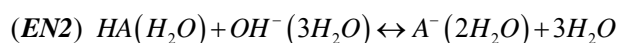
Dep. exp.



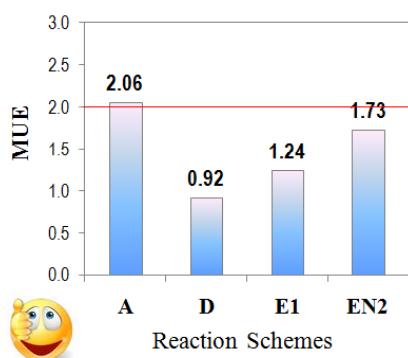
Dep. exp.



Indep. exp.



Indep. exp.



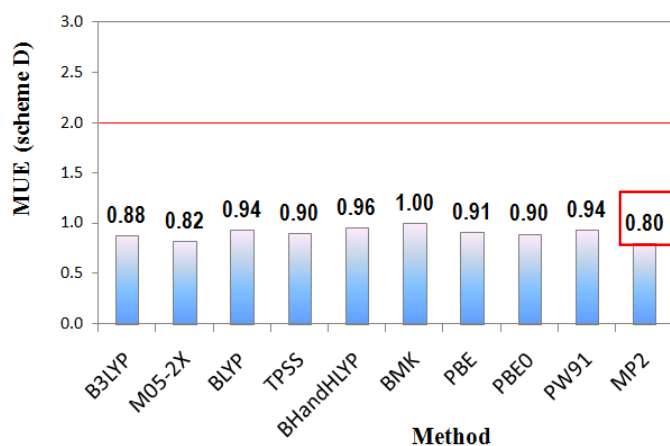
*D*, *E1* y *EN2* conservan el número y tipo de especies cargadas a ambos lados del equilibrio.

Esta parece ser la característica más importante a tener en cuenta al seleccionar el esquema de reacción.



*J. Chem. Theory Comput.* **2011**, 7, 2528

### Caso particular: fenoles

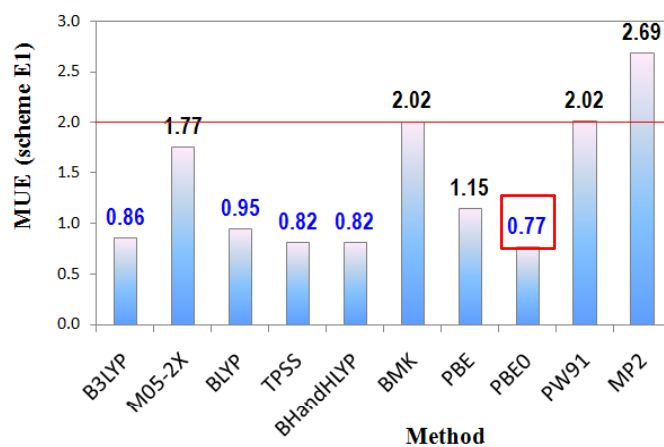


La calidad de los resultados con el esquema D es prácticamente independiente del método de cálculo utilizado y en todos los casos  $MUE \leq 1$ .

Es muy dependiente de la exactitud del  $pK_a$  experimental de  $H_{ref}$  y de cuán parecidos sean el  $H_{ref}$  y el que se está estudiando.

*J. Chem. Theory Comput.* **2011**, 7, 2528

### Caso particular: fenoles



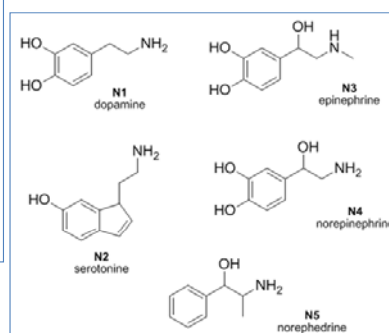
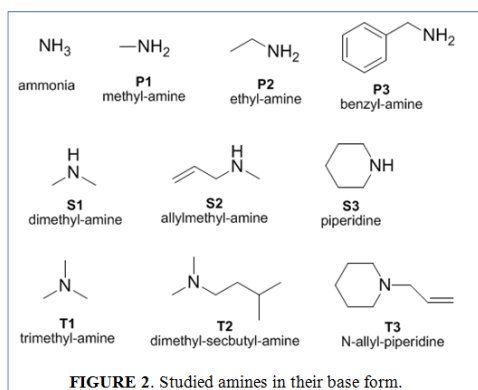
Es más sensible al método de cálculo utilizado.

MP2 → los peores resultados

Los mejores resultados ( $MUE \leq 1$ ) se obtuvieron con PBE0, TPSS, BHandHLYP, B3LYP y BLYP, en se orden.

*J. Chem. Theory Comput.* **2011**, 7, 2528

### Caso particular: aminas



**Caso particular: aminas**

Scheme	Equilibrium	$pKa =$	Used for*
A	$NH_4^+ \rightleftharpoons H^+ + NH_3$	$\frac{\Delta G_i}{RT \ln(10)}$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
B	$NH_4^+ + H_2O \rightleftharpoons H_3O^+ + NH_3$	$\frac{\Delta G_i}{RT \ln(10)} - \log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
C1	$NH_4^+(H_2O) \rightleftharpoons H_3O^+ + NH_3$	$\frac{\Delta G_i}{RT \ln(10)}$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
C2	$NH_4^+(2H_2O) \rightleftharpoons H_3O^+ + NH_3 + H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + \log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$
C3	$NH_4^+(3H_2O) \rightleftharpoons H_3O^+ + NH_3 + 2H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 2\log[H_2O]$	$NH_4^+$ , $HP^+$
C4	$NH_4^+(4H_2O) \rightleftharpoons H_3O^+ + NH_3 + 3H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 3\log[H_2O]$	$NH_4^+$
CN1	$NH_4^+(H_2O) + H_2O \rightleftharpoons H_3O^+ + NH_3(H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - \log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
CN2	$NH_4^+(2H_2O) \rightleftharpoons H_3O^+ + NH_3(H_2O)$	$\frac{\Delta G_i}{RT \ln(10)}$	$NH_4^+$ , $HP^+$ , $HS^+$
CN3	$NH_4^+(3H_2O) \rightleftharpoons H_3O^+ + NH_3(H_2O) + H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + \log[H_2O]$	$NH_4^+$ , $HP^+$
CN4	$NH_4^+(4H_2O) \rightleftharpoons H_3O^+ + NH_3(H_2O) + 2H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 2\log[H_2O]$	$NH_4^+$
D	$NH_4^+ + Ref \rightleftharpoons NH_3 + HRef$	$\frac{\Delta G_i}{RT \ln(10)} + pKa(HRef)$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$

**Caso particular: aminas**

E1	$NH_4^+(H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3 + 5H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 5\log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
E2	$NH_4^+(2H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3 + 6H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 6\log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$
E3	$NH_4^+(3H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3 + 7H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 7\log[H_2O]$	$NH_4^+$ , $HP^+$
E4	$NH_4^+(4H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3 + 8H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 8\log[H_2O]$	$NH_4^+$
EN1	$NH_4^+(H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3(H_2O) + 4H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 4\log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
EN2	$NH_4^+(2H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3(H_2O) + 5H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 5\log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$
EN3	$NH_4^+(3H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3(H_2O) + 6H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 6\log[H_2O]$	$NH_4^+$ , $HP^+$
EN4	$NH_4^+(4H_2O) + OH^-(3H_2O) \rightleftharpoons NH_3(H_2O) + 7H_2O$	$\frac{\Delta G_i}{RT \ln(10)} + 14 + 7\log[H_2O]$	$NH_4^+$
F1	$NH_4^+ + 4H_2O \rightleftharpoons NH_3 + H_3O^+(3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - 4\log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
F2	$NH_4^+(H_2O) + 3H_2O \rightleftharpoons NH_3 + H_3O^+(3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - 3\log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
F3	$NH_4^+(2H_2O) + 2H_2O \rightleftharpoons NH_3 + H_3O^+(3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - 2\log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$
F4	$NH_4^+(3H_2O) + H_2O \rightleftharpoons NH_3 + H_3O^+(3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - \log[H_2O]$	$NH_4^+$ , $HP^+$

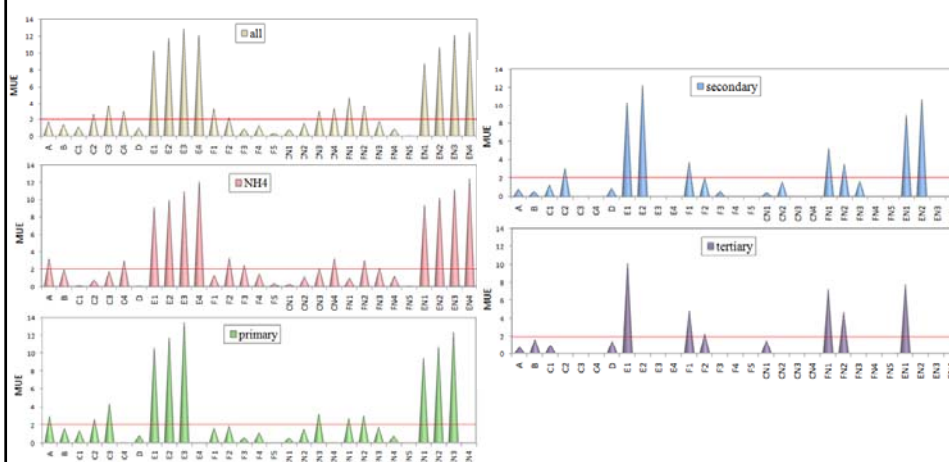


### Caso particular: aminas

F5	$NH_4^+ (4H_2O) \rightleftharpoons NH_3 + H_3O^+ (3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)}$	$NH_4^+$
FN1	$NH_4^+ + 5H_2O \rightleftharpoons NH_3(H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - 5 \log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
FN2	$NH_4^+ (H_2O) + 4H_2O \rightleftharpoons NH_3(H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - 4 \log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$ , $HT^+$
FN3	$NH_4^+ (2H_2O) + 3H_2O \rightleftharpoons NH_3(H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - 3 \log[H_2O]$	$NH_4^+$ , $HP^+$ , $HS^+$
FN4	$NH_4^+ (3H_2O) + 2H_2O \rightleftharpoons NH_3(H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - 2 \log[H_2O]$	$NH_4^+$ , $HP^+$
FN5	$NH_4^+ (4H_2O) + H_2O \rightleftharpoons NH_3(H_2O) + H_3O^+ (3H_2O)$	$\frac{\Delta G_i}{RT \ln(10)} - \log[H_2O]$	$NH_4^+$

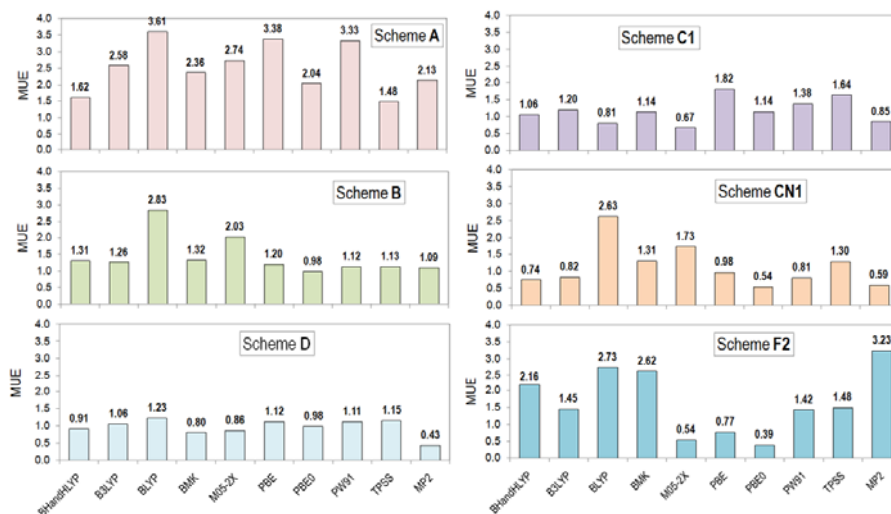
\*  $HP^+$ =protonated primary amines,  $HS^+$ =protonated secondary amines,  $HT^+$ =protonated tertiary amines

### Caso particular: aminas



Mean unsigned errors (MUE) for the  $pK_a$  values calculated with BHandHLYP.

### Caso particular: aminas



Mean unsigned errors (MUE) for the calculated  $pK_a$  values of the whole set of amines

### Caso particular: aminas

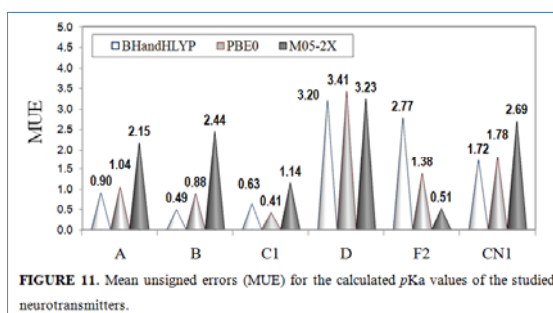


FIGURE 11. Mean unsigned errors (MUE) for the calculated  $pK_a$  values of the studied neurotransmitters.

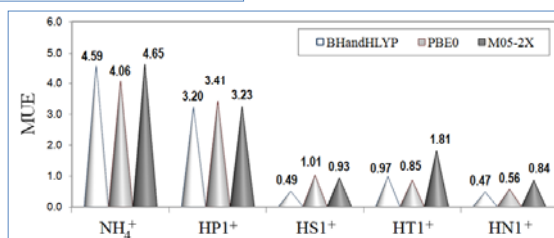
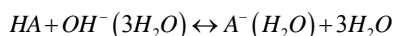


FIGURE 12. Mean unsigned errors (MUE) for the  $pK_a$  values of the studied neurotransmitters calculated with scheme D and different  $H_{Ref}$ .

***fenoles***

Los mejores resultados ( $MUE \leq 1$ ) se obtuvieron con el esquema **E1** y PBE0, TPSS, BHandHLYP, B3LYP y BLYP, en se orden.



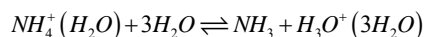
La calidad de los resultados con el esquema D es prácticamente independiente del método de cálculo utilizado y en todos los casos  $MUE \leq 1$ .

Es muy dependiente de la exactitud del  $pK_a$  experimental de Href y de cuan parecidos sean el HRef y el que se está estudiando.

conservar el número y tipo de especies cargadas a ambos lados del equilibrio parece ser la característica más importante a tener en cuenta al seleccionar el esquema de reacción

***aminas***

Los mejores resultados ( $MUE \leq 1$ ) se obtuvieron con el esquema **F2** y M05-2X.



La calidad de los resultados con el esquema D es prácticamente independiente del método de cálculo.

Es muy dependiente de la amina que se escoge como referencia, se pueden obtener valores de  $MUE > 4$  unidades de  $pK_a$  en neurotransmisores usando una amina primaria y  $MUE > 3$  unidades de  $pK_a$  usando una amina secundaria.

**Recomendaciones Generales:****Equilibrios de Acidez**

- Describir órdenes de deprotonación (descripción cualitativa) para sistemas polipróticos puede hacerse de manera confiable con varias metodologías.
- Estimar constantes de acidez o  $pK_a$ s (descripción cuantitativa) es un reto computacional. Lograr exactitudes comparables a la experimental es difícil de lograr. En general requiere probar diferentes metodologías y validar la calidad de los resultados obtenidos contra valores experimentales de sistemas similares.
- Probar varios modelos para cada tipo de sistema concreto (No hay una estrategia única que sea la que de mejores resultados para todos los sistemas. Ej fenoles vs. aminas)
- Utilizar para  $H^+$  los valores de energías libres de Gibbs experimentales (ciclo A)
- Las estrategias que hasta ahora han probado una mayor eficacia son **OPT + FREQ en solución (SMD) para obtener directamente  $\Delta G_s$**  y **OPT + FREQ en fase gas para obtener la  $\Delta G_g$  y puntos simples en solución (PCM) para obtener  $\Delta \Delta G_{solv}$**  (Hess:  $\Delta G_s = \Delta G_g + \Delta \Delta G_{solv}$ )
- Ambas estrategias pueden utilizarse incluyendo o no moléculas de agua explícitas. La segunda variante ha probado ser mejor para sistemas con grupos fenólicos.
- Escoger muy cuidadosamente el Href si se va a utilizar el ciclo D, no siempre es suficiente que sea el mismo tipo de compuesto (Ej. Neurotransmisores, aminas secundarias)
- El número óptimo de aguas explícitas dependerá del compuesto (Ej. Fenoles,  $NH_4^+$ , aminas primarias, secundarias y terciarias)