

Cálculos Cuánticos de Resonancia Magnética Nuclear Empleados en el Desarrollo de una Probabilidad para Asignación Estereoquímica

Zanardi, María Marta[†]; Sarotti, Ariel Marcelo[‡].

[†]Facultad de Química e Ingeniería del Rosario, Pontificia Universidad Católica Argentina, Av. Pellegrini 3314, Rosario 2000, Argentina.

[‡]Instituto de Química Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario 2000, Argentina

Resumen. *La caracterización y elucidación estructural de nuevas moléculas orgánicas complejas es una tarea altamente desafiante. Entre las diferentes técnicas espectroscópicas para llevar a cabo este propósito, la Resonancia Magnética Nuclear es la más útil. Una vez que se ha determinado la conectividad total de la molécula, el trabajo más difícil es determinar su estructura 3D sin errores estereoquímicos. La química computacional moderna ha contribuido significativamente a prevenir estos errores. En este contexto, hemos desarrollado algunas herramientas para interpretar la correlación entre los datos de RMN experimentales y calculados que contribuyen significativamente al proceso de elucidación estructural.*

La probabilidad de DP4 es uno de los enfoques más sofisticados y populares para la asignación estereoquímica de moléculas orgánicas usando cálculos de desplazamiento químico GIAO NMR cuando solo está disponible un conjunto de datos experimentales. Para mejorar el rendimiento del método, hemos desarrollado una probabilidad modificada (DP4+), cuyas diferencias principales con respecto a la DP4 original son la inclusión de datos sin escala y el uso de niveles de teoría más elevados para el procedimiento de cálculo de RMN. Con estas modificaciones, se logró una mejora significativa en el rendimiento general, proporcionando resultados precisos y confiables en el establecimiento de la estereoquímica de 48 compuestos isoméricos desafiantes.

Palabras Clave: Química computacional, RMN, Asignación estereoquímica, DP4+.

Abstract. *The characterization and structural elucidation of isolated complex organic molecules are a highly challenging task. Among the different spectroscopic techniques to carry out this purpose, the Nuclear Magnetic Resonance is the most useful one. Once the total connectivity of the molecule has been determined, the hardest job is to determine its 3D structure without stereochemical errors. Modern computational chemistry has contributed significantly to preventing these errors. In this context, we have developed some tools to interpret the correlation between experimental and calculated NMR data that contribute significantly to structural elucidation process.*

The DP4 probability is one of the most sophisticated and popular approaches for the stereochemical assignment of organic molecules using GIAO NMR chemical shift calculations when only one set of experimental data is available. In order to improve the performance of the method, we have developed a modified probability (DP4+), whose main differences from the original DP4 are the inclusion of unscaled data and the use of higher levels of theory for the NMR calculation procedure. With these modifications, a significant improvement in the overall performance was achieved providing accurate and confident results in establishing the stereochemistry of 48 challenging isomeric compounds.

Keywords: Computational chemistry, NMR, Stereochemical Assignment, DP4+.

Introducción.

Antiguamente la ciencia química era considerada como un campo experimental alejado de consideraciones matemáticas. Sin embargo con la llegada de la Mecánica Cuántica los pioneros

en ese campo comenzaron a darse cuenta que tenía el potencial de ser una teoría predictiva en química [1].

Actualmente, la investigación científica necesita cada vez más de una integración multidisciplinar que permita aprovechar las distintas herramientas que ofrece cada área científica en post de un objetivo común. Así surge la química computacional como una disciplina que se extiende más allá de los límites tradicionales que separan la química, la física, la biología y la ciencia de la computación; permitiendo la investigación del comportamiento de la materia a nivel molecular mediante un sistema de ordenadores, cuando la investigación de laboratorio sea inapropiada, poco práctica o imposible. Por este motivo, la química computacional se encuentra estrechamente relacionada con la síntesis orgánica ya que permite establecer puntos de partida para trabajos sintéticos, determinar estructuras moleculares y mecanismos de reacción, estudiar especies muy reactivas que no pueden ser aisladas, interpretar y predecir datos experimentales, así como identificar correlaciones entre estructuras químicas y sus propiedades [2]. La química computacional se encarga de incorporar la química teórica a un software informático con el objeto de calcular las estructuras y propiedades de moléculas, ya sea empleando la mecánica molecular o las teorías estructurales electrónicas basadas en la mecánica cuántica [3]. Dada la gran precisión en las simulaciones alcanzada con el modelado molecular, es importante poder combinarlas con métodos quimiométricos que permitan analizar bases de datos y detectar patrones de comportamiento entre variables calculables o medibles con propiedades o actividades químicas o biológicas de interés, y en base a ellos poder realizar predicciones sobre nuevos casos.

Un área que ha sido altamente beneficiada con el advenimiento de la química computacional ha sido la caracterización estructural de nuevos compuestos orgánicos. Cada nuevo producto natural o sintético que se determina presenta alguna actividad interesante es necesario que sea adecuadamente caracterizado ya que sus propiedades químicas y biológicas están íntimamente relacionadas con su estructura tridimensional.

Si bien existen una gran variedad de técnicas instrumentales para efectuar dicha tarea, sin dudas la espectroscopia de resonancia magnética nuclear (RMN) constituye la herramienta más valiosa que disponemos los químicos orgánicos para determinar la naturaleza estructural y estereoquímica de los compuestos que se sintetizan o aíslan de fuentes naturales.

En los últimos años se ha dado un gran desarrollo para el perfeccionamiento de las técnicas experimentales de RMN, pero aún con el advenimiento de espectrómetros cada vez más poderosos y de nuevas y sofisticadas secuencias de pulsos, la asignación incorrecta de estructuras moleculares continúa siendo una situación muy común [4]. Esto se ve reflejado con los cientos de revisiones estructurales publicados en los últimos años de productos naturales que fueron erróneamente asignados [5]. Si bien el origen de los errores puede ser muy variado, la ambigüedad de señales, impurezas en la muestra, errores humanos y estructuras moleculares muy complejas suelen ser las causas más comunes. Esto se debe a que el análisis de datos espectroscópicos es un proceso complejo, que involucra múltiples etapas y requiere de una gran experiencia y entrenamiento, sumados a buenas herramientas y datos confiables. Teniendo en cuenta que las discrepancias a menudo se detectan después de la síntesis total de la estructura originalmente propuesta (incorrecta), no es descabellado suponer que la arquitectura molecular real de muchos productos naturales publicados continúe siendo desconocida.

Con el objetivo de reducir esta problemática y su impacto sobre la comunidad científica, en los últimos años se han desarrollado distintas metodologías que emplean cálculos computacionales basados en métodos de química cuántica. Estos enfoques se han potenciado conforme avanza la tecnología informática, con computadoras cada vez más potentes y el desarrollo de paquetes de software más eficientes. Hoy día muchas propiedades moleculares y espectroscópicas pueden ser predichas con notables niveles de precisión y exactitud. En particular, los cálculos de resonancia magnética nuclear (incluyendo desplazamientos químicos y constantes de acoplamiento), pueden emplearse para simplificar los procesos de asignación espectroscópica (es decir determinar qué señal de resonancia corresponde a qué núcleo), tarea que suele ser compleja aún con la asistencia de técnicas de RMN bidimensionales. También revisten especial importancia en los procesos de elucidación, validación y revisión estructural de moléculas orgánicas [6]. Recientemente [7] hemos desarrollado una novedosa estrategia de validación estructural basada en redes neuronales artificiales, que permite correlacionar un set de datos experimentales con un set de datos calculados arrojando según el ajuste entre ellos si la estructura propuesta es correcta o no. Esta herramienta es empleada una vez que se ha determinado completamente la estructura, pero muchas veces antes de llegar a la estructura propuesta se debe realizar la asignación estereoquímica, es decir en función del número de estereocentros (n) que posee la molécula cuál de los 2^n posibles isómeros corresponde a la estructura a determinar. Para ello hemos trabajado en una probabilidad llamada DP4+ que permite asignar una probabilidad a cada uno de los isómeros [8].

Aplicación de cálculos teóricos de RMN en la asignación estereoquímica de compuestos orgánicos

En principio, se contaba con dos estrategias principales con respecto al uso de los cálculos químicos cuánticos de los desplazamientos de RMN en la elucidación estructural. Por un lado, la exactitud de una estructura putativa dada (solo una) se evalúa con la única información proporcionada por la RMN experimental registrada para ese compuesto y los desplazamientos químicos calculados para la propuesta estructural. Recientemente hemos demostrado que el análisis de reconocimiento de patrones a través de redes neuronales artificiales resultó en un método prometedor en este campo.⁴ Por otro lado, en un enfoque conceptualmente diferente, dos o más candidatos se evalúan y se correlacionan con al menos un conjunto de cambios experimentales siguiendo una metodología basada en la comparación. En este sentido, Smith y Goodman han logrado un gran avance al introducir primero el parámetro CP3 y más tarde la probabilidad DP4 [9,10]. El parámetro CP3 se diseñó para asignar dos conjuntos de datos experimentales (una situación común encontrada en las reacciones estereoselectivas) a dos estructuras posibles mediante la comparación de las diferencias en los desplazamientos calculados entre los dos isómeros con las correspondientes diferencias en los cambios experimentales de ambos.⁹ Una situación mucho más compleja surge cuando se dispone de un único conjunto de datos experimentales y se tienen dos o más posibles candidatos, en este caso el parámetro CP3 no puede ser empleado. Este es a menudo el caso de productos naturales, pero también de reacciones orgánicas con niveles perfectos de estereoselectividad. De esta situación surgió la probabilidad DP4 como una poderosa herramienta empleada en los últimos años, para confirmar o corregir la arquitectura molecular de numerosos compuestos orgánicos.

En nuestro grupo de trabajo hemos desarrollado una nueva probabilidad mejorada (DP4+), erigida sobre los cimientos matemáticos del DP4, que garantiza un nivel de predicción y confianza muy superiores a los establecidos por el formalismo original [8].

Resultados y discusión

La probabilidad DP4 tiene su base en la observación llevada a cabo por Smith y Goodman de que los errores e entre desplazamientos químicos experimentales y calculados (escalados) ($e = \delta_{sc} - \delta_{exp}$) para un conjunto de moléculas orgánicas obedece a la distribución t con una media $\mu = 0$ (como consecuencia del procedimiento de escalado lineal), desviación estándar σ , y de grados de libertad ν . Entonces, para una molécula dada con N núcleos, se puede calcular la probabilidad de cada uno de los i errores. Suponiendo que el conjunto de errores es una variable aleatoria independiente, la multiplicación de las probabilidades individuales asigna la probabilidad total de esa estructura candidata. Finalmente, las probabilidades resultantes se convierten en un conjunto de probabilidades porcentuales para cada isómero utilizando el teorema de Bayes. Matemáticamente, la probabilidad de DP4 se definió de la siguiente manera:

$$P(i) = \frac{\prod_{k=1}^N (1 - T^{\nu}(\delta_{s,k}^j - \delta_{exp,k})/\sigma)}{\sum_{j=1}^m [\prod_{k=1}^N (1 - T^{\nu}(\delta_{s,k}^j - \delta_{exp,k})/\sigma)]} \quad \text{Eq. 1}$$

T^{ν} , σ y δ_s^j calculados sobre geometrías MMFF

Donde $P(i)$ es la probabilidad de que la estructura i (de m candidatos posibles) sea la correcta. T^{ν} da la función de distribución t acumulativa con ν grados de libertad y varianza σ . $\delta_{exp,k}$ es el desplazamiento químico experimental para el núcleo k (de los N) y $\delta_{s,k}$ representa el desplazamiento calculado para el núcleo k después del procedimiento de escalado para eliminar los errores sistemáticos. Esto se hace según $\delta_s = (\delta_{calc} - b) / m$, donde b y m son la intersección y la pendiente de una gráfica de δ_{calc} contra δ_{exp} . [10]. De la Eq. 1, σ y ν son los términos clave en el cálculo de la probabilidad DP4 y se debe determinar ajustando los datos (errores entre δ escalados y experimentales) de un conjunto de datos grande a una distribución utilizando programas estadísticos específicos.

Un minucioso análisis del método nos permitió identificar dos posibles inconvenientes en la arquitectura DP4: el nivel de teoría empleado en los cálculos y el uso exclusivo de los desplazamientos químicos escalados. Con respecto al primer problema, se utilizó el nivel

B3LYP/6-31G**//MMFF en la formulación DP4 original para proporcionar buenos resultados en el cálculo de desplazamientos de RMN a bajo costo computacional (evitando principalmente costosos tratamientos *ab initio* o DFT para la optimización de la geometría). Sin embargo, desde nuestra experiencia, está bastante lejos de ser el método más preciso para los cálculos de RMN, principalmente en la predicción de los δ de ^1H . Esto puede verse reflejado en la Figura 1 donde se comparan los Errores Promedio Corregidos (CMAE por sus siglas en inglés *Corrected Mean Average Errors*), los cuales van disminuyendo a medida que se incrementa el costo computacional del nivel de teoría utilizado en los cálculos.

level of theory	CMAE		CMAE	
	^1H	^{13}C	^1H	^{13}C
B3LYP/6-31G** //MMFF	0.12	1.5	0.21	1.9
B3LYP/6-31G** //B3LYP/6-31G*	0.08	0.9	0.12	1.2
PCM/mPW1PW91/6-31+G** //B3LYP/6-31G*	0.06	0.6	0.09	1.1

Figura 1. Errores medios promedios corregidos (CMAE) para los compuestos 4 y 11 calculados en diferentes niveles de teoría.

La principal contribución a estos errores puede ser atribuida al paso de optimización de la geometría (MMFF- Método basado en la mecánica molecular) y no al método empleado para realizar los cálculos de RMN, ya que incluso pequeños errores en la geometría molecular pueden llevar a grandes errores en los desplazamientos químicos. Por lo tanto, consideramos que si se optimizara la geometría molecular y se calcularan los δ de RMN a niveles de teoría superiores permitiría predicciones más precisas y, por lo tanto, probabilidades de DP4 más confiables.

Por otro lado, planteamos si el uso exclusivo de los δ escalados para calcular la probabilidad de DP4 no conllevaría a la pérdida de información importante para la asignación. El escalado lineal se emplea típicamente para eliminar errores sistemáticos, lo que lleva a δ corregidos que están más cerca de los valores experimentales [11]. Sin embargo, esta práctica supone que la magnitud de un error es independiente del entorno químico (por ejemplo, la hibridación ^{13}C) que, en general, no es así. Eventualmente puede suceder que uno de los candidatos presente un mejor ajuste que el isómero correcto, por ello consideramos que incluyendo desplazamientos químicos sin escalar, por ende más dependientes del entorno, se podría enfatizar el contraste producido por el entorno.

Por lo expuesto, consideramos que el uso de niveles de teoría más precisos para el procedimiento de cálculo de RMN y la inclusión de los desplazamientos sin escalar podría dar lugar a una mejor probabilidad de tipo DP4. La formulación básica de la nueva probabilidad DP4+ para asignar un conjunto de datos experimentales a una de muchas estructuras diferentes se muestra en la Ecuación 2. La probabilidad de que el candidato i (de m isómeros) represente la estructura correcta, $P(i)$, se da como una función de las probabilidades correspondientes calculadas usando δ escalados y sin escalar, $P(i)_s$ y $P(i)_u$. A su vez se pueden calcular utilizando el formalismo DP4 estándar (ver Eq 1).

$$\begin{array}{ccc}
 \text{Empleo de} & \leftarrow \boxed{\text{DP4+}} \rightarrow & \text{Cálculo de } \delta \\
 \text{cálculos de RMN} & & \text{escalados y no} \\
 \text{más exactos} & & \text{escalados} \\
 & & \text{Eq 2.} \\
 \hline
 P(i) = f [P(i)_s ; P(i)_u] \\
 \hline
 P(i)_s = f [\mu_s, \sigma_s, \delta_s] \text{ empleando errores escalados} \\
 P(i)_u = f [\mu_u, \sigma_u, \delta_u] \text{ empleando errores no escalados} \\
 \hline
 \end{array}$$

Para construir la probabilidad DP4+, se calcularon los términos T^v_s , σ_s , T^v_u , σ_u y μ_u (dependientes del nivel de teoría empleado en el procedimiento de cálculo de RMN) para ^{13}C y ^1H (tener en cuenta que μ_s es cero como consecuencia de la procedimiento de escalado). Para ello, se seleccionó un gran conjunto de datos constituido por 72 compuestos (Figura 2) con una amplia variedad de grupos funcionales y complejidad estructural para la cual los espectros ^1H y ^{13}C son conocidos y totalmente asignados. Los desplazamientos químicos de geometrías optimizadas a nivel B3LYP/6-31G* se calcularon a 24 niveles de teoría diferentes, combinando 2 funcionales

(B3LYP y mPW1PW91) y 6 conjuntos de bases (6-31G*, 6-31G**, 6-31+G**, 6-311G*, 6-311G** y 6-311+G**). Los cálculos de RMN, también se computaron en fase gaseosa y en solución (PCM, solvente: cloroformo).¹⁶ En el caso de moléculas conformacionalmente flexibles, se tomaron todas las conformaciones dentro de 2 kcal/mol del mínimo global, y la contribución de cada conformero fue ponderada usando Boltzmann promediando. Una vez que se hicieron los cálculos, se contaba con 1219 y 1123 errores de ^{13}C y ^1H , respectivamente, para cada nivel de teoría (tanto escalada como sin escalar) como fueron previamente definidos.

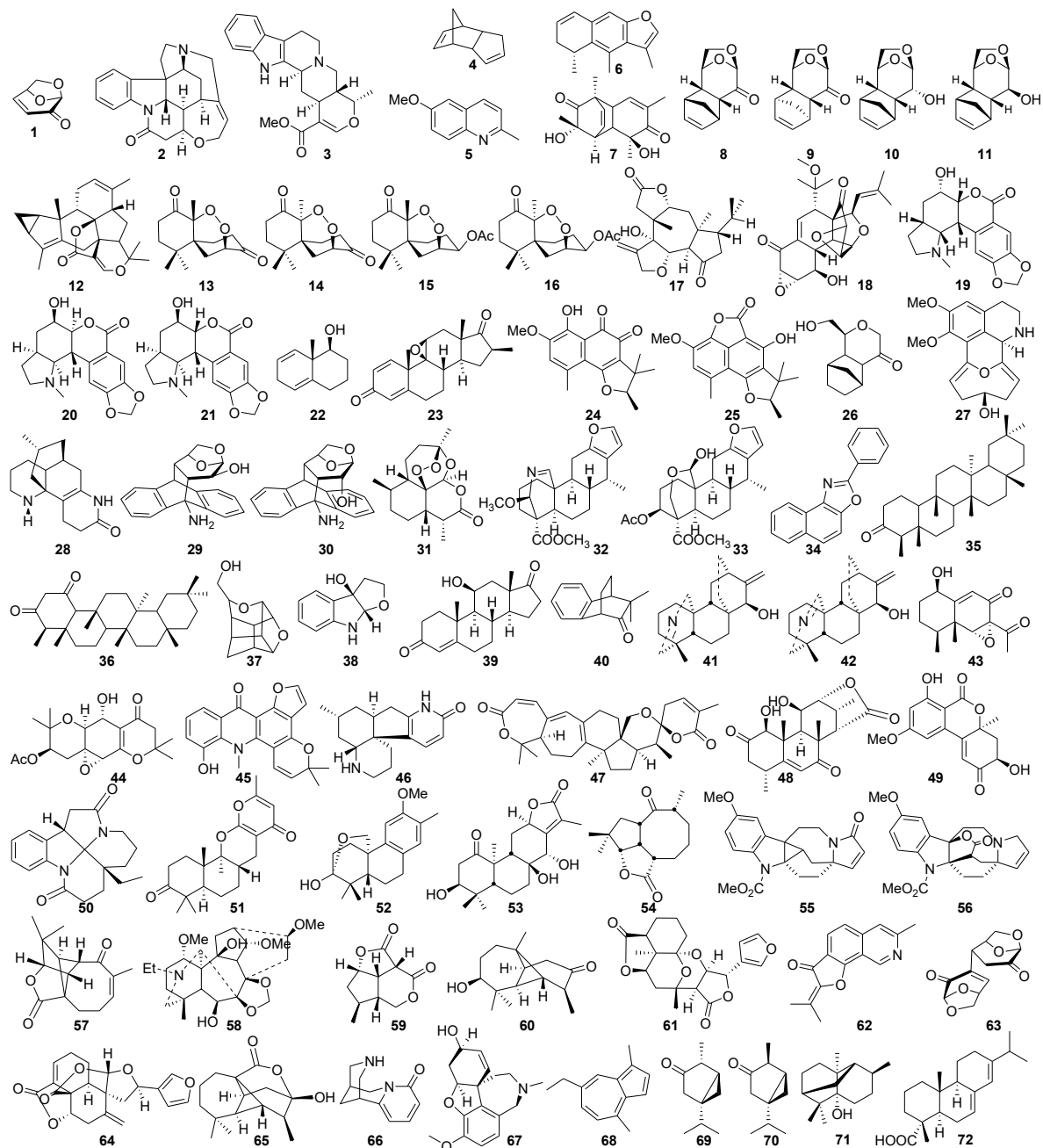


Figura 2. Set de compuestos empleados para calcular los parámetros estadísticos.

Con estos conjuntos de errores, se evaluó si obedecen una distribución t , un requisito primario para la probabilidad de DP4+. A pesar de que este fue el caso de los errores escalados (Figura 3A), encontramos que los errores no escalados no siguen esta distribución. En contraste, los histogramas correspondientes parecían estar formados por la superposición de dos series normalmente distribuidas (Figura 3B). Teniendo en cuenta que el rendimiento de TMS como patrón de referencia depende principalmente de la hibridación de los núcleos en cuestión (origen del enfoque multistandard) [12], especulamos que la serie podría derivarse de los errores de carbonos sp^2 y sp^3 (o protones unidos a los mismos). De hecho, después de separar los datos, nos alegramos de encontrar que cada serie derivada de sp^2 y sp^3 se ajustaba sin problemas a dos distribuciones t (Figura 3B). Es importante señalar que este comportamiento se observó tanto

para ^1H como para ^{13}C en los 24 niveles de teoría bajo estudio.

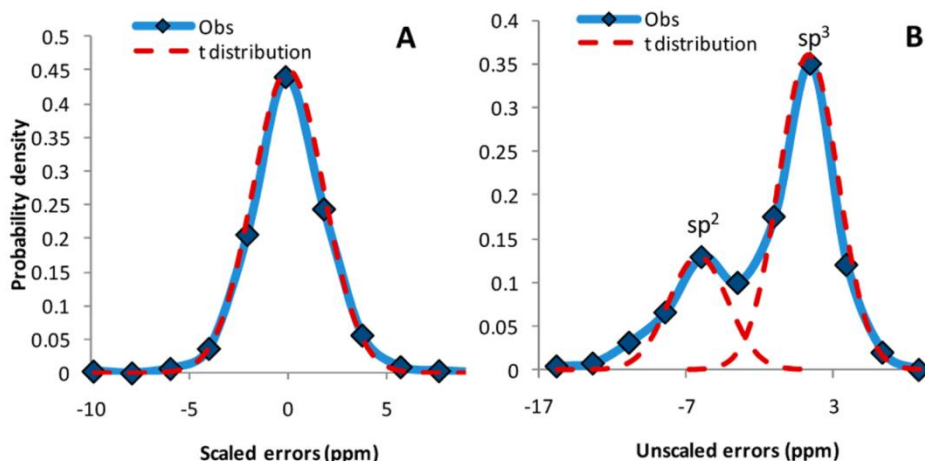


Figura 3. Distribución de los errores para los distintos tipos de desplazamientos químicos.

Este hallazgo nos permitió postular la formulación matemática de nuestra nueva probabilidad DP4+, representada en la ecuación 3.

$$\begin{aligned}
 & \text{DP4+} \\
 P(i) = & \frac{\prod_{k=1}^N [1 - T^v_s(e^j_{s,k}/\sigma_s)] \prod_{k=1}^N [1 - T^v_{u\text{-sp}x}(e^j_{u,k} - \mu_{u\text{-sp}x})/\sigma_{u\text{-sp}x}]}{\sum_{j=1}^m \prod_{k=1}^N [1 - T^v_s(e^j_{s,k}/\sigma_s)] \prod_{k=1}^N [1 - T^v_{u\text{-sp}x}(e^j_{u,k} - \mu_{u\text{-sp}x})/\sigma_{u\text{-sp}x}]} \quad \text{Eq 3.} \\
 & T^v, \sigma, \mu \text{ y } e^j \text{ calculados sobre geometrías B3LYP/6-31G}^+ \\
 & \begin{array}{cc}
 \text{sDP4+} & \text{uDP4+} \\
 \frac{\prod_{k=1}^N [1 - T^v_s(e^j_{s,k}/\sigma_s)]}{\sum_{j=1}^m \prod_{k=1}^N [1 - T^v_s(e^j_{s,k}/\sigma_s)]} & \frac{\prod_{k=1}^N [1 - T^v_{u\text{-sp}x}(e^j_{u,k} - \mu_{u\text{-sp}x})/\sigma_{u\text{-sp}x}]}{\sum_{j=1}^m \prod_{k=1}^N [1 - T^v_{u\text{-sp}x}(e^j_{u,k} - \mu_{u\text{-sp}x})/\sigma_{u\text{-sp}x}]}
 \end{array}
 \end{aligned}$$

Bajo el supuesto de que la estructura propuesta i es correcta, la probabilidad de obtener un set de errores escalados ($e_s = \delta_s - \delta_{\text{exp}}$) y sin escalar ($e_u = \delta_u - \delta_{\text{exp}}$) viene dada por la multiplicación de cada probabilidad independiente $[1 - T^v(e - \mu) / \sigma]$, calculada para cada δ escalado y sin escalar (numerador de la ecuación 3). Entonces, suponiendo que la estructura correcta se encuentra entre los m candidatos, la probabilidad de que i sea el isómero correcto, $P(i)$, se obtiene dividiendo por la suma de las probabilidades de todos los m candidatos (denominador de Eq 3). Una simple mirada a la Eq. 3 revela que la probabilidad DP4+ puede ser descompuesta en dos contribuciones principales: sDP4+ y uDP4+. El primer término es la probabilidad que se obtiene al usar exclusivamente desplazamientos escalados (como en el caso de DP4), mientras que el segundo proporciona la probabilidad cuando solo se utilizan datos sin escalar. Teniendo en cuenta que en este caso, los valores de $\mu_{u\text{-sp}x}$, $\sigma_{u\text{-sp}x}$ y $v_{u\text{-sp}x}$ dependen de la hibridación de los núcleos. De este modo, $T^v_{u\text{-sp}2}$, $\mu_{u\text{-sp}2}$ y $\sigma_{u\text{-sp}2}$ constituyen la función t acumulativa con v grados de libertad centrados en μ y la varianza σ correspondiente a los carbonos $\text{sp}2$ no escalados (o hidrógenos unidos a los carbonos $\text{sp}2$), y $T^v_{u\text{-sp}3}$, $\mu_{u\text{-sp}3}$ y $\sigma_{u\text{-sp}3}$ corresponden a los parámetros análogos de la distribución correspondiente a los núcleos $\text{sp}3$.

El rendimiento de nuestra probabilidad de DP4+ se evaluó con un desafiante set de 48 ejemplos (Figura 4) para los cuales DP4 original proporcionó resultados insatisfactorios.

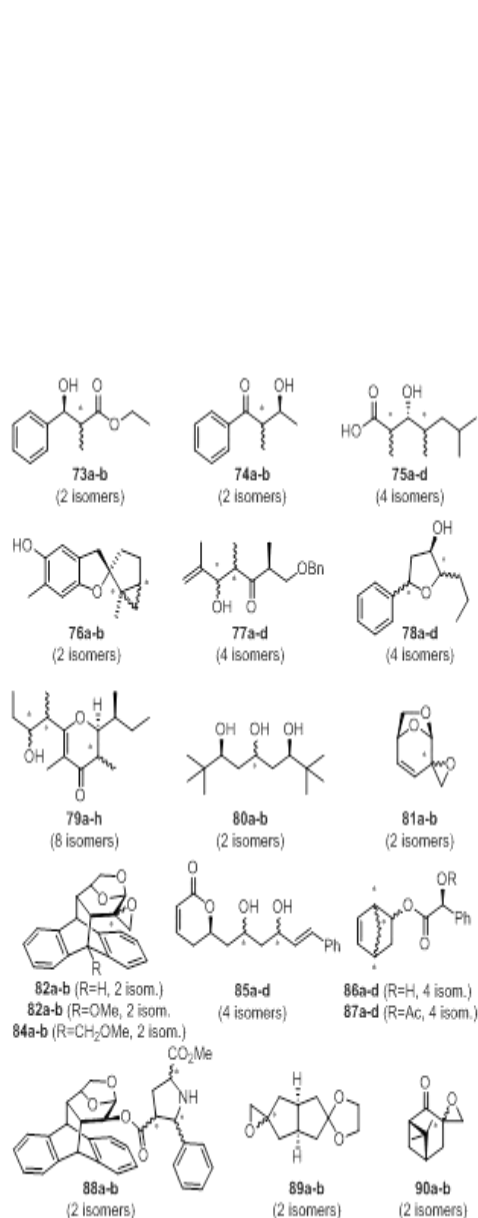


Figura 4. Set de moléculas para la evaluación del DP4+

DP4	B3LYP												mPW1PW91											
	Gas Phase						PCM						Gas Phase						PCM					
	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
73a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
73b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
74a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
74b	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
75a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
75b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
75c	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
75d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
76a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
76b	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
77a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
77b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
77d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
78a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
78b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
78c	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
78d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79c	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79e	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79f	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79g	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
79h	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
80a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
80b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
81a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
82b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
83b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
84b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
85a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
85b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
85c	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
85d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
86a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
86b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
86c	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
86d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
87a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
87b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
87c	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
87d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
88a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
88b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
89a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
90a	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	
90b	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	

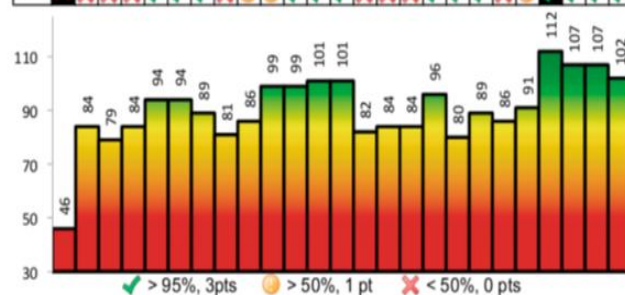


Figura 5. Desempeño general de DP4 original (primera columna) y las nuevas probabilidades de DP4+ calculadas para los compuestos 73-90 (Figura 4) en los 24 niveles de teoría bajo estudio (columnas 2 a 25). En el sombreado negro se destacan el DP4 original y el nivel que ofrece el mejor resultado de DP4+ (PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*). Conjuntos de bases: A: 6-31G*; B: 6-31G**; C: 6-31 + G**; D: 6-311G*; E: 6-311G**; F: 6-311 + G**.

Los resultados de las probabilidades de DP4+ (calculadas en los 24 niveles de teoría discutidos

anteriormente utilizando datos de protones y carbono) en la estereoasignación de los 48 ejemplos que se muestran en la Figura 4 se muestran en la Figura 5, incluyendo los valores de DP4 original correspondientes. Para facilitar una discusión adicional, se utilizó un sistema de puntuación simple introducido arbitrariamente en base al valor de probabilidad DP4+ calculado para un compuesto dado. Dependiendo de la confianza en la asignación correcta, se identificaron tres intervalos principales: > 95% (bueno), 50% -95% (medio) y <50% (malo), y a cada uno se le dio un puntaje diferente: 3, 1, y 0 puntos, respectivamente. Este estándar de "tres puntos para ganar", inspirado en muchas ligas deportivas, se implementó para resaltar asignaciones correctas altamente confiables, lo que representa el escenario óptimo.

A partir de los datos mostrados en la Figura 5, se pueden extraer varias conclusiones: (a) Las 24 nuevas probabilidades de DP4+ se comportaron mejor que DP4 original (hasta 2,4 veces), lo que indica una clara superioridad en la asignación estereoquímica de compuestos isoméricos. Es interesante observar que, en general, cuando el DP4 original apunta exitosamente al isómero correcto, nuestra probabilidad modificada también lo hace. Por el contrario, muchos casos de asignación incorrecta por DP4 podrían ser revertidos por DP4 + (por ejemplo, compuestos **73b**, **75b**, **76a**, **77b**, **78c**, **82b** y **86d**). En algunos ejemplos, tanto DP4 como DP4+ falló sistemáticamente al apuntar hacia el isómero correcto (por ejemplo, los compuestos **79a**, **79h** y **85c**).

Orígenes del rendimiento mejorado de DP4 +

El presente trabajo se basó en las principales hipótesis de que se podría desarrollar un mejor análisis estadístico similar al DP4 utilizando cálculos de RMN más precisos e incluyendo desplazamientos químicos no escalados en las ecuaciones de probabilidad. Tras establecer la clara mejora del método resultante (Figura 5), intentamos comprender mejor la contribución real de cada factor en el resultado del DP4+.

Primero, volvimos a calcular las probabilidades de los 48 ejemplos de validación que se muestran en la Figura 4 utilizando solo cambios desplazamientos escalados o sin escalar de los términos $sDP4+$ y $uDP4+$, respectivamente, de la ecuación 3. La Figura 6 muestra la performance de los cálculos resultantes promediados en los 24 niveles de teoría en estudio, junto con los valores correspondientes de DP4 y DP4 +.

Comparando los resultados obtenidos con $uDP4+$ y $sDP4+$ con los valores DP4+ correspondientes, resulta claro que la combinación de desplazamientos de RMN escalados y no escalados proporciona la capacidad de asignación más alta.

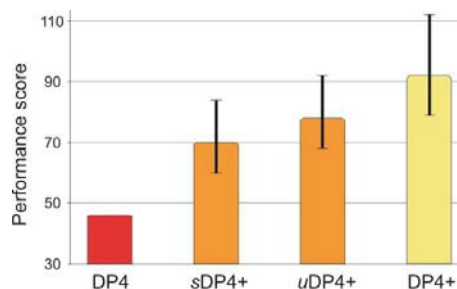


Figura 6. Los puntajes de desempeño general de DP4, $sDP4+$, $uDP4+$ and DP4+, promediadas para los 24 niveles de teoría bajo estudio.

Segundo teniendo en cuenta que $sDP4+$ se comporta mejor que DP4 independientemente del nivel de teoría (~1.5 veces en promedio), se hace evidente que computar los δ de RMN desde geometrías más robustas (B3LYP/6-31G* frente a MMFF) dio como resultado un efecto significativo en el resultado de probabilidad. Además, también se investigó la influencia del nivel de la teoría en los valores $[\sigma, \nu]$.

Finalmente, otro observación relevante es que el uso de δ no escalados por $uDP4+$ ofrece incluso mejores resultados (~1,7 veces en promedio) que los obtenidos con $sDP4+$, lo que sugiere que durante el proceso de escalado por regresión lineal algunos datos valiosos para diferenciar entre las estructuras candidatas se pierde.

¿Datos de protones o carbono? En esta etapa del estudio, calculamos las probabilidades de DP4+ utilizando datos de ^1H y ^{13}C . Los argumentos que respaldan esta elección se basan en la suposición principal de que se podría hacer una asignación mejor con la mayor cantidad de información disponible. Sin embargo, en este trabajo encontramos que los niveles de teoría con una distribución de errores t más estrecha (para ^1H pero no necesariamente para la serie ^{13}C)

condujeron a mejores probabilidades de DP4+ en términos de asignación estereoquímica correcta. Para determinar la importancia relativa de cada núcleo (^1H y ^{13}C), se descompusieron las probabilidades de DP4+ calculadas para los 48 compuestos de la Figura 4 en las correspondientes probabilidades de ^1H y ^{13}C . Los puntajes de rendimiento (promediados en los 24 niveles de teoría) se muestran en la Figura 7.

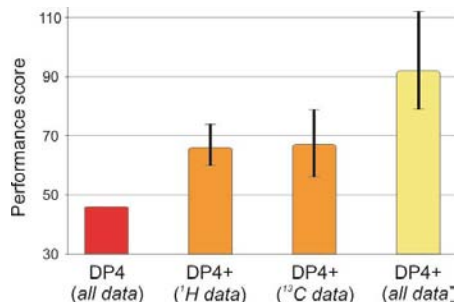


Figura 7. Puntajes de desempeño general de DP4, ^1H -DP4+, ^{13}C -DP4+ and DP4+, promediadas para los 24 niveles de teoría bajo estudio.

Curiosamente, la capacidad de asignación de DP4+ usando solo δ de ^1H o ^{13}C es en realidad mejor que la de DP4 con todos los datos, proporcionando evidencia adicional de la mejora ejercida por nuestras modificaciones en la formulación original. En promedio, ^1H -DP4+ y ^{13}C -DP4+ mostraron resultados globales similares, lo que sugiere que ninguno de los dos núcleos discrimina más que el otro. Naturalmente, la combinación de ambos ofrece una clara mejora en el rendimiento de DP4+ resultante, lo que indica que todos los datos son importantes y deben utilizarse cuando estén disponibles.

El análisis cuidadoso de los resultados nos permitió observar que una mala asignación estereoquímica eventualmente hecha por ^1H -DP4+ a menudo puede corregirse con ^{13}C -DP4+ (y viceversa).

Recientemente, en un interesante debate sobre qué núcleo es más relevante para la asignación estereoquímica, se descubrió que el protón es el más determinante [13]. Sin embargo, a partir de los datos presentados aquí, ambos núcleos son importantes y deben usarse cuando sea posible para calcular la probabilidad DP4+.

Metodología General

Las búsquedas conformacionales de compuestos se llevaron a cabo empleando el método de mecánica molecular MM+ y/o MMFF implementado en Spartan. Todos los cálculos semiempíricos, *ab initio* o DFT se llevaron a cabo empleando Gaussian 09. Se realizaron cálculos de frecuencia para caracterizar las estructuras optimizadas como mínimos de energía (ausencia de frecuencias imaginarias) o como estructuras de transición (una frecuencia imaginaria). Los cálculos de RMN se hicieron empleando el método GIAO, empleando funcionales DFT (mPW1PW91, B3LYP, etc) combinados con diversas funciones base, tanto en fase gas como solución.

Conclusión

Hemos desarrollado una nueva probabilidad (DP4+) como herramienta para la importante y difícil tarea de la asignación estructural o estereoquímica basada en la correlación de datos experimentales y calculados de RMN de moléculas orgánicas, cuando se dispone de un único set de datos experimentales. La inclusión de desplazamientos químicos no escalados en la formulación de probabilidad y el uso de niveles de teoría superiores tanto en el procedimiento de cálculo de RMN como en la optimización de la geometría dio como resultado una mejora significativa en el rendimiento general de la probabilidad DP4+, proporcionando resultados precisos y confiables en el establecimiento de la estereoquímica de 48 desafiantes compuestos isoméricos.

Simplificación del Procedimiento

Para simplificar el procedimiento de cálculo desarrollamos un archivo de *Excel* el cual se encuentra disponible como parte del [Supporting Information](#) de la publicación del trabajo [8] o en la

web del grupo de investigación en sarotti-NMR.weebly.com. Con sólo completar los *desplazamientos químicos experimentales* y los *tensores calculados*, indicando que tipo de núcleo se trata (^1H o ^{13}C) y la hibridación sp^2 cuando corresponda, la hoja de cálculo calcula los desplazamientos químicos escalados y no escalados, los errores y calcula la probabilidad asociada a cada uno de los posibles candidatos, siendo el que mejor ajuste los datos el isómero más probable.

	A	B	C	D	E	F	G	H	I
1	Functional		Solvent?		Basis Set		Type of Data		
2	mPW1PW91		PCM		6-31+G(d,p)		Shielding Tensors		
3									
12			DP4+	100.00%	0.00%	-	-	-	-
14	Nuclei	sp2?	Experimental	Isomer 1	Isomer 2	Isomer 3	Isomer 4	Isomer 5	Isomer 6
15	C		73.7	122.0	119.2				
16	C		46.5	146.8	147.0				
17	C	x	175.5	17.4	19.3				
18	C		11	184.2	177.7				
19	C	x	141.5	56.5	52.6				
20	C	x	125.9	72.2	73.7				
21	C	x	128.1	71.5	71.0				
22	C	x	127.3	72.0	72.3				
23	C		60.5	133.3	134.0				
24	C		13.9	181.0	180.8				
25									
26	H		5.09	26.22	26.66				
27	H		2.77	28.89	28.68				
28	H		1.12	30.57	30.04				
29	H	x	7.29	23.82	23.91				
30	H	x	7.29	23.79	23.88				
31	H	x	7.29	23.99	23.98				
32	H		4.12	27.32	27.50				
33	H		1.21	30.26	30.38				

Figura 8. Esquema de la planilla de Excel que permite realizar automáticamente el estudio DP4+.

Referencias.

1. Head-Gordon, M.; Artacho, E. *Phys. Today* **2008**, *61*, 58
2. Lewars, E. G. *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*; Springer, **2016**
3. (a) Foresman, J.; Frish, E. *Gaussian Inc., Pittsburg, USA 1996* (b) Koch, W.; Holthausen, M. C. *A chemist's guide to density functional theory*; John Wiley & Sons, **2015** (c) Cramer, C. J. *Essentials of computational chemistry: theories and models*; John Wiley & Sons, **2013**.
4. (a) Nicolaou, K. C.; Snyder, S. A. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 1012 (b) Suyama, T. L.; Gerwick, W. H.; McPhail, K. L. *Bioorg. Med. Chem.* **2011**, *19*, 6675 (c) Maier, M. E. *Nat. Prod. Rep.* **2009**, *26*, 1105.
5. a) Sikorska, J.; Hau, A. M.; Anklin, C.; Parker-Nance, S.; Davies-Coleman, M. T.; Ishmael, J. E.; McPhail, K. L. *J. Org. Chem.* **2012**, *77*, 6066 (b) Lei, H.; Yan, J.; Yu, J.; Liu, Y.; Wang, Z.; Xu, Z.; Ye, T. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 6533 (c) Khambay, B. P.; Beddie, D. G.; Hooper, A. M.; Simmonds, M. S. *Tetrahedron* **2003**, *59*, 7131 (d) Nishiwaki, H.; Fujiwara, S.; Wukirsari, T.; Iwamoto, H.; Mori, S.; Nishi, K.; Sugahara, T.; Yamauchi, S.; Shuto, Y. *J. Nat. Prod.* **2015**, *78*, 43 (e) Lee, S. U.; Asami, Y.; Lee, D.; Jang, J. H.; Ahn, J. S.; Oh, H. *J. Nat. Prod.* **2011**, *74*, 1284 (f) Lorenzo, P.; Álvarez, R.; de Lera, Á. R. *Eur. J. Org. Chem.* **2014**, *2014*, 2557 (g) Hamada, T.; White, Y.; Nakashima, M.; Oiso, Y.; Fijita, M. J.; Okamura, H.; Iwagawa, T.; Arima, N. *Molecules* **2012**, *17*, 9931 (h) Maeda, K.; Hamada, T.; Onitsuka, S.; Okamura, H. *J. Nat. Prod.* **2017**, *80*, 1446 (i) Festa, C.; De Marino, S.; Sepe, V.; D'Auria, M. V.; Bifulco, G.; Debitus, C.; Bucci, M.; Vellecco, V.; Zampella, A. *Org. Lett.* **2011**, *13*, 1532 (j) Kashinath, K.; Jachak, G. R.; Athawale, P. R.; Marelli, U. K.; Gonnade, R. G.; Reddy, D. S. *Org. Lett.* **2016**, *18*, 3178.
6. (a) Grimblat, N.; Sarotti, A. M. *Chemistry* **2016**, *22*, 12246 (b) Bagno, A.; Saielli, G. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015**, *5*, 228 (c) Tantillo, D. J. *Nat. Prod. Rep.* **2013**, *30*, 1079 (d) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839.

7. a) Zanardi, M. M.; Sarotti, A. M. *J. Org. Chem.* **2015**, *80*, 9371. b) Sarotti, A.M. *Org. Biomol. Chem.* **2013**, *11*, 4847.
8. Grimblat, N.; Zanardi, M. M.; Sarotti, A. M. *J. Org. Chem.* **2015**, *80*, 12526.
9. Smith, S. G.; Goodman, J. M. *J. Org. Chem.* **2009**, *74*, 4597.
10. Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946.
11. (a) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* 2012, *112*, 1839. (b) Bagno, A.; Saielli, G. *Wiley Interdisciplinary Reviews: Computational Molecules Science* 2015, *5*, 228. (c) Tantillo, D. J. *Nat. Prod. Rep.* 2013, *30*, 1079. (d) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* 2007, *107*, 3744–3779. (e) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. *Nat. Protoc.* 2014, *9*, 643.
12. (a) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254. (b) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2012**, *77*, 6059.
13. (a) Chini, M. G.; Riccio, R.; Bifulco, G. *Eur. J. Org. Chem.* **2015**, *6*, 1320. (b) Marell, D. J.; Emond, S. J.; Kulshrestha, A.; Hoye, T. R. *J. Org. Chem.* **2014**, *79*, 752.