

## Redefiniendo el impacto del análisis conformacional por Boltzmann en la asignación estereoquímica de moléculas polares y flexibles mediante cálculos de RMN.

Zanardi, María Marta<sup>†</sup>; Maribel Oriana Marcarino<sup>‡</sup>; Sarotti, Ariel Marcelo<sup>‡</sup>.

<sup>†</sup>Instituto de Ingeniería Ambiental, Química y Biotecnología Aplicada (INGEBIO), Facultad de Química e Ingeniería del Rosario, Pontificia Universidad Católica Argentina, Av. Pellegrini 3314, (S2002QEO) Rosario, Argentina.

<sup>‡</sup>Instituto de Química Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario 2000, Argentina

### Resumen

Actualmente la Resonancia Magnética Nuclear constituye una de las herramientas más poderosas para la elucidación estructural de nuevas moléculas orgánicas complejas. A pesar de los enormes avances en la disciplina la determinación de la estructura tridimensional de nuevos compuestos de modo exclusivamente experimental es muchas veces extremadamente difícil o imposible; lo que conlleva frecuentemente a la publicación de estructuras erróneas. La química computacional, ha contribuido enormemente a prevenir estas situaciones, mediante la construcción de distintas herramientas de correlación de datos experimentales con cálculos cuánticos, las cuales permiten reforzar la confianza en la asignación estereoquímica de nuevos productos tanto naturales como sintéticos. Entre ellas se destaca la probabilidad que hemos desarrollado denominada DP4+. Sin embargo, la asignación *in silico* de compuestos polihidroxilados continúa representando un gran desafío debido al problema no resuelto hasta ahora de la inapropiada descripción de las distribuciones conformacionales. Notamos que la calidad de las predicciones dependía en gran medida de la distribución conformacional proporcionada por las energías DFT. En este trabajo, desarrollamos un enfoque estocástico conceptualmente novedoso basado en la creación y evaluación de conjuntos artificiales aleatorios, que podría proporcionar un nuevo paradigma para calcular las propiedades de RMN de moléculas flexibles. La estrategia se probó con éxito en las plataformas DP4 y DP4+ utilizando un gran conjunto de compuestos pertenecientes a la familia de las Hiacintacinas. Éstas son miembros importantes de las pirrolizidinas, con varios compuestos cuyas estructuras resultaron ambiguas, fueron revisadas o no han sido aún verificadas.

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

### Abstract

Currently, Nuclear Magnetic Resonance is one of the most powerful tools for the structural elucidation of new complex organic molecules. Despite the enormous advances in the discipline, the determination of the three-dimensional structure of new compounds exclusively in experimental way is often extremely difficult or impossible; which frequently leads to the publication of erroneous structures. Computational chemistry has contributed greatly to prevent these situations, through the construction of different tools for correlating experimental data with quantum calculations, which increase the confidence in the stereochemical assignment of new natural and synthetic products. Among them, the probability we have developed named DP4+ is highlighted. However, *in silico* assignment of polyhydroxylated compounds represents a major challenge given the thus far unsolved problem of inappropriate description of the conformational amplitudes. We noticed that the quality of the predictions strongly depended on the conformational landscape provided by DFT energies. Herein, we report a conceptually novel stochastic approach based on the creation and evaluation of random artificial ensembles, which could provide a new paradigm for computing NMR properties of flexible molecules. The strategy was successfully tested under the DP4/DP4+ platforms using a large set of compounds belonging to the hyacinthacine family. Hyacinthacines are important members of the pyrrolizidines, with several compounds having ambiguous, revised or unverified structures.

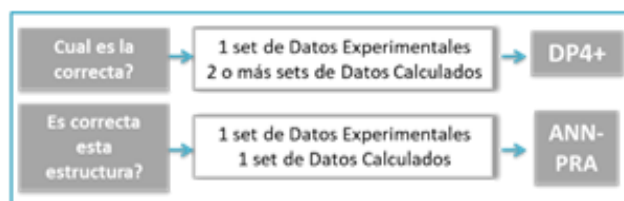
**Keywords:** Computational chemistry, NMR, Stereochemical Assignment, DP4+.

### Introducción

La elucidación estructural de nuevas moléculas complejas es una etapa fundamental (y muchas veces crítica) en el descubrimiento de nuevos compuestos activos, ya que su estructura tridimensional se encuentra fuertemente ligada con sus propiedades químicas, físicas y biológicas. La espectroscopia de resonancia magnética nuclear (RMN) es sin duda la técnica más poderosa de determinación estructural. A pesar de los continuos avances en el campo, la cantidad de asignaciones erróneas que se publican demuestra que continúa habiendo desafíos en el área. El empleo de cálculos cuánticos de RMN para facilitar la etapa de elucidación ha crecido exponencialmente en los últimos años, consolidándose como un complemento ideal de los estudios experimentales. En la actualidad, se dispone de una amplia pléthora de poderosas metodologías, que constituyen herramientas extremadamente valiosas para resolver problemas estructurales.<sup>1</sup>

En los últimos años hemos realizado importantes avances en el empleo de modelado molecular en combinación con

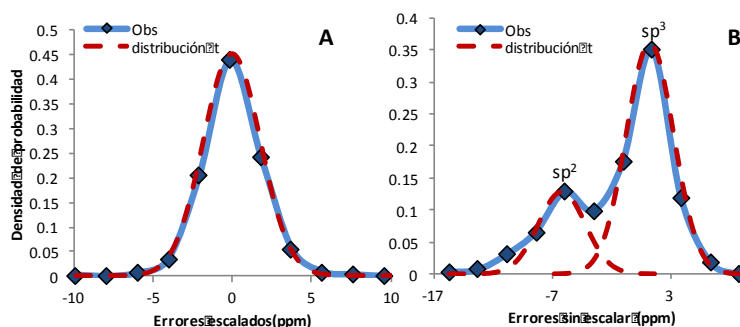
métodos estadísticos y quimiométricos para la elucidación estructural de moléculas orgánicas complejas. El uso de cálculos cuánticos de propiedades espectroscópicas ha resultado una herramienta muy valiosa para complementar los experimentos de Resonancia Magnética Nuclear (RMN), muchas veces insuficientes o ambiguos en la determinación de aspectos estereoquímicos en moléculas quirales. Entre los desarrollos más destacados se encuentran las ANN-PRA (*Artificial Neural Network Pattern Recognition Analysis*)<sup>2</sup> empleadas en validación estructural y la probabilidad DP4+<sup>3</sup> utilizada para la asignación de configuraciones relativas de compuestos quirales (Figura 1). Para validación estructural, es decir determinar si la estructura propuesta es correcta o no, desarrollamos un método rápido, sencillo y confiable, que mediante la utilización de cálculos de RMN bidimensionales combinado con reconocimiento de patrones vía redes neuronales artificiales (ANN-PRA), permite comparar único *set* de datos experimentales con un *set* de datos calculados.<sup>2</sup>



**Figura 1.** Representación esquemática de las herramientas desarrolladas para elucidación estructural

Estos trabajos han tenido muy buena repercusión dentro de la comunidad científica, pero sin dudas DP4+ ha sido la más utilizada por distintos grupos para la elucidación estructural de nuevos productos naturales y sintéticos. Esta probabilidad permite correlacionar dos o más estructuras candidatas a un sólo *set* de datos experimentales asignando a cada una su probabilidad (*i* tiene más chance que *j* de ser la estructura correcta). DP4+ mostró un incremento notable en su performance respecto a la formulación original debido fundamentalmente al nivel de teoría empleado en los cálculos y la introducción de datos no escalados.

Brevemente, la asignación estereoquímica realizada tiene su origen en la distribución estadística de errores entre los datos experimentales y calculados para los distintos estereoisómeros, *e* (definidos como  $\delta_{calc} - \delta_{exp}$ ). La probabilidad DP4+ se sustenta sobre la base que los errores escalados ( $\delta_s - \delta_{exp}$ ) son variables aleatorias normalmente distribuidas (en realidad, ajustan mejor a una distribución *t*), y que cuando se emplean  $\delta$  no escalados hay una distribución *t* para núcleos  $sp^2$  y otra para  $sp^3$  (Figura 2).



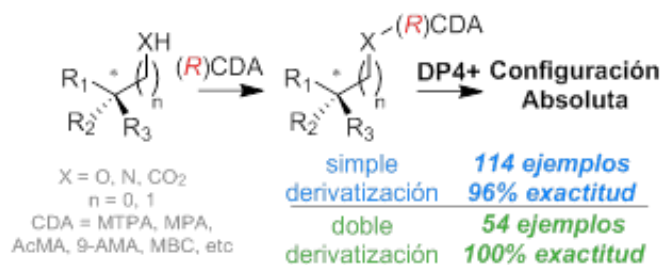
**Figura 2.** Distribución de errores escalados y no escalados.

Luego, la probabilidad asociada al *i*-ésimo error puede calcularse conociendo los parámetros [ $\mu, \sigma, \nu$ ] que definen dichas distribuciones. La probabilidad global correspondiente a cada isómero se construye multiplicando todas las probabilidades individuales, y empleando el Teorema de Bayes, se asigna un valor porcentual a cada candidato (Eq. 1).<sup>3</sup>

$$\begin{aligned}
 & \text{DP4+} \\
 & P(i) = P(i)_s \cdot P(i)_u \quad \text{Eq. 1} \\
 & \uparrow \text{sDP4+} \\
 & P(i)_s = \frac{\prod_{k=1}^N (1 - T^v_s(\delta^i_{s,k} - \delta_{exp,k})/\sigma_s)}{\sum_{j=1}^m [\prod_{k=1}^N (1 - T^v_s(\delta^j_{s,k} - \delta_{exp,k})/\sigma_s)]} \\
 & \uparrow \text{uDP4+} \\
 & P(i)_u = \frac{\prod_{k=1}^N (1 - T^v_u(\delta^i_{u,k} - \delta_{exp,k}) - \mu_u)/\sigma_u)}{\sum_{j=1}^m [\prod_{k=1}^N (1 - T^v_u(\delta^j_{u,k} - \delta_{exp,k}) - \mu_u)/\sigma_u]} \\
 & T^v, \sigma, \mu, \text{ y } \delta_s \text{ calculados sobre geometrías B3LYP/6-31G*}
 \end{aligned}$$

Según el nivel de teoría empleado para el cálculo de los tensores magnéticos los parámetros estadísticos se ven afectados y de este modo varía la capacidad predictiva de DP4+. Para demostrar su gran capacidad en la asignación estereoquímica se decidió evaluar su performance en casos muy desafiantes, empleando el nivel de teoría que mejores resultados arrojó durante la etapa de validación del desarrollo. De los 24 niveles analizados para DP4+, se decidió emplear la optimización de la geometría a nivel B3LYP/6-31G\* y el cálculo de RMN a nivel PCM/mPW1PW91/6-31+G\*\* empleando cloroformo como solvente. Demostramos que pueden obtenerse excelentes resultados incluso en la asignación estereoquímica de moléculas muy complejas de determinar experimentalmente como son los espiroepóxidos y epóxidos terminales.<sup>4</sup> Luego nos enfrentamos a dos desafíos evaluar la performance de DP4+ en moléculas más flexibles (conteniendo estereoclusters) y su comportamiento en sistemas altamente polares (como poli alcoholes) cuyos espectros deben ser registrados en solventes próticos deuterados. Respecto al primer tópico, realizamos los estudios de evaluación del empleo de DP4+ para la determinación de la configuración absoluta (CA) una de las etapas más importantes y desafiantes durante la elucidación estructural de las moléculas quirales.

Como alternativa a los análisis clásicos de los métodos tipo Mosher, propusimos una novedosa estrategia que consiste en resolver la CA más probable de los compuestos quirales mediante experimentos de simple derivatización, acoplados a cálculos cuánticos de los  $\delta$  de RMN y el análisis de DP4+. Se seleccionó un conjunto de compuestos variados, entre ellos alcoholes y aminas primarios y secundarios, ácidos carboxílicos y cianhidrinas terciarias de diversa complejidad estructural y libertad conformacional. Respecto a la naturaleza del CDA, nuestra selección cubrió los más populares, incluyendo MTPA, MPA, MA, Ac-MA, 9-AMA y MBC. Los resultados obtenidos en este estudio fueron excelentes tanto para simple como doble derivatización (Figura 3).<sup>5</sup>



**Figura 3.** Metodología empleada para la determinación de la CA y resultados obtenidos para ambos métodos DP4+ y DIP.

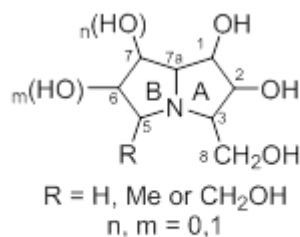
Es bien sabido que los cálculos de sistemas flexibles ofrecen una complicación adicional. En la escala de tiempo de RMN, la interconversión conformacional es generalmente rápida; por lo tanto, los desplazamientos químicos observados serán un promedio ponderado de todas las conformaciones. En un esfuerzo por reproducir este efecto mediante cálculos de RMN, los  $\delta$  y los valores  $J$  (constante de acoplamiento) de cada conformación individual deben calcularse por separado para producir un valor promedio del conjunto. Suponiendo equilibrio térmico, la contribución de cada conformero viene dada por el factor de ponderación de Boltzmann correspondiente ( $w_i$ ), estimado a partir de las energías relativas calculadas típicamente en los niveles de DFT. Por lo tanto, si la predicción de esas energías es inadecuada posiblemente nos encontremos con inconvenientes en los cálculos de RMN.

De este modo, independientemente de la estrategia matemática empleada en la correlación de datos calculados y experimentales, debe enfatizarse que en todos los casos la certeza de la asignación estereoquímica está fuertemente relacionada con la calidad de la predicción de RMN en sí misma. Es decir si los tensores magnéticos calculados son poco fiables impactará sobre el poder de clasificación de las distintas herramientas actualmente disponibles como CP3, DP4, DU8+, CASE-3D, DiCE, etc.<sup>1</sup>

## Resultados y discusión

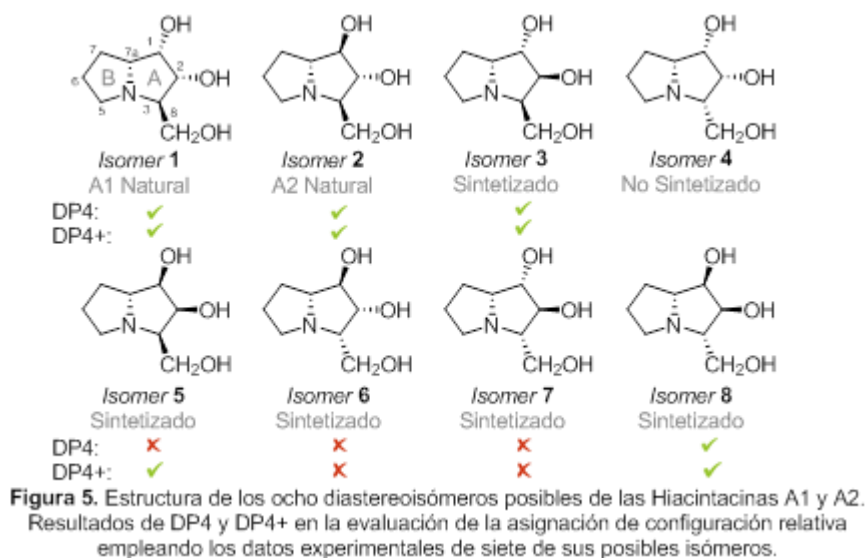
Para la evaluación del comportamiento de la probabilidad DP4+ en moléculas polares, se eligió como caso testigo las iminoazúcares y dentro de ellas las Hiacintacinas, una familia con varios miembros cuya estructura debió ser revisada, fue rechazada o aún no ha sido verificada. La característica estructural de esta familia es un esqueleto de 7a-R-1,2-hidroxi-3-hidroximetilpirrolizidina. Luego pueden clasificarse dependiendo del patrón de sustitución en C-5 (H, Me o -CH<sub>2</sub>OH), C-6 (H o OH) y C-7 (H o OH) en el anillo B, en A, B y C dependiendo del número de grupos hidroxilo/hidroximetilo presentes en el anillo B, 0, 1 o 2 respectivamente. Hasta la fecha, se aislaron 19 miembros diferentes de la familia de extractos de la familia *Hyacinthaceae*, y se denominaron A1-7, B1-7 y C1-5 (Figura 4).<sup>6</sup>

Estas características hacen suponer que su asignación estereoquímica es *a priori* compleja. En las últimas décadas, los iminoazúcares (también denominados aza azúcares) han recibido mucho interés principalmente debido a sus prometedoras actividades biológicas. En particular, los alcaloides polihidroxilados que imitan las estructuras de los azúcares están muy extendidos en las plantas y se ha demostrado que inhiben las glucosidasas.<sup>7</sup>

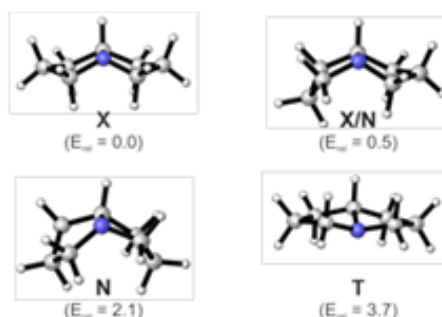


**Figura 4.** Estructura general de las Hyacintacinas

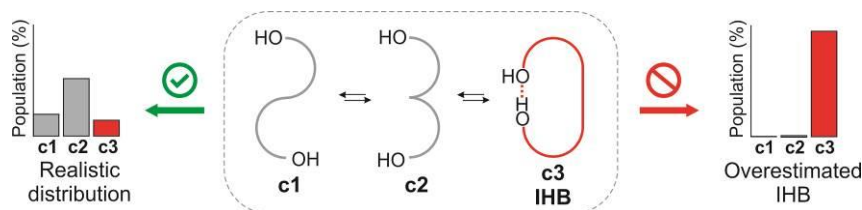
Se evaluó la capacidad de DP4 y DP4+ para establecer la configuración relativa de 7 isómeros de los miembros más simples de la familia de las hiacintacinas, las cuales contienen sustituyentes solo en el anillo A dando lugar a 8 posibles diastereoisómeros (Figura 5).



Empleando niveles estándar para los cálculos B3LYP/6-31G\*\*//MMFF (para DP4) y PCM/mPW1PW91/6-31+G\*\*//B3LYP/6-31G\* (para DP4+), ni DP4 ni DP4+ permitieron asignar correctamente todos los isómeros, identificando erróneamente 3 y 2 compuestos respectivamente. Cabe destacar que ambas herramientas arrojaron probabilidades sorprendentemente bajas (<0,1%) para los isómeros 6 y 7, siendo las estructuras correctas clasificadas entre las últimas posiciones. Lo cual representó un resultado francamente inusual. Estas observaciones nos llevaron a estudiar profundamente cuales eran los motivos de esta disminución en la performance. Se realizó un estudio conformacional de esta serie de compuestos conteniendo anillos de pirrolidina, y encontramos que el sistema bicíclico podía adoptar cuatro posibles conformaciones (tres con fusión *cis*: **X** *exo*, **X/N** *exo-endo*, **N** *endo* y una con fusión *trans* **T**) (Figura 6).<sup>6</sup> Y luego las correlacionamos con las conformaciones más estables obtenidas para cada uno de los isómeros. Esto nos permitió llegar a la conclusión que los métodos DFT, al menos a los niveles de teoría evaluados, tienden a sobreestimar la contribución de las conformaciones con *enlaces de hidrógeno intramoleculares* (IHB) por una descripción inadecuada de la interacción soluto-solvente (Figura 7).



**Figura 6.** Geometrías de las distintas conformaciones posibles para el anillo de pirrolidina a B3LYP/6-31G\*. Las energías relativas están informadas en Kcal/mol.

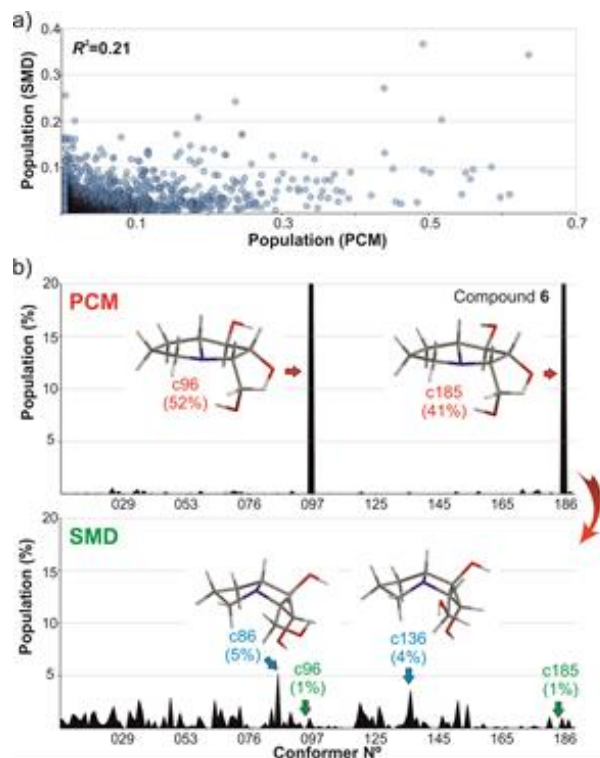


**Figura 7.** Distribución conformacional simplificada de compuestos polihidroxiados flexibles

Como la distribución conformacional real en solventes prácticos surge del delicado equilibrio entre enlaces H intra e intermoleculares, estos errores en las energías derivan en una mala descripción de la distribución de Boltzmann para algunos diastereoisómeros, y por ende se traduce en errores en los desplazamientos químicos calculados. Las malas descripciones de datos de RMN que se obtienen para algunos estereoisómeros dificultan a las herramientas probabilísticas basadas en la comparación, como DP4+ poder efectuar una buena predicción.

La remoción de las conformaciones espurias por distintas estrategias (como remoción manual, o eliminación sistemática de las conformaciones más estables para todos los isómeros de un compuesto) permitió asignar correctamente todos los ejemplos ensayados. Estos resultados fueron el puntapié para intentar encontrar una solución más generalizada a los errores de subestimación de las energías por los métodos DFT en moléculas polihidroxiadas con múltiples enlaces de hidrógeno intramoleculares. Para ello, se decidió estudiar la familia completa de las Hiacintacinas a fin de encontrar una posible estrategia para prevenir malas asignaciones.

La limitación de DFT para reproducir correctamente la dinámica conformacional de los sistemas flexibles ha sido ampliamente reconocida y se desarrollaron diferentes enfoques para mitigar esta fuente de error. Una de las estrategias implica calcular las energías relativas utilizando niveles teóricos más apropiados. El otro implica refinar el análisis de Boltzmann aliviando el efecto de conformaciones inadecuadas mediante el uso de información adicional de RMN, como NOE o constantes de acoplamiento  $J$ . Sin embargo, la asignación precisa *in silico* de moléculas flexibles y polares que presentan redes IHB complejas empleando solo datos de RMN de  $^1\text{H}$  y  $^{13}\text{C}$  no se ha resuelto hasta ahora.<sup>8</sup>

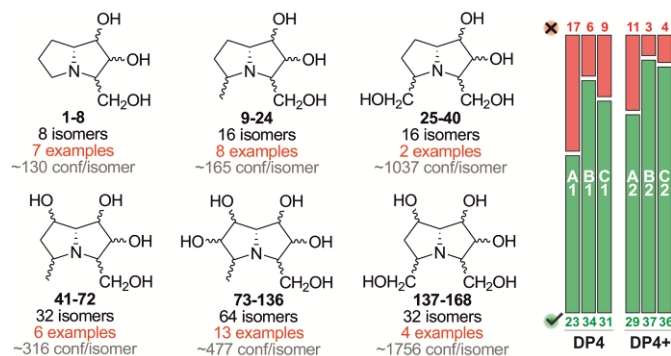


**Figura 8.** a) Diagrama de dispersión de las amplitudes de Boltzmann calculadas para los compuestos **1-168** a PCM/mPW1PW91/6-31+G\*\*//B3LYP/6-31G\* (eje x) y SMD/B3LYP/6-31+G\*\*//B3LYP/6-31G\* (eje y) levels. b) Distribución conformacional del compuesto **6** calculado a los niveles de teoría PCM/mPW1PW91/6-31+G\*\*//B3LYP/6-31G\* y SMD/mPW1PW91/6-31+G\*\*//B3LYP/6-31G\*. Entre paréntesis se indican algunas conformaciones seleccionadas con su población relativa.

Una posible ruta para mitigar la fuente de error provocados por los IHB, fue evaluar diferentes niveles de teoría para refinar las amplitudes de Boltzmann, lo que además permitiría una mejor estimación de la incertidumbre energética proporcionada por el tratamiento DFT estándar. Inicialmente, exploramos a fondo el análisis conformacional de **1-8** con cuatro funcionales DFT populares (mPW1PW91, B3LYP, M06-2X y  $\omega$ B97XD) y dos modelos de solvatación ampliamente empleados: el PCM y SMD.

Los resultados indicaron marcadas diferencias al correlacionar las distribuciones de Boltzmann calculadas con PCM y SMD ( $0.04 < R^2 < 0.27$ - Fig. 8a), mientras que se observaron efectos más moderados con el funcional empleado. Esto se validó aún más con el conjunto completo de isómeros de la familia de las hiacintacinas (compuestos **1-168**). Como tendencia general, la serie PCM mostró ventanas de energía más grandes, con pozos más profundos en la superficie potencial y menos conformaciones contribuyentes que las calculadas con SMD, consistente con hallazgos anteriores. El modelo de solvatación más realista (SMD) se atenuó la tendencia de PCM a sobreestimar la estabilidad de las formas IHB. Este interesante efecto se puede ilustrar en la Figura 8.

El efecto del modelo de solvatación sobre las distribuciones de Boltzmann tuvo una gran influencia en la asignación estereoquímica de 40 miembros naturales y no naturales de la familia de las hiacintacinas validadas por síntesis. Como se muestra en la Figura 9, el enfoque DP4 estándar (conjunto A1) clasificó exitosamente 23 ejemplos y falló al asignar con baja probabilidad el isómero correcto en 17 casos. Como era de esperar, los resultados obtenidos con el análisis DP4+ estándar (conjunto A2) fueron superiores (con 29 aciertos y 11 fallos), aunque lejos de proporcionar un resultado óptimo. El refinamiento de energía SMD mejoró considerablemente los resultados, especialmente cuando se combinó con B3LYP (conjuntos B1 y B2), pero aún presentó asignaciones incorrectas (6-9 para DP4, 3-4 para DP4+). Después de un análisis detallado de todos los datos, llegamos a la conclusión de que esto se debía a que SMD no pudo revertir con éxito la tendencia energética de algunas conformaciones espurias en esos casos.

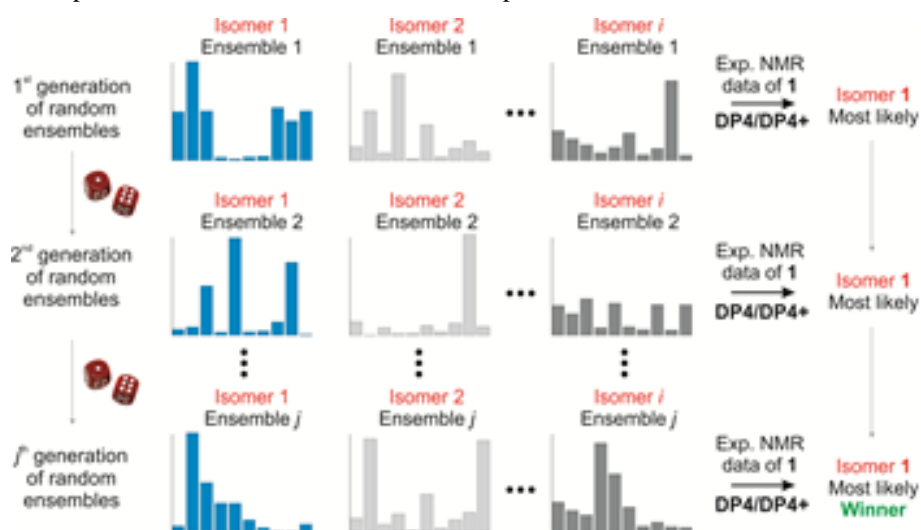


**Figura 9.** Izquierda) Representación de todos los posibles isómeros de la familia de las Hiacintacinas (**1-168**). Derecha) Resultados de la clasificación obtenidos mediante el uso de 3 arreglos seleccionados. Para **DP4**: A1) B3LYP/6-31G\*\*//MMFF; B1) SMD/B3LYP/6-31+G\*\*//MMFF; C1) SMD/M06-2X/6-31+G\*\*//MMFF. For **DP4+**: A2) PCM/mPW1PW91/6-31+G\*\*//B3LYP/6-31G\*; B2) SMD/B3LYP/6-31+G\*\*//B3LYP/6-31G\*; C2) SMD/M06-2X/6-31+G\*\*//B3LYP/6-31G\*

Si bien fue sencillo determinar que las energías PCM pueden estar lejos de ser aceptables para los sistemas que presentan interacciones IHB, y pueden corregirse parcialmente por la inclusión de SMD, la tarea de resolver el problema aún no era clara. Con la incertidumbre energética proporcionada por estos métodos (al menos  $\pm 2$  kcal/mol), cualquier conformación errónea puede competir con una adecuada (o incluso anularla). El problema se basa en el hecho de que la decisión sobre la correcta contribución de una conformación determinada puede ser difícil (o imposible) de tomar sin más pruebas experimentales de RMN, como el acoplamiento  $J$  o la información NOE. Establecer las distribuciones conformacionales ajustando los datos experimentales y computados surge como una estrategia útil para abordar el problema cuando el número de conformaciones es bajo. Sin embargo, un mayor número de conformaciones podría conducir a un sobreajuste con el riesgo de que un isómero incorrecto acabe mostrando un mejor ajuste fortuito que el correcto.

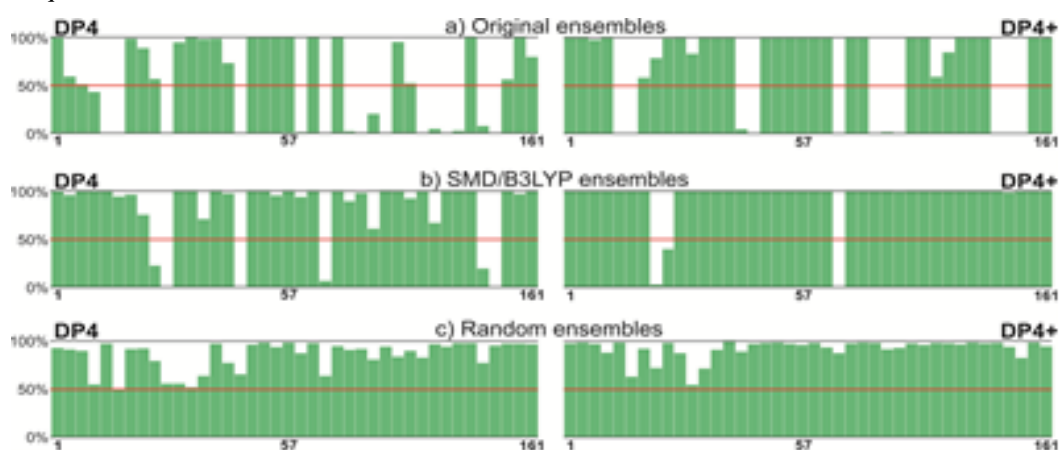
En este contexto, exploramos un enfoque más drástico al eliminar por completo las energías relativas dadas por los cálculos de DFT (claramente, la fuente principal del problema) y, alternativamente, crear y evaluar diferentes arreglos (ensembles) con distribuciones conformacionales aleatorias seguidas de promediar el DP4/DP4+ resultante (Figura 10). Esto implica romper con el enfoque tradicional de confiar en los resultados proporcionados por un único arreglo (ya sea generado con energías DFT o por ajuste). Si bien un arreglo individual podría eventualmente conducir a una predicción incorrecta similar a la de DP4, especulamos que los resultados proporcionados por un gran conjunto de arreglos podrían reflejar la asignación correcta. Las distribuciones barajadas se generaron mediante la combinación de dos componentes: generación aleatoria de subconjuntos eliminando conformaciones seguido de asignación aleatoria de energía relativa de

las formas restantes. Consideramos 15 estrategias alternativas dentro de este marco, y cada una se probó después de 1.000.000 de iteraciones por isómero candidato utilizando un script Matlab.<sup>8</sup>



**Figura 10.** Generación de arreglos al azar.

Como se muestra en la Figura 11c, el rendimiento de clasificación general promedio de este enfoque fue excelente, con el 100% de los compuestos clasificados correctamente tanto por DP4 como por DP4+, una puntuación notable dada la aleatoriedad de las amplitudes de la población. La naturaleza estocástica del enfoque no condujo a valores de probabilidad extremos (> 99,9%) incluso para aquellos compuestos asignados con alta confianza, lo que por el contrario representa una situación bastante común en el tratamiento estándar DP4/DP4+. Sin embargo, esta fue una propiedad valiosa para evitar la conocida tendencia de ambos métodos a sobrestimar la tasa de probabilidad. Las diferentes estrategias para generar distribuciones aleatorias arrojaron resultados diferentes, pero aun cualitativamente similares, lo que indica la posibilidad de realizar el análisis con solo pocas definiciones. Además, las tendencias continúan utilizando iteraciones más bajas durante el procedimiento (encontramos que solo 1.000 iteraciones dieron resultados estadísticamente significativos), que se pueden implementar fácilmente en Matlab, Python o entornos relacionados, requiriendo solo unos minutos usando computadoras estándar. Sobre la base de lo dicho anteriormente, llegamos a la conclusión de que este enfoque representa una estrategia simple, factible y conceptualmente novedosa para fortalecer o desafiar la asignación estereoquímica realizada por los formalismos estándar DP4/DP4+, o sus variantes SMD actualizadas aquí.



**Figura 11.** Resultados DP4/DP4+ obtenidos al correlacionar los datos de RMN experimentales de 40 Hiacintacinas conocidas con los datos calculados para los compuestos **1-168** empleando distintos ensembles.

También exploramos estas metodologías para establecer la estructura correcta de 8 Hiacintacinas naturales cuyas estructuras han sido rechazadas por síntesis total (C1, C3, C5 y B7), o permanecieron sin verificar (A4, A5, B6 y C4), Figura 12.<sup>8</sup>

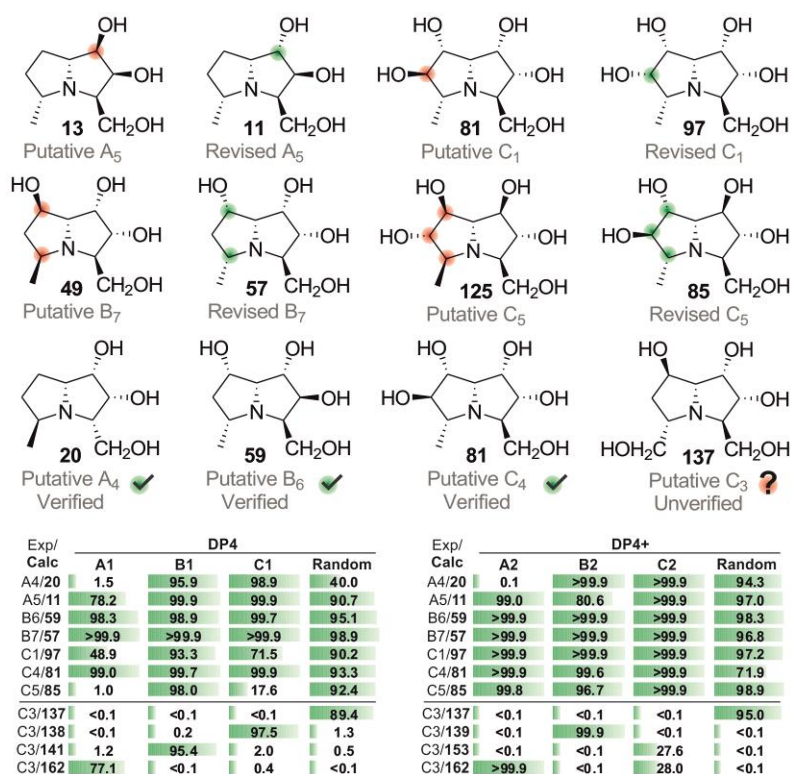


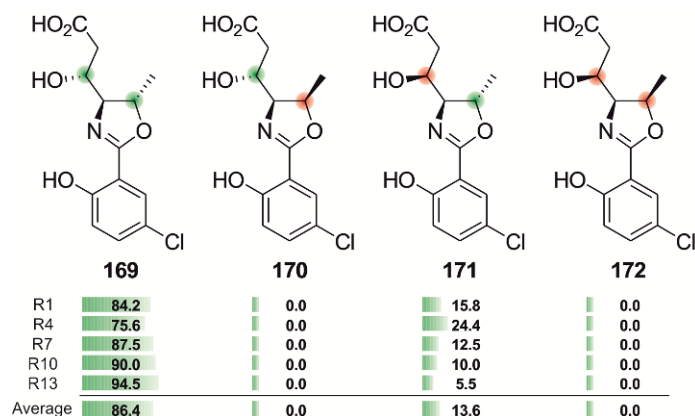
Figura 12. Stereochemical verification or reassignment of natural hyacinthacines

Las Hiacinthacinas C1 y C4 fueron aisladas de *Hyacinthoides non-scripta* y *Scillasocialis* en 1999, y 2007, respectivamente. A estos compuestos se les han atribuido estructuras idénticas (**81**), a pesar de que poseen datos espectroscópicos de RMN  $^1\text{H}$  y RMN  $^{13}\text{C}$  diferentes. La estructura de C1 fue revisada por síntesis como el epímero C-6 de **81** (compuesto **97**), mientras que la estructura de C4 permaneció desconocida. En el caso de Hiacinthacina C5 (coaislado con C4), diferentes grupos lograron la síntesis total de la estructura propuesta (**125**) encontrando discrepancias en los datos de RMN entre muestras sintéticas y originales. Finalmente, en 2018 el grupo de Pyne lo reasignó como **85** después de la síntesis de siete isómeros. En excelente concordancia con estas observaciones, nuestros cálculos sugirieron que las estructuras correctas de C1, C4 y C5 son **97**, **81** y **85**, respectivamente.

Luego probamos aquellos ejemplos que fueron rechazados por síntesis pero no reasignados, como el caso de B7 y C3 (ambos aislados en 2007). Pyne y colaboradores demostraron por síntesis total que el supuesto B7 (**49**) era incorrecto, lo que sugiere que la estructura real debería ser **57** (la misma propuesta para B5) dada la estrecha coincidencia en los datos de RMN de  $^{13}\text{C}$  de ambos productos naturales. Un análisis más detallado de los datos de RMN de  $^1\text{H}$  arrojó diferencias más significativas (hasta 0,2 ppm), pero los autores invocaron el efecto conocido del pH y la concentración sobre los cambios químicos de los protones de este tipo de compuestos. De acuerdo con la propuesta de Pyne, empleando análisis DP4 / DP4+ la estructura más probable de B7 fue **57**. En el caso de Hiacinthacina C3, observamos una alta heterogeneidad de resultados al cambiar el método para estimar las amplitudes conformacionales. Curiosamente, los cálculos aleatorios de DP4/ P4+ apoyaron el supuesto **137** como el C3 correcto, aunque esto está en conflicto con los hallazgos sintéticos de Yoda et al. Las diferencias en los datos de RMN de muestras de C3 naturales y sintéticas podrían atribuirse a los efectos del pH y la concentración, pero en nuestra opinión son demasiado grandes. Sin tener una muestra auténtica del producto natural, ni copias de los espectros de RMN 1D y 2D, la determinación inequívoca de la estructura real de C3 permanece hasta ahora sin resolver.

Finalmente, exploramos las estructuras de A4, A5 y B6, cuyas estructuras fueron sugeridas por el equipo de aislamiento pero no verificadas ni rechazadas por síntesis. En los casos de A4 y B6 nuestros cálculos apoyaron fuertemente las asignaciones originales (**20** y **59**, respectivamente), pero este no fue el caso de A5. La estructura propuesta (**13**) se encontró altamente improbable por DP4/DP4+, identificando el epímero C-1 (**11**) como el A5 real con los cuatro métodos aquí estudiados para calcular las amplitudes conformacionales. El análisis de los datos de acoplamiento  $J$  reportados para el producto natural apoyó nuestra asignación, con  $J_{1,7a}=4.5$  Hz consistente con hiacinthacinas naturales con configuración 1S (A1, A6, A7, B1, B2, B3, B5, C1 y C2, con Rango  $J_{1,7a}$  en 3.9-4.5 Hz). Por el contrario, se informaron valores más altos de  $J_{1,7a}$  para los compuestos con configuración 1R (como el rango A2, A3, B4 y C5 en 7.1-7.6 Hz).





**Figura 13.** Probabilidades generales de DP4+ calculadas para los compuestos **169-172** utilizando cinco tipos diferentes de arreglos aleatorios (10.000 iteraciones cada uno).

Para extrapolar esta metodología a otros sistemas distintos de hincintacinas evaluamos la performance del nuevo enfoque en un caso conflictivo como fue la Leptazolina A (**169**).<sup>9</sup> Al hacer coincidir los datos experimentales de RMN reportados para **169** con los valores de desplazamiento químico de cuatro posibles diastereoisómeros recalculados en el nivel PCM /mPW1PW91/ 6-31+G\*\*//B3LYP/6-31G\* y promediados usando cinco tipos de conjuntos aleatorios (10,000 iteraciones cada una), el candidato correcto fue identificado por DP4+ con un promedio de confianza alto, lo que sugiere la amplia aplicabilidad de la metodología.

## Conclusión

Los ejemplos mostrados demuestran los posibles inconvenientes que se encuentran al seguir los procedimientos estándar para los cálculos de RMN cuando se aplican a moléculas flexibles y polihidroxiadas. Al calcular las energías relativas utilizando el modelo de solvatación SMD se obtienen resultados sistemáticamente mejorados. Además, la aplicación de la estrategia de conjunto aleatorio brinda la oportunidad de introducir una alternativa nueva y emocionante que se aplicará en la elucidación estructural mediante cálculos de RMN al reconocer las limitaciones de los enfoques actuales de DFT para estimar adecuadamente la distribución conformacional de los sistemas polares flexibles en el agua (o solventes relacionados). Dichos procedimientos son de muy bajo requerimiento en términos de costos computacionales adicionales, pero podrían mejorar en sustancialmente la confianza en las conclusiones.

## Referencias

- 1 N. Grimblat and A. M. Sarotti, *Chem. - A Eur. J.*, 2016, **22**, 12246–12261.
- 2 M. M. Zanardi and A. M. Sarotti, *J. Org. Chem.*, 2015, **80**, 9371–9378.
- 3 N. Grimblat, M. M. Zanardi and A. M. Sarotti, *J. Org. Chem.*, 2015, **80**, 12526–12534.
- 4 M. M. Zanardi, A. G. Suárez and A. M. Sarotti, *J. Org. Chem.*, 2017, **82**, 1873–1879.
- 5 M. M. Zanardi, F. A. Biglione, M. A. Sortino and A. M. Sarotti, *J. Org. Chem.*, 2018, **83**, 11839–11849.
- 6 M. M. Zanardi, M. A. Sortino and A. M. Sarotti, *Carbohydr. Res.*, 2019, **474**, 72–79.
- 7 C. Sousa, R. Mendes, F. Costa, V. Duarte, A. Fortes and M. Alves, *Curr. Org. Synth.*, 2013, **11**, 182–203.
- 8 M. M. Zanardi, M. O. Marcarino and A. M. Sarotti, *Org. Lett.*, 2020, **22**, 52–56.
- 9 J. Bhandari Neupane, R. P. Neupane, Y. Luo, W. Y. Yoshida, R. Sun and P. G. Williams, *Org. Lett.*, 2019, **21**, 8449–8453.